frontiers | Research Topics

Gene and genetic studies of tumor microenvironment

Edited by
Yi Yao, Minglun Li, Jian Li, Zhongliang Zheng and Shun Lu

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-3690-2
DOI 10.3389/978-2-8325-3690-2

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

Gene and genetic studies of tumor microenvironment

Topic editors

Yi Yao — Renmin Hospital of Wuhan University, China
Minglun Li — LMU Munich University Hospital, Germany
Jian Li — University Hospital Bonn, Germany
Zhongliang Zheng — Wuhan University, China
Shun Lu — University of Electronic Science and Technology of China, China

Citation

Yao, Y., Li, M., Li, J., Zheng, Z., Lu, S., eds. (2023). *Gene and genetic studies of tumor microenvironment*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3690-2

Table of contents

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 September 2022
DOI 10.3389/fgene.2022.1008416

OPEN ACCESS

EDITED BY
Shun Lu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Qiang Yan,
Taiyuan University of Technology, China
Marc De Perrot,
University Health Network (UHN), Canada

*CORRESPONDENCE
Ling-Xia Zhao,
shan2013le@126.com
Sheng-Xiao Zhang,
shengxiao_zhang@163.com
Xiao-Fei Zhuang,
xiaofei_zhuang01@hotmail.com

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

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

RECEIVED 31 July 2022
ACCEPTED 31 August 2022
PUBLISHED 16 September 2022

CITATION
Li F, Song Q-Z, Zhang Y-F, Wang X-R, Cao L-M, Li N, Zhao L-X, Zhang S-X and Zhuang X-F (2022), Identifying the EMT-related signature to stratify prognosis and evaluate the tumor microenvironment in lung adenocarcinoma.
Front. Genet. 13:1008416.
doi: 10.3389/fgene.2022.1008416

Identifying the EMT-related signature to stratify prognosis and evaluate the tumor microenvironment in lung adenocarcinoma

Feng Li[1†], Qing-Zhen Song[2†], Yi-Fan Zhang[3,4†], Xing-Ru Wang[5], Li-Min Cao[3], Nan Li[6], Ling-Xia Zhao[7*], Sheng-Xiao Zhang[4,8*] and Xiao-Fei Zhuang[9*]

[1]Department of Cell Biology, Shanxi Province Cancer Hospital, Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences, Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China, [2]Department of Special Geriatrics, Shanxi Province Cancer Hospital, Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences, Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China, [3]The First Clinical Medical College, Shanxi Medical University, Taiyuan, China, [4]Key Laboratory of Cellular Physiology at Shanxi Medical University, Ministry of Education, Taiyuan, China, [5]The Second Clinical Medical College, Shanxi Medical University, Taiyuan, China, [6]The School of Basic Medicine of Shanxi Medical University, Taiyuan, Shanxi, China, [7]Department of Endocrinology and Metabolism, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, China, [8]Department of Rheumatology, The Second Hospital of Shanxi Medical University, Taiyuan, Shanxi, China, [9]Department of Thoracic Surgery, Shanxi Province Cancer Hospital, Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences, Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China

Background: Epithelial-mesenchymal transition (EMT) is a critical process in tumor invasion and metastasis. EMT has been shown to significantly influence the invasion, metastasis, and poor prognosis in lung adenocarcinoma (LUAD). This study aimed to develop a novel EMT-related prognostic model capable of predicting overall survival (OS) in patients with LUAD.

Methods: A total of *283* LUAD patients from TCGA RNA-seq dataset were assigned to a training cohort for model building, and 310 LUAD patients from GEO RNA-seq dataset were assigned to a validation cohort. EMT genes were acquired from MsigDB database and then prognosis-related EMT genes were identified by univariate Cox regression. Lasso regression was then performed to determine the genes and the corresponding variables to construct a prognosis risk model from the training cohort. Furthermore, characteristics of the tumor microenvironment (TME), mutation status and chemotherapy responses were analyzed to assess the differences between the two risk groups based on the prognostic model. In addition, RT-qPCR was employed to validate the expression patterns of the 6 genes derived from the risk model.

Results: A six-gene EMT signature (PMEPA1, LOXL2, PLOD2, MMP14, SPOCK1 and DCN) was successfully constructed and validated. The signature assigned the LUAD patients into high-risk and low-risk groups. In comparison with the low-risk group, patients in the high-risk group had a significantly lower survival rate. ROC curves and calibration curves for the risk

model demonstrated reliable stratification and predictive ability. The risk model was robustly correlated with multiple TME characteristics. Besides, the data showed that patients in the low-risk group had more immune activities, higher stemness scores and cytolytic activity scores and higher TMB. In addition, RT-qPCR results revealed that PMEPA1, LOXL2, PLOD2, MMP14, and SPOCK1 were notably upregulated in LUAD tissues, while DCN was downregulated.

Conclusion: Our study successfully developed a novel EMT-related signature to predict prognosis of LUAD patients and guide treatment strategies. The six genes derived from the prediction signature might play a potential role in antitumor immunity and serve as promising therapeutic targets in LUAD.

KEYWORDS

epithelial-mesenchymal transition, lung adenocarcinoma, prognosis, tumor immune microenvironment, immunotherapy, chemotherapy

Introduction

Lung cancer is the leading cause of death from cancer worldwide, which contributed to approximately 1,800,000 deaths in 2020 (Sung et al., 2021). The majority of the lung cancers are lung adenocarcinoma (LUAD), which are highly invasive, with a rapid metastatic spread, highlighting the systemic threat of the disease (Devarakonda et al., 2015).

In the past two decades, there have been important advances in characterization of mutational spectrum and molecular subtypes of LUAD, which have led to development of targeted therapies resulting in dramatically improved patient outcomes (Rotow and Bivona, 2017). However, clinical application in targeting RAS signaling or rescuing the tumor suppressor TP53 gene, which have been recommended as LUAD therapies, remains challenging (Kim et al., 2021). Besides, treatment regimens that target epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase have only benefitted a small percentage of LUAD patients (Wang Q. et al., 2020). Hence, there is an urgent need for identification of prognostic biomarkers as well as effective drug targets.

Spread of cancer cells due to metastasis is the leading cause of death in patients with primary lung cancer (Prateep et al., 2018). Epithelial-mesenchymal transition (EMT) is an important mechanism driving the tumor metastasis process, where epithelial cells lose their morphology and subsequently change to mesenchymal phenotypes, thereby acquiring features of mesenchymal cells such as motility and invasiveness (Iwatsuki et al., 2010). Loss of E-cadherin is a hallmark of EMT, which leads to decreased intercellular adhesion and enhanced cell motility (Thiery et al., 2009). Previous data have suggested that EMT is associated with neoplastic aggressiveness and progression in various malignancies including LUAD.

Tumor microenvironment (TME) is comprised of stromal and immune cells that secrete a variety of cytokines, chemokines and growth factors, which have been shown to induce EMT in nearby cancer cells through direct activation of various EMT-induced transcription factors (EMT-TFs) or inhibition of expression of effector molecules that promote mesenchymal cell state (Gupta and Massagué, 2006; Grivennikov et al., 2010). Besides, EMT has also been reported to play a vital role in tumor malignancy, immune regulation and initiation of therapeutic responses by inducing cell phenotypic plasticity, inflammatory, and immunosuppressive TME, leading to resistance to immunotherapy and chemotherapy (Terry et al., 2017). In addition, EMT status was associated with the activation of varied immune checkpoint molecules (Datar and Schalper, 2016). Therefore, it is important to understand the underlying mechanisms mediating the interaction between EMT and TME. The development of an EMT-related signature may contribute to the provision of potential biomarkers for LUAD and help enhance the understanding of the immunogenomic profile of LUAD.

In this study, we developed an EMT-related signature (ERGS) related to prognosis based on The Cancer Genome Atlas (TCGA) database which was validated by the Gene Expression Omnibus (GEO) database. The results demonstrated that LUAD patients with high-risk scores were strongly associated with shorter overall survival (OS) compared with patients in the high low-risk score group. We then explored the difference in immune infiltration and mutation landscape between the two risk groups and analyzed patients' response to the immune checkpoint inhibitor (ICI) therapy and chemotherapy. Together, the high-risk group was more likely to experience immunosuppression, thus less likely to benefit from either of the treatment options, which is consistent with the EMT features. In a nutshell, our study highlights a functional role of the ERGs and uncovers a potential prognostic biomarker for individualized treatment of LUAD.

Methods

Collection of the clinical samples

A total of 10 pairs of LUAD tissues and adjacent non-tumor tissues were collected from patients who received surgical

resection at Shanxi Cancer Hospital (Shanxi, China) from January to September 2021. The patients were not subjected to any anti-cancer treatment before surgery. Tissue specimens were frozen in liquid nitrogen within 30 min of resection and stored at −80°C for analysis. Our study was approved by the ethics committee of Shanxi Cancer Hospital (sxszl-F-375), and was conducted in accordance with the principles of Declaration of Helsinki.

Dataset acquisition and processing

The RNA-sequencing data and corresponding clinical information of 341 samples in the TCGA-LUAD cohort were obtained from the UCSC Cancer Genomics Browser database (https://genome-cancer.ucsc.edu/) and used as the training cohort, which included 283 tumor samples and 58 tumor-adjacent tissue samples. Another 310 LUAD samples in the GSE72094 from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) were used as an external validation cohort. Gene expression profiles were all quantified with fragments per kilobase of transcript per million mapped reads (FPKM) and normalized using log2-based transformations. Because the data from TCGA and GEO datasets are publicly available, there was no requirement for institutional review board approval and informed consent from the patients. EMT-related genes (ERGs) from the gene set "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION" were downloaded from GSEA (https://www.gsea-msigdb.org/gsea/index.jsp) as shown in Supplementary Table S1.

Identification of differentially expressed ERGs

The "limma" package in R software was used to identify differentially expressed genes (DEGs) between LUAD and tumor-adjacent tissue samples, where $|\log2FC| > 0.32$ and $FDR < 0.05$ were set as filters. The Venn diagram was generated using the "Venn Diagram" R package to identify DEGs related to the EMT process. The DE-ERGs were visualized as volcano plots and heat maps by "pheatmap" and "ggplot2" R packages, respectively.

Construction and establishment of the ERG prognostic signature

Univariate Cox regression analysis was employed to determine the DE-ERGs associated with OS of the patients. The DE-ERGs most related to OS with a $p < 0.05$ were further picked out for least absolute shrinkage and selector operator (LASSO) Cox regression. The analysis narrowed down the candidate DE-ERGs which were used to construct a prognostic model. To find optimal penalization terms, a penalty regularization parameter (λ) was determined by ten-fold cross validation following the minimum criteria (i.e., the value of λ corresponding to the lowest partial likelihood deviance). The risk score of each patient was calculated based on the normalized expression level of each gene in the prognostic signature and its relevant regression coefficients. The formula of the model was: Risk score = ∑(expression of signature genes × corresponding coefficient). Based on the median of the risk score, LUAD patients were assigned into high- or low-risk groups.

Evaluation and validation of the ERG prognostic signature

To determine the potential value of the ERG prognostic signature in predicting the prognosis in LUAD patients, Kaplan-Meier survival curves were utilized to assess the prognostic value in the high- and low-risk groups using the "survminer" R package. The predictive ability of the ERG prognostic signature was assessed using the AUC values of the time-dependent receiver operating characteristic (ROC) curves generated by the "pROC" package. To explore the distribution of different risk groups, we performed t-SNE for cluster visualization using the "Rtsne" R package. In addition, univariate and multivariate Cox analyses were conducted to determine whether the risk score was independent from other clinicopathological characteristics (age, gender, T stage, N status and smoking status). Thereafter, the ERG prognostic signature was validated in an independent cohort obtained from the GEO database.

Construction of the nomogram

Based on all independent prognostic factors determined by the multivariate analysis, we established a nomogram to predict 1-, 3-, and 5-years OS of LUAD patients using the "rms" R package. The concordance index (C-index) was calculated to appraise the discriminative ability of the nomogram while the calibration curve was performed to evaluate the accuracy of the nomogram.

Mutation analysis

The "TCGAbiolinks" R package was used to download the somatic mutation profiles based on the segment mean value log2 (copy-number/2) of LUAD patients from the TCGA cohort, while the "maftools" package was employed for analysis and visualization of the data. We then used Oncoplot to display the top 20 genes with high mutation frequency in LUAD patient

samples in high- and low-risk groups. Besides, the tumor mutation burden (TMB) was calculated as the total number of somatic, coding, indel mutations and base substitution for each mega-base of the genome under analysis. The correlation between TMB levels and risk score was analyzed using the Spearman's correlation. In addition, we assessed the effect of the risk score combined with the TMB on the survival of LUAD patients.

Analysis of tumor immune microenvironment

To identify the immune infiltration characteristics in LUAD, the 22 immune cells from each LUAD sample from the TCGA cohort was quantified based on standardized gene expression profile using the CIBERSORT algorithm. The "ESTIMATE" R package was employed to compute the stromal score, immune score, and estimate score. As a measure of inflammation, the cytolytic activity (CYT) score was computed as the geometric mean of the RPKM expression of granzyme A (GZMA) and perforin-1 (PRF1) mRNA expression levels in the tumor tissues (Wakiyama et al., 2018). We also computed mRNA stemness index (mRNAsi) using the "TCGAbiolinks" R package based on the mRNA levels obtained from one-class logistic regression machine learning (OCLR) algorithm (Guo et al., 2021), where a higher mRNAsi represents a greater tumor dedifferentiation and higher cancer stem cell levels.

Assessment of immunotherapy response

Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was employed to evaluate the likelihood of each sample to respond to immunotherapy. TIDE algorithm is a computational method used to model two primary mechanisms of tumor immune evasion: the induction of T cell dysfunction in tumors with high infiltration of cytotoxic T lymphocytes (CTL) and prevention of T cell infiltration in tumors with low CTL level. Immune checkpoint expression has a significant impact on the immunotherapy treatment responses. To further investigate the influence of the ERG scores on immunotherapy, the differential expression of immune checkpoint-related genes between the two ERG subgroups were analyzed.

Estimation of chemotherapy response

We also predicted the chemotherapy response of each LUAD patient based on information obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database. Four common chemotherapeutic agents (cisplatin, paclitaxel, gemcitabine, and docetaxel) and two small molecule inhibitors targeting EGFR (erlotinib and gefitinib) were selected and used at default parameters, which are approved in the treatment of LUAD cases. The prediction procedure was conducted using the "pRRophetic" R package where sensitivity to the drug was quantified by half-maximal inhibitory concentration (IC50) predicted through ridge regression. A low IC50 indicates that the patients are more sensitive to the drug.

TABLE 1 Primers used in RT-qPCR.

Gene	Primer sequences (5′-3′)
PMEPA1	FORWARD CGTAGGTGAAAAGGCAGAACA
	REVERSE GACACAGCTCAACAAAGAAACGT
LOXL2	FORWARD ACAGAATGTGAAGGAGACATCC
	REVERSE TGATGTTGTTGGAGTAATCGGA
PLOD2	FORWARD GGATGCAGATGTTGTTTTGACA
	REVERSE GCTTTCCATGACGAGTTACAAG
MMP14	FORWARD CAAGATTGATGCTGCTCTCTTC
	REVERSE ACTTTGATGTTCTTGGGGTACT
SPOCK1	FORWARD CAGAAACTGGAATCCCAACAAG
	REVERSE TTGCACTTGACCAAATTCGAAG
DCN	FORWARD GACAACAACAAGCTTACCAGAG
	REVERSE TGAAAAGACTCACACCCGAATA
GAPDH	FORWARD TGACTTCAACAGCGACACCCA
	REVERSE CACCCTGTTGCTGTAGCCAAA

Gene set enrichment analysis

For further exploration of differences in biological pathways between high-risk and low-risk groups, GSEA was performed to assess Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) using "clusterProfiler" R package.

Reverse transcription-quantitative PCR

RT-qPCR was employed to quantify the expression of genes in ERGS in clinical specimens. We extracted total RNA from the LUAD and adjacent non-tumor tissues using the TRIzol reagent (Invitrogen, CA, United States). We synthesized cDNA from the total mRNA using PrimeScriptTM RT Master Mix (RR036B, Takara). Quantitative PCR was performed to analyze the mRNA expression levels of the ERGs genes using GoTaq® qPCR Master Mix (Promega, A6001). The RT-qPCR was performed in ABI Vii7 Sequence detection system (ABI, United States). We then compared the mRNA expression levels of PMEPA1, LOXL2, PLOD2, MMP14, SPOCK1, and DCN using the 2-ΔΔCT method. The primer sequences are shown in Table 1.

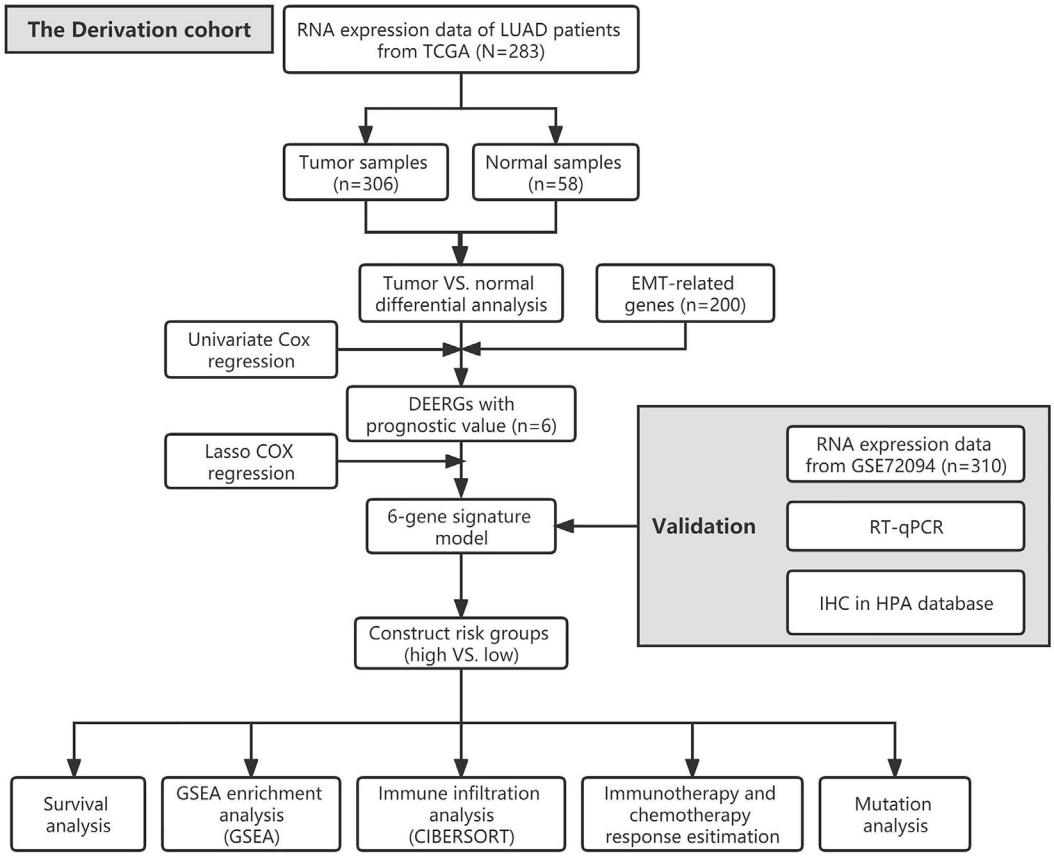


FIGURE 1
The flow chart of our study.

Immunohistochemical analysis

To further validate the expression f the signature genes, we analyzed immunohistochemistry (IHC) staining data of ERG proteins in lung cancer and normal lung tissues from the Human Protein Atlas (HPA) online database (https://www.proteinatlas.org/).

Statistical analysis

All statistical analyses and visualization were performed using R version 4.1.3. Differences between two groups were compared via the Wilcoxon rank-sum test or Kruskal-Wallis test in cases where the data did not follow normal distribution and the variance was unknown. The survival difference was evaluated using log-rank tests. In addition, correlation analyses between two continuous variables were evaluated by Spearman rank correlation test while K-nearest neighbor (k-NN) imputation was performed to impute the missing AUC values. A P value of less than 0.05 (two-sided) was considered statistically significant.

Results

Patient characteristics

The flow chart of our study is shown in Figure 1. A total of 341 LUAD patients from the TCGA cohort were defined as a training set, while 310 patients from the GSE72094 cohort were used for external validation. The detailed clinical information of these patients is as summarized in Supplementary Table S2.

Identification of differentially expressed ERGs

Through the differential gene screening analysis, we retrieved 149 differentially expressed ERGs, which included 83 downregulated and 66 upregulated genes as shown in Figure 2A. The expression of the ERGs in LUAD samples and normal samples was displayed in a heat map (Figure 2B).

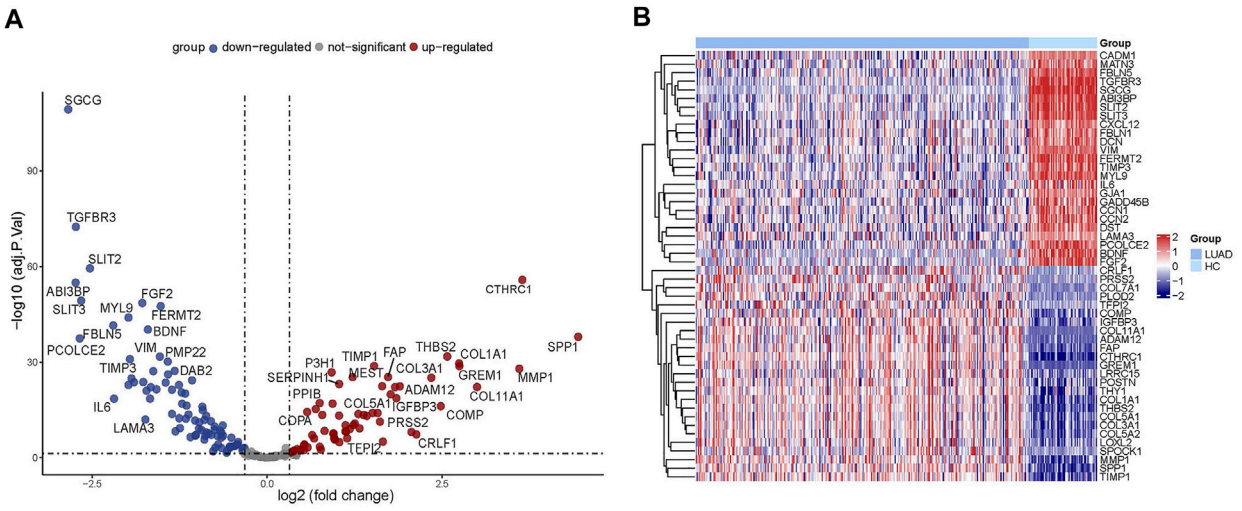


FIGURE 2
Analysis of differentially expressed ERGs. **(A)** Volcano plot showing the downregulated and upregulated ERGs. **(B)** Heatmap of the differentially expressed ERGs in LUAD.

Construction and assessment of ERGs

To establish a prognostic model, Cox regression and LASSO regression were performed in the training set. First, we employed univariate Cox proportional hazard regression to identify prognosis-related genes from 149 DE-ERGs. Using a p value <0.05, 6 prognosis-related genes were identified (Figure 3A) (Supplementary Table S3). The 6 prognosis-related genes were further included in LASSO Cox regression based on the optimal value of λ to eliminate overfit genes and narrow down the range of model genes (Figures 3B,C) (Supplementary Table S4). After the LASSO regression analysis, 6 ERGs were used to construct the prognostic model consisting of PMEPA1, LOXL2, PLOD2, MMP14, SPOCK1 and DCN genes (Figure 3D). According to the coefficients and standardized expression of the six genes, the risk score of each LUAD patient from the TCGA dataset was calculated as follows: Risk Score = (−0.282*DCN) + (0.105*LOXL2) + (0.041*MMP14) + (0.071*PLOD2) + (0.149*PMEPA1) + (0.03*SPOCK1) (Supplementary Table S5). Taking the median risk score of 0.104 as cut-off, the patients were divided into high- (N = 142) and low-risk groups (N = 141). The t-SNE analysis revealed that the patients in different risk groups were distributed in two directions (Figure 4A). We defined the risk scores rank distribution, survival status and expression patterns of the 6 ERGs in LUAD patients (Figure 4B). Besides, the Kaplan-Meier survival analysis demonstrated that patients in the high-risk group had significantly poorer OS compared to those in the low-risk group (p < 0.001) (Figure 4C). In addition, as depicted in Figure 4D, the AUC value of the ROC curves for 1-year, 3-year, and 5-year OS was 0.685, 0.705 and 0.620, respectively, in the TCGA cohort.

Validation of the prognostic signature

Using the same formula and the cut-off value described above, patients in the GSE72094 cohort (N = 310) were stratified into the high-risk group (N = 118) and low-risk group (N = 192). Likewise, t-SNE analysis in the GSE72094 cohort confirmed that the patients in different risk groups were distributed into two directions (Figure 5A). The risk scores rank distribution, survival status and expression patterns of the 6 ERGs in the LUAD patients were shown in Figure 5B. In line with the results from the TCGA dataset, patients in the high-risk group showed significantly worse OS as opposed to those in the low-risk group (p = 0.00048; Figure 5C). ROC analysis demonstrated that the ERGs exhibited precise predictive capacity. AUCs at 1-, 3- and 5-year OS was 0.621, 0.670, and 0.878, respectively (Figure 5D). Together, these results indicated that the established prognostic model was capable of universal application.

Construction and evaluation of the nomogram

To further evaluate the effect of the ERGs in predicting prognosis, we employed univariate and multivariate Cox regression analyses. Results from the univariate Cox regression analysis showed that T, N, stage and risk score were all significantly associated with OS, while the multivariate Cox regression analysis indicated that N stage and risk score were correlated with OS in patients with LUAD (HR:1.681, 95% CI (1.226–2.307), p < 0.001; HR: 1.302, 95% CI (0.936–1.811), p <

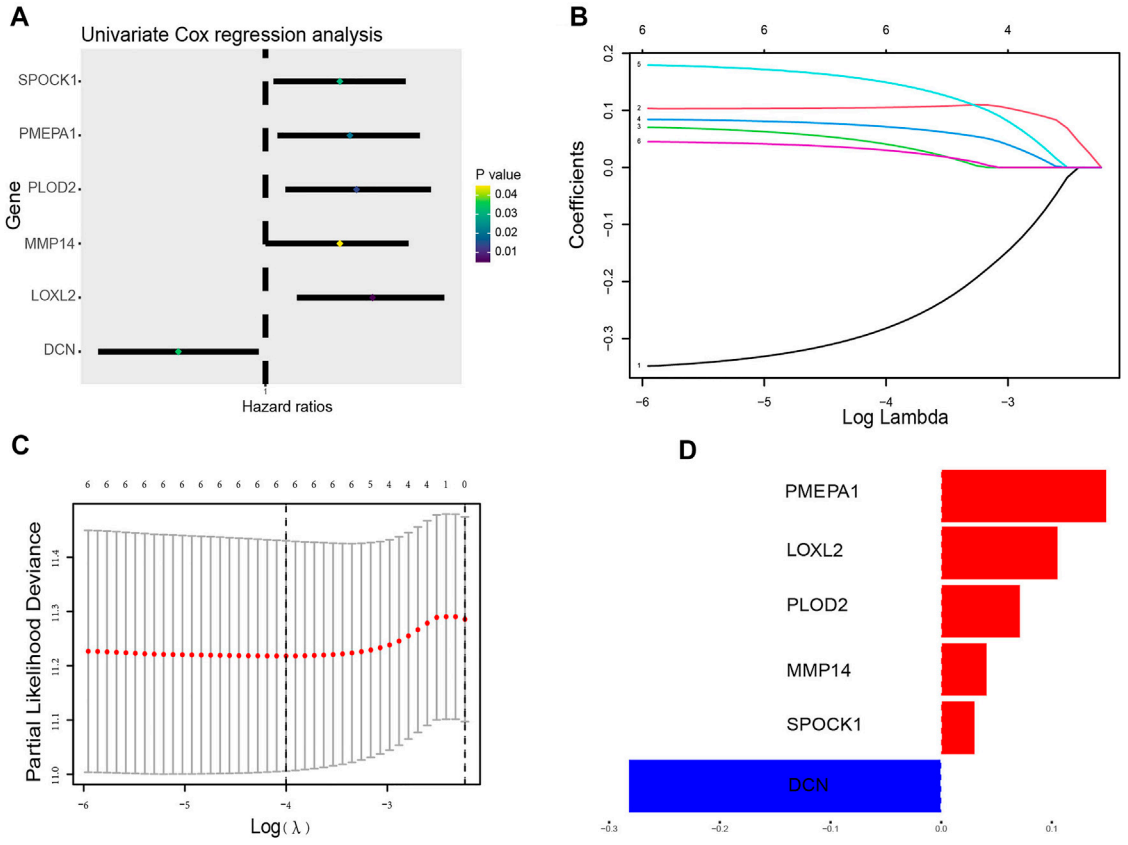


FIGURE 3
Construction of the ERG prognostic signature. **(A)** The forest plots illustrate univariate Cox analysis of the six genes significantly associated with OS. **(B)** Ten-time cross-validation for tuning parameter selection in the LASSO model. **(C)** LASSO coefficient profiles of the six ERGs significantly associated with OS. **(D)** The expression level of the 6 genes identified by Lasso regression analysis.

0.001) (Figure 6A). Moreover, the regression analyses in the GSE72094 indicated that the pathological stage and risk score were independent prognostic factors for OS (Figure 6B). These data demonstrated that the signature was an independent risk factor for survival in patients with LUAD.

We then developed a nomogram premised on the results from the multivariate Cox regression including the N stage and risk score, to predict 1-year, 3-year and 5-year OS, which contributed to defining higher risk scores (range 0–100 points) for the worse OS (Figure 6C). Each variable was allocated a score on the point scale. After summing up the points, the estimation of the survival likelihood was achieved by drawing a vertical line between the total points axis and the survival probability axes. Consistent with our previous findings, the nomogram illustrated the risk score as the prevailing contribution to prognosis compared with conventional clinical characteristics. In addition, the C-index was 0.715,896 and the calibration curves demonstrated that the predicted survival probability were highly consistent with the actual one (Keynesian cross) for 1-, 3-, and 5-year OS, which showed that this nomogram had great prediction performance (Figure 6D).

Gene set enrichment analyses

To evaluate the potential mechanisms of the ERGs in the high and low risk groups, we performed GSEA to identify GO terms and KEGG pathways in the TCGA cohort ($p < 0.05$). The GO analysis showed that the most concentrated biological processes in the high-risk group included cell aggregation, detection of chemical stimulus involved in sensory perception of bitter, detection of chemical stimulus involved in sensory perception of taste, immature T cell proliferation, and tertiary alcohol metabolic process (Figure 7A), while the most enriched biological processes in the low-risk group were cytoplasmic translation, formation of cytoplasmic translation initiation complex, negative regulation of chromatin silencing, ribosomal small subunit assembly and SRP-dependent co-translational protein targeting to membrane (Figure 7B). On the other hand, the KEGG analysis showed that the genes in the high-risk group were mainly enriched in pathways such as chemical carcinogenesis-DNA adducts, drug metabolism-cytochrome P450, glutathione metabolism, metabolism of

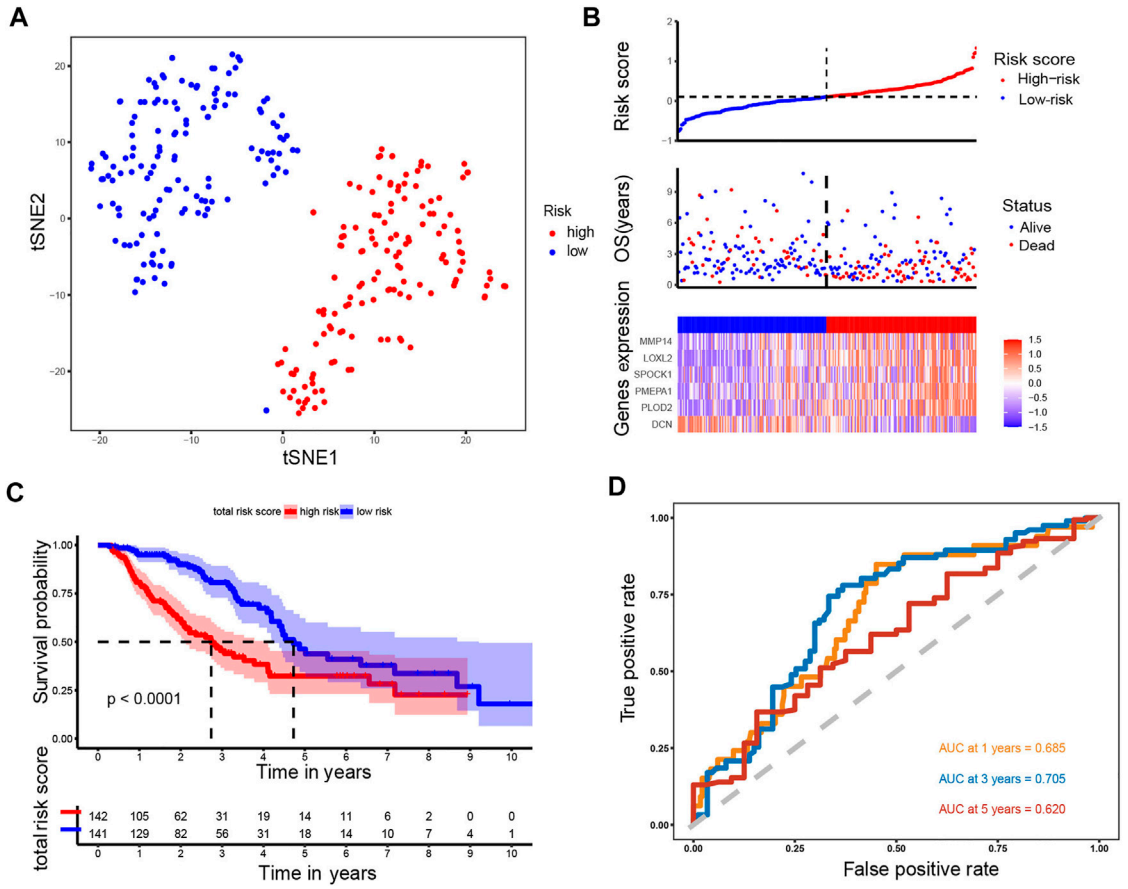


FIGURE 4
Analysis of the prognostic model in the training set. **(A)** t-SNE was used to evaluate whether the samples could be grouped correctly based on the ERGs risk score. **(B)** Heatmap showing the expression of six crucial genes in high- and low-risk groups and the distribution of risk scores and survival status of the LUAD patients with increasing risk score. **(C)** KM survival analysis between the high- and low-risk groups. **(D)** ROC curves analysis of the ERGs on OS at 1 year, 3 years, and 5 years.

xenobiotics by cytochrome P450 as well as protein digestion and absorption (Figure 7C). The genes in the low-risk group were mainly enriched in aminoacyl-tRNA biosynthesis, citrate cycle (TCA cycle), proximal tubule bicarbonate reclamation, ribosome biogenesis in eukaryotes and RNA degradation (Figure 7D). Moreover, most of the GO terms and KEGG pathways enriched in our analysis were closely associated with the occurrence and development of LUAD, which indicated that the ERGs may play a key role in cancer development and revealed potential pathways that could serve as therapeutic targets in LUAD.

Analysis of the mutational landscape based on ERGS

In further characterize the high- and low-risk groups at the genomic level, the mutation status of both groups was analyzed. Results showed that missense mutations were the most prevalent of all mutation types followed by nonsense and frameshift deletions in both groups. The main variant was single nucleotide polymorphism (SNP), in which the single-nucleotide variant (SNV) with T>G was the most frequent (Figure 8A,B). The top 20 most common mutated genes in the high- and low-risk groups ranked based on percentages are shown in Figures 8C,D. In the low-risk group, 86.93% of the samples carried mutations. The top 10 mutated genes were TTN, CSMD3, MUC16, LRP18, RYR2, USH2A, TP53, FLG, ZFHX4, and ZNF536. The mutation frequency was higher in the high-risk group (94.74%) compared with the low-risk group. The top 10 factors linked to mutations were TTN, MUC16, RYR2, CSMD3, TP53, USH2A, ZFHX4, LRP1B, KRAS, and FLG. Subsequently, the relationship between the TMB and the risk score was examined. It was found that the risk score was positively correlated with TMB ($R = 0.14$, $p < 0.05$, Figure 8E), suggesting that the risk score could be an

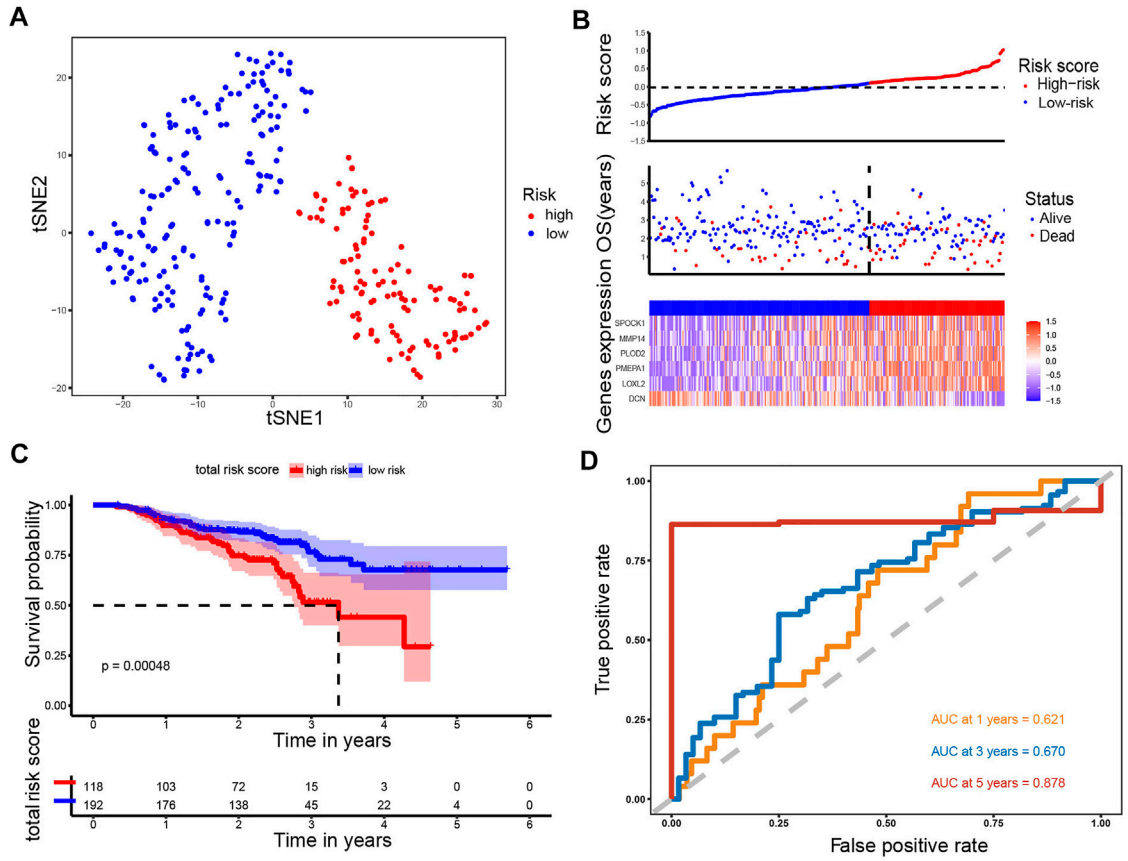


FIGURE 5
Analysis of the prognostic model in the test set. **(A)** t-SNE was used to evaluate whether the samples could be grouped correctly based on the ERGs risk score. **(B)** Heatmap showing the expression of six crucial genes in high- and low-risk groups and the distribution of risk scores and survival status of the LUAD patients with increasing risk score. **(C)** KM survival analysis between the high- and low-risk groups. **(D)** ROC curves analysis of the ERGs on OS at 1 year, 3 years, and 5 years.

accurate indicator of the characteristics and performance of TMB in tumors. Further analysis revealed that the TMB status did not affect the survival outcome predicted using the risk score. The KM curve showed that the survival outcome was different between the subgroups (high-risk and high TMB vs. high-risk and low TMB, low-risk and high TMB vs. low-risk and low TMB, $p = 0.018$; Figure 8F). The low-risk and low TMB subgroup showed the highest overall survival rate, whereas the high TMB and high-risk group had the lowest survival rate.

Evaluation of immune infiltration status

The CIBERSORT tool was utilized to calculate the infiltration degree of 22 immune cells (Figure 9A). Results showed that the resting memory CD4 T cells, gamma delta T cells, monocytes, resting mast cells, and resting dendritic cells were higher in low-risk group compared with the high-risk group. On the contrary, the high-risk group had high proportions of activated memory CD4 T cells, resting NK cells, M0 macrophages, activated mast cells and neutrophils compared with the low-risk group. Subsequently, we explored the relationship between immune cells and the risk score. It was found that the risk score was positively associated with the level of activated memory CD4 T cells, M0 macrophages, activated dendritic cells, activated mast cells, and resting NK cells ($p < 0.05$, Supplementary Figure S1A–E), however, it was negatively correlated with the level of T cells CD4 memory resting, macrophages M2, eosinophils, mast cells resting, and dendritic cells resting ($p < 0.05$, Supplementary Figure S1F–J). Having identified the effects of ERGs on the regulation of TME remodeling, we further investigated whether ERGs expression levels were correlated with the abundance of immune cells. It was identified that these six prognostic genes were differentially correlated with immune cell infiltration (Supplementary Figure S2). Furthermore, the estimate score, immune core, and stromal score were higher in

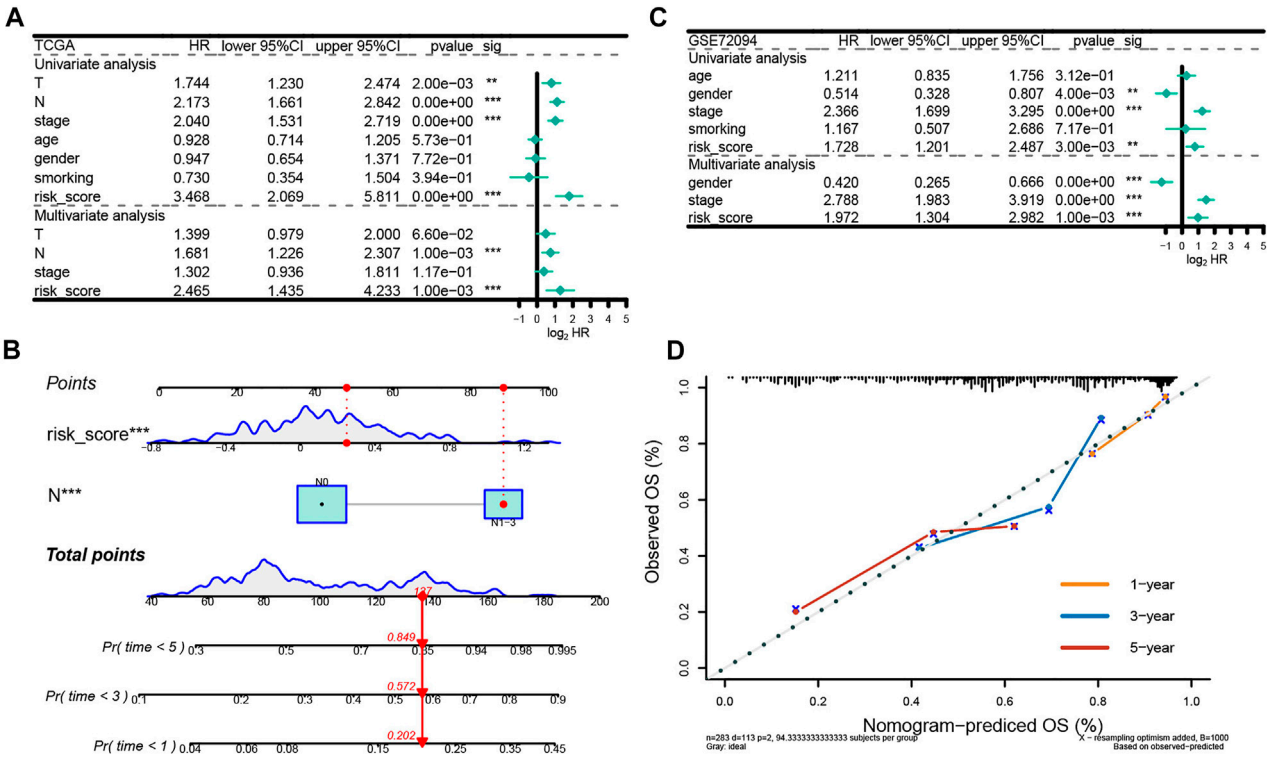


FIGURE 6
Independent prognostic analysis of the prognostic model. **(A)** Independent prognostic factors in the training set. **(B)** Independent prognostic factors in the test set. **(C)** The nomogram to predict overall survival was created based on independent prognostic factors. The 1-year, 3-year, and 5-year survival rate is predicted according to the total score. **(D)** The correction curve based on the prediction model.

the low-risk group compared with the high-risk group ($p < 0.001$, Figures 9B–D), suggesting that the infiltration levels of immune and stromal cells were higher in the low-risk group. In comparison, the high-risk group had low CYT score ($p < 0.05$) and high mRNAsi score ($p < 0.001$, Figures 9E,F). This demonstrated that patients in the high-risk group had lower antitumor immunity and higher neoplastic stemness. Consequently, tumor cells in these patients had stronger self-renewal, differentiation, and proliferation ability, which may explain their worse OS(Wang et al., 2021b).

Assessment of chemotherapy efficacy

In subsequent analyses, we further explored the association between the ERGS and efficacy of chemotherapy in LUAD. Results showed that patients in the low-risk group had significantly lower IC50 values and were more sensitive to paclitaxel ($p < 0.0001$), docetaxel ($p < 0.0001$), and gemcitabine ($p < 0.05$) compared with those in the low-risk group, which suggested that the constructed model could effectively predict efficacy and sensitivity to chemotherapy (Figure 10E).

Validation of gene expression

To validate the expression profile of ERGs in LUAD patients, clinical specimens were collected from LUAD patients, together with adjacent normal tissue. These specimens were analyzed using RT-qPCR. It was found that LOXL2, PLOD2, MMP14 and SPOCK1 were upregulated in tumor samples, whereas DCN was downregulated in tumor specimens (Figure 11).

Validation of protein expression

The HPA is a public database with millions of immunohistochemical images and is used by researchers to compare protein expression patterns between normal and tumor tissues. Because the lung cancer data were not classified according to histological type in HPA, we analyzed the IHC staining of six ERGs in lung cancer to verify their expression levels. Notably, only protein expression staining images of five genes (DCN, LOXL2, PLOD2, MMP14, and SPOCK1) were found in HPA. Moreover, the results showed that the expression levels of LOXL2, PLOD2, MMP14 and

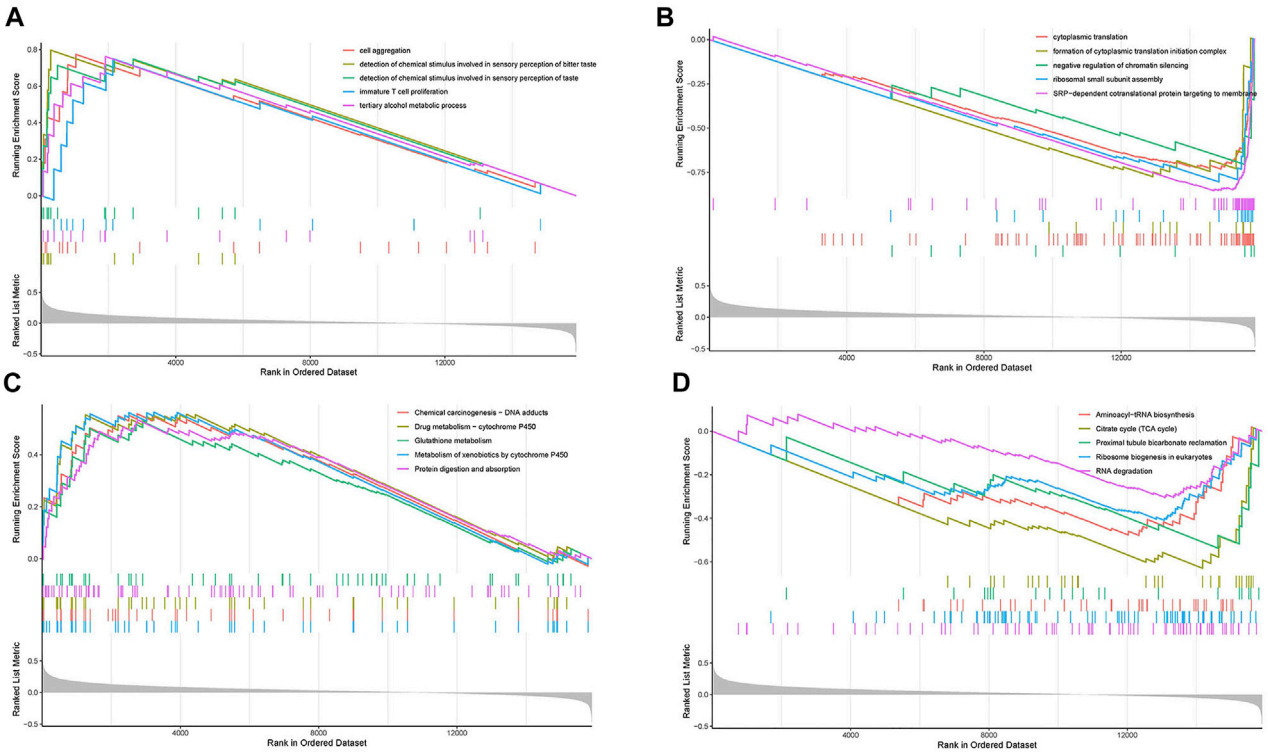


FIGURE 7
Gene set enrichment analysis (GSEA). **(A)** The top five enriched GO pathways in the high-risk group. **(B)** The top five enriched GO pathways in the low-risk group. **(C)** The top five enriched KEGG pathways in the high-risk group. **(D)** The top five enriched KEGG pathways in the low-risk group.

SPOCK1 was higher whereas that of DCN was lower in tumor tissues compared with normal tissues (Figure 12).

Discussion

The pathologic stage is a critical marker used for prediction of the prognosis of LUAD in routine clinical practice. However, the progression of LUAD is highly heterogenous in terms of genetic and epigenetic presentations (Lu et al., 2021). For this reason, patients with the same stage of disease may have different clinical outcomes (Yao et al., 2021). Accurate prognosis analysis is a critical factor of precision medicine in stratifying risks and developing an optimal management plan. A growing body of research has revealed that EMT process regulate several aspects of cancer cells including invasion, metastasis, refractory responses to chemotherapy and immunotherapy, immunosuppression, and acquisition of stem cell-like properties. Studies have also reported that the EMT process has been linked to metastasis and treatment resistance in LUAD (Wang et al., 2021a). Therefore, we constructed and validated a comprehensive signature based on EMT-related genes to predict the prognosis of LUAD patients using GEO and TCGA dataset.

The proposed signature consisted of six genes, DCN, PMEPA1, LOXL2, PLOD2, MMP14, and SPOCK1. Among them, PMEPA1, LOXL2, PLOD2, MMP14, and SPOCK1 were associated with poor outcomes whereas DCN was correlated with good prognosis of LUAD patients. PMEPA1 has been shown to induce tumorigenesis by interfering with several signaling cascades such as mutated p53, Hippo signaling, Wnt, and EGF (Qiu et al., 2021). It has also been reported to promote malignant behavior and enhance tumorigenic ability by activating MAPK/JNK signaling pathways (Tan et al., 2021). LOXL2 regulates collagen cross-linking and deposition in primary tumor tissue. In previous studies, it was found to promote tumor cell survival and development of drug resistance, regulate cell adhesion, motility, and invasion. Upregulation of LOXL2 enhanced the invasion and metastasis of lung cancer (Peng et al., 2017). Moreover, LOXL2 has been incorporated in prognostic models to predict late recurrence in LUAD patients (Zhao et al., 2021). PLOD2 is one of the members of the PLOD family that encodes the lysyl hydroxylase 2. It modulates collagen cross-link formation in the extracellular matrix. In previous studies, PLOD2 was reported to promote tumor metastasis by inducing collagen cross-linking (Yamauchi and Sricholpech, 2012; Chen et al., 2015). A bioinformatics study

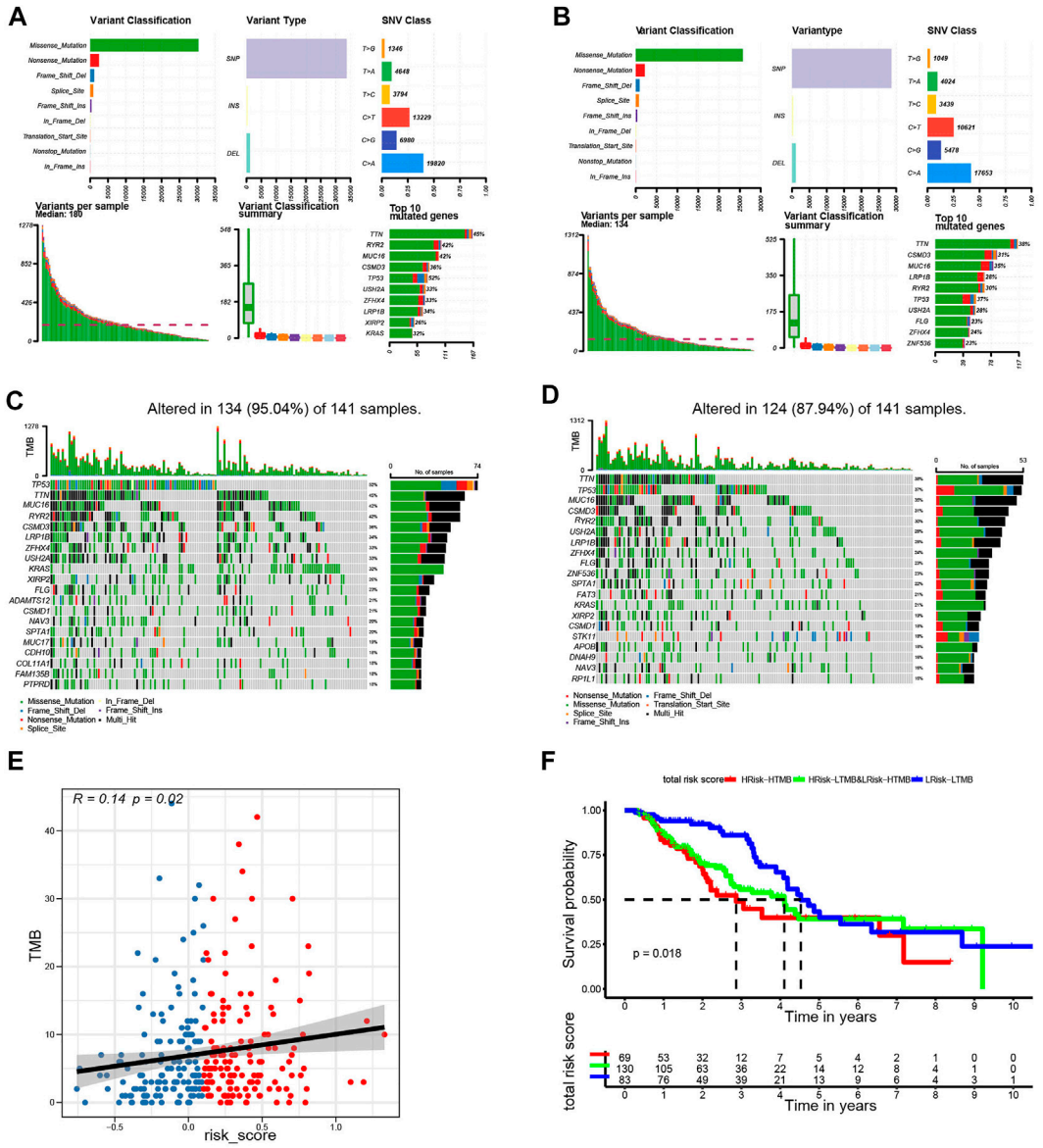


FIGURE 8
Landscape of mutation profiles in the low- and high-risk groups. **(A)** Overview of mutation types in the high-risk group. **(B)** Overview of mutation types in the low-risk group. **(C)** Waterfall Plot of the top 20 genes with the most mutations in the high-risk group. **(D)** Waterfall Plot of the top 20 genes with the most mutations in the low-risk group. **(E)** The correlation between risk score and TMB. **(F)** Kaplan-Meier curves for patients stratified by risk score combined with TMB.

showed that the PLOD family members can be novel biomarkers for predicting LUAD prognosis (Meng et al., 2021). Being a member of the first membrane matrix metalloproteinases (MMPS), MMP14 has been shown to promote extracellular matrix (ECM) degradation to accelerate tumor cell migration, inflammation, invasion, angiogenesis and metastasis. Its expression was found to be increased in colorectal cancer, lung cancer, and nasopharyngeal carcinoma, leading to enhanced tumor progression (Yan et al., 2015; Stawowczyk et al., 2017; Cui et al., 2019). SPOCK1 encodes a matricellular glycoprotein belonging to a new Ca (2+)-binding proteoglycan family, which promotes cell proliferation, adherence, and migration (Bradshaw and Sage, 2001). High SPOCK1 expression has been associated with increased invasiveness, growth, and metastatic potential (Wang et al., 2018). In lung cancer, high expression of SPOCK1 correlated with poor prognosis. SPOCK1 is a novel TGF-β-targeted gene that regulates lung cancer epithelial cells (Basu et al., 2018). As a

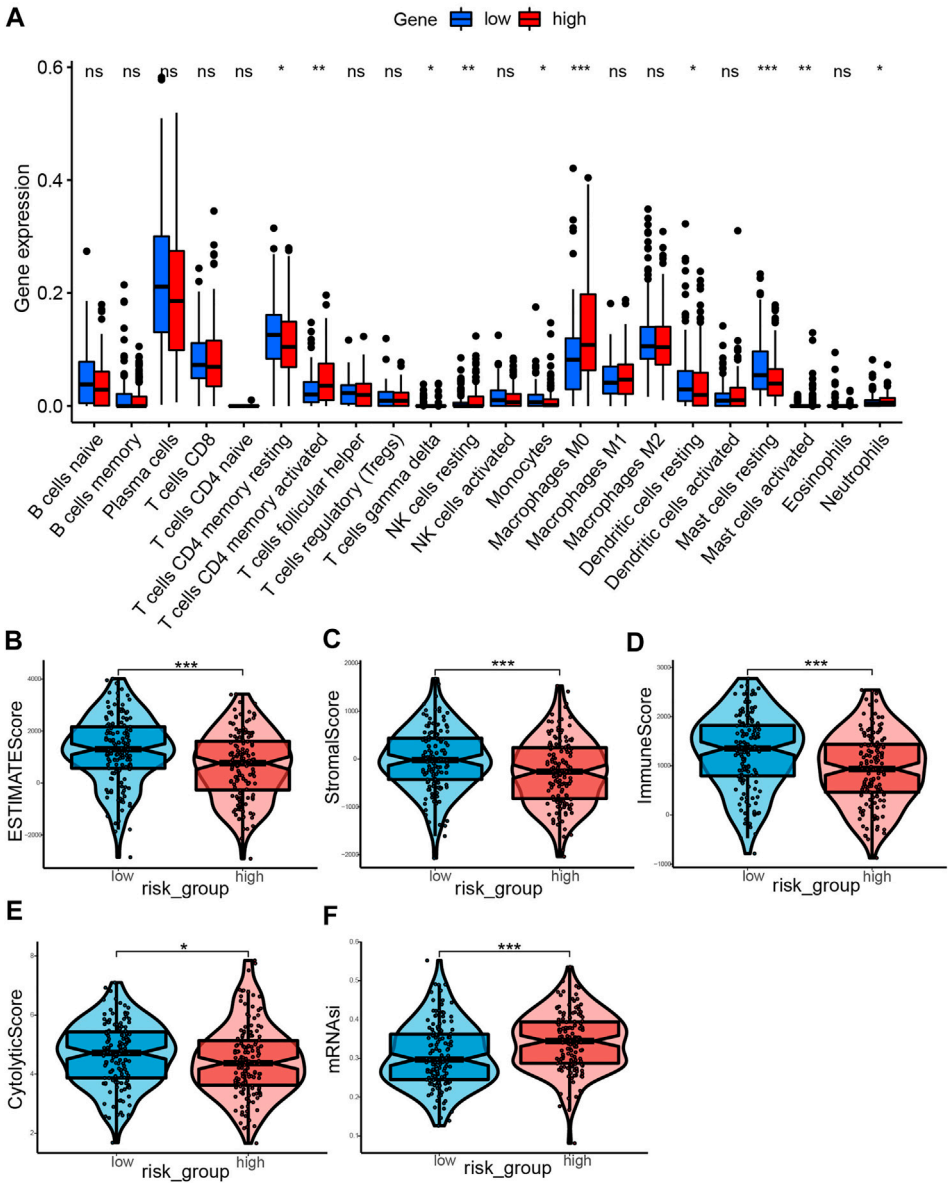


FIGURE 9
Landscape of immune Infiltration profiles in the low- and high-risk groups. **(A)** The ratio differentiation of 22 kinds of immune cells between the low- and high-risk group, and the Wilcoxon rank sum was used for the significance test. **(B–F)** The violin plot showed the differences of ESTIMATE score, stromal score, immune score, CYT score and mRNAsi score between the low- and high-risk groups. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

protective factor, DCN has been suggested to block receptor tyrosine kinases thereby suppress lung cancer progression (Horváth et al., 2014). This view is consistent with that of previous articles.

The immune components of the TME can promote or inhibit or tumor progression (Papait et al., 2020). It is important to understand the underlying mechanisms that are involved between EMT and the TME. In this study, the TME was assessed using CIBERSORT and the ESTIMATE algorithm. The relationship between the risk score and immune-infiltrating cells was evaluated. Results showed that tumors in the high-risk group showed higher infiltration of immunosuppressive cells such as macrophages, neutrophils, and mast cells compared with tumors in the low-risk group. This indicated that the EMT process may protect tumors from the intrinsic anti-tumor immune response by creating an immunosuppressive microenvironment. Therefore, the EMT process may explain the poor prognosis of the high-risk group. Previously, it was found that patients with high levels of M0 macrophages had enhanced EMT

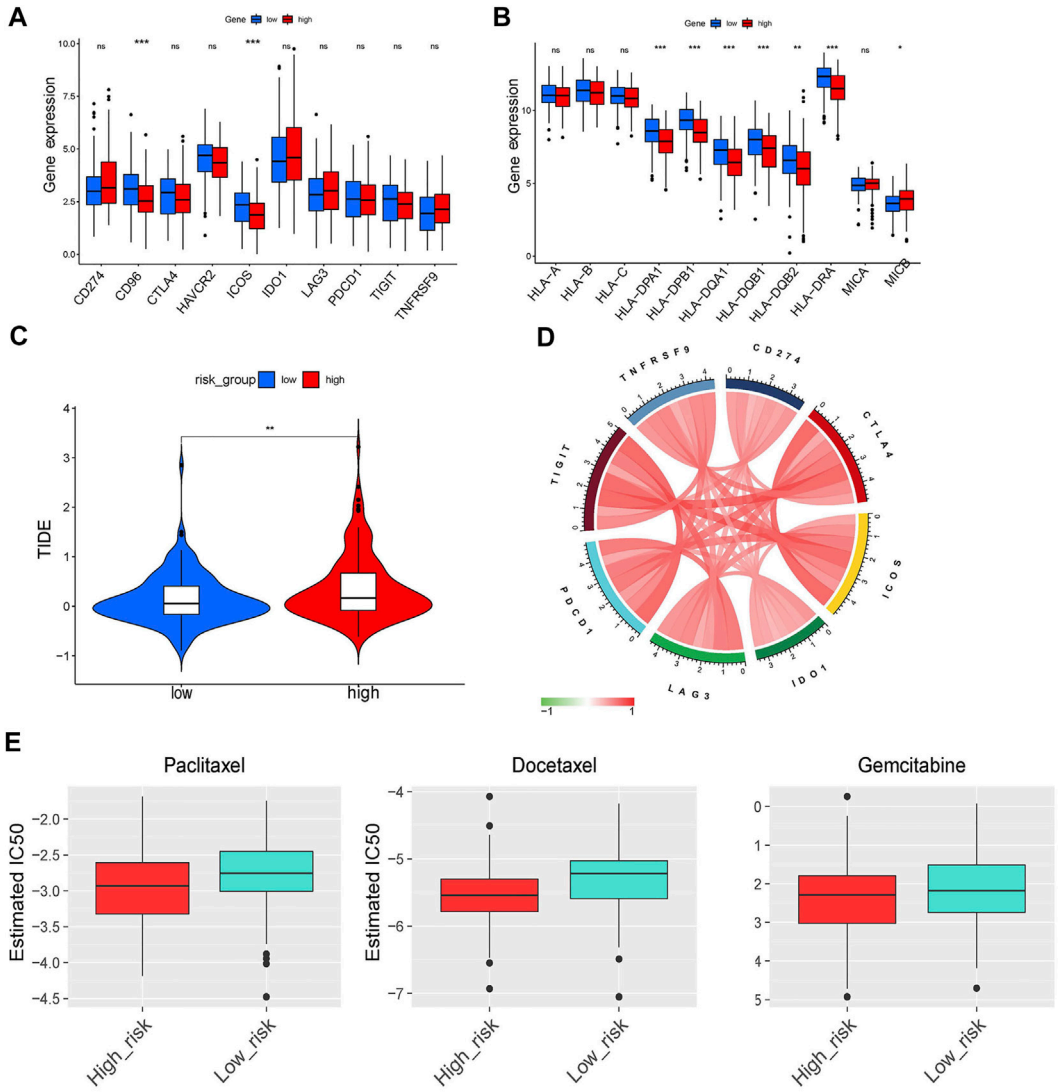


FIGURE 10
Immune checkpoints analysis and evaluation of response to ICI therapy and chemotherapy. **(A,B)** The differentiation of immune checkpoints between the low- and high-risk group. **(C)** The comparison of TIDE score between the low- and high-risk group. **(D)** Circos plot showing the interconnectivity among ERGS genes. The thickness and color of the ribbons depend on the correlation between the signature gene expression. **(E)** The sensitivity to paclitaxel, docetaxel and gemcitabine of patients with LUAD.

process, hence poor prognosis (Dong et al., 2021). Our findings corroborate the results that mast cell-derived extracellular vesicles induced EMT by signaling cascades (Yin et al., 2020). Activated memory $CD4^+$ T cells were also highly expressed in the high-risk group. We speculate that cytokines released by activated T cells, such as IL-6, TNF, and TGFβ, can promote EMT development (Cohen et al., 2015). It was reported that CD4+T cells can induce EMT-like features in clear cell renal carcinoma cells by secreting IL-6 (Chen et al., 2017).

In this study, the low-risk subgroup showed higher immunoactivity of immune checkpoint molecules, immune score, and cytolytic activity. The higher mRNAsi was found in the high-risk group, which may explain why EMT results in poor prognosis. The tumor immune escape decreases the expression of HLA which enables tumor cells to avoid the cytolysis function of T cells (Garcia-Lora et al., 2003). We found that expression of HLA genes was significantly downregulated in the high-risk group, which indicated that immune escape occurred in the high-risk group. Previous study also reported that EMT inhibits the formation of immune synapses between cancer cells and T cells leading the immune escape (Akalay et al., 2013). Based on the above results, we speculated that there might be significant

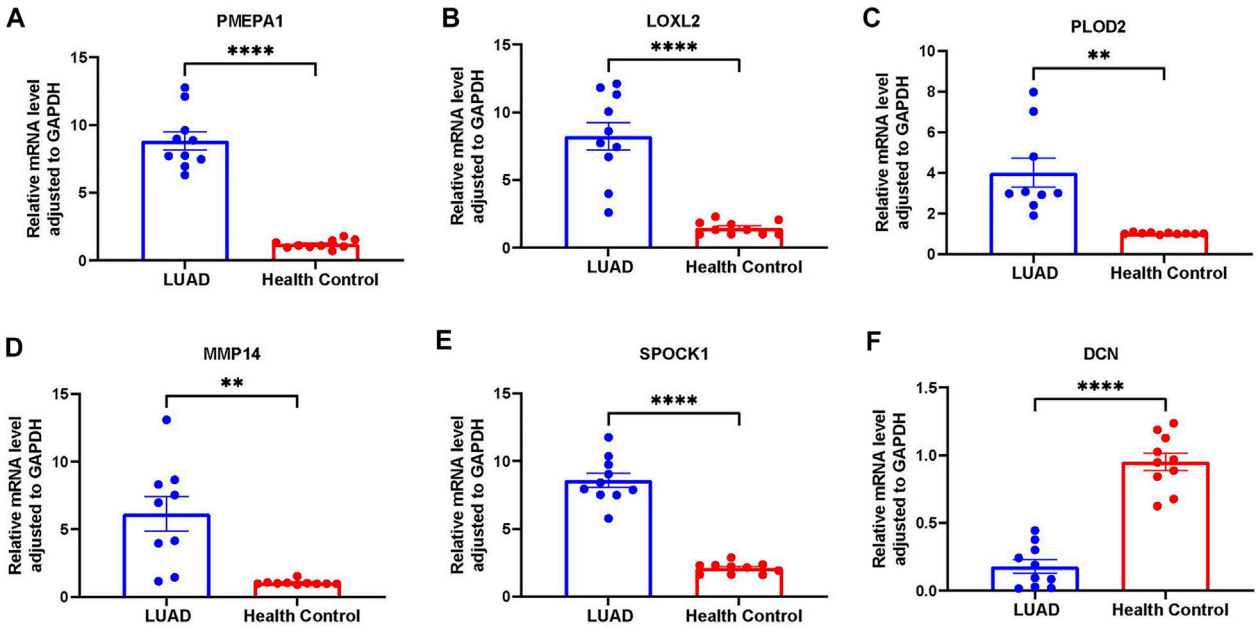


FIGURE 11
Further verification of the overexpression of six genes in RT-qPCR analysis. **(A)** PMEPA1 expression level in LUAD and health control tissues. **(B)** LOXL2 expression level in LUAD and health control tissues. **(C)** PLOD2 expression level in LUAD and health control tissues. **(D)** MMP14 expression level in LUAD and health control tissues. **(E)** SPOCK1 expression level in LUAD and health control tissues. **(F)** DCN expression level in LUAD and health control tissues.

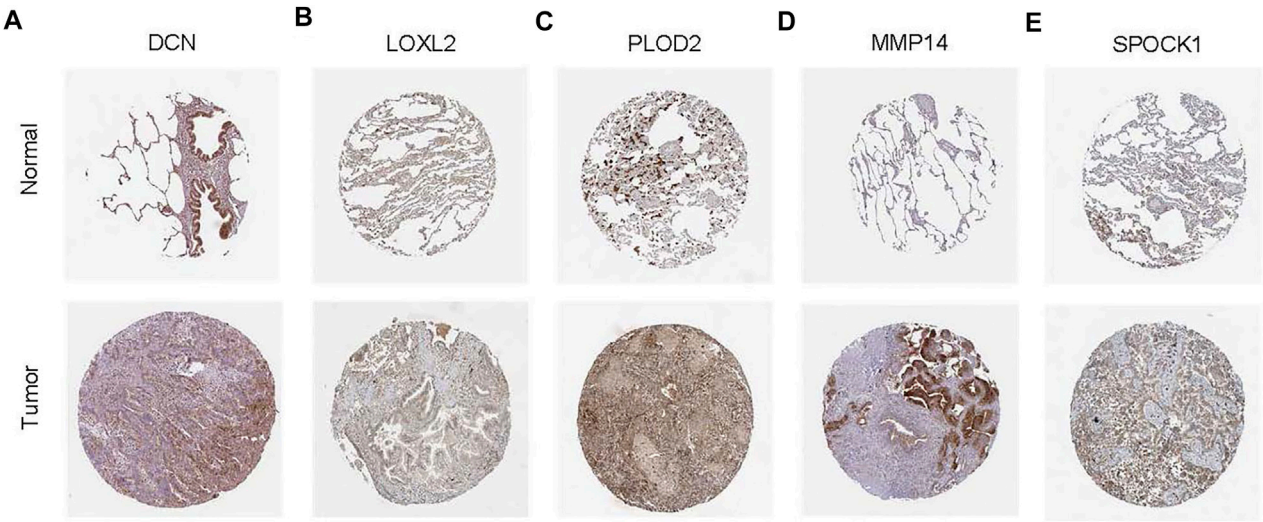


FIGURE 12
Representative immunohistochemical stains of the five prognostic genes in the HPA database. **(A)** Expression of DCN protein in LUAD and normal control samples. **(B)** Expression of LOXL2 protein in LUAD and normal control samples. **(C)** Expression of PLOD2 protein in LUAD and normal control samples. **(D)** Expression of MMP14 protein in LUAD and normal control samples. **(E)** Expression of SPOCK1 protein in LUAD and normal control samples.

differences in the efficacy of immunotherapy between the two groups. Therefore, we examined the response to ICI therapy using the TIDE algorithm. Results showed that patients in the low-risk group were more likely to respond to immunotherapy.

Further analysis showed that the six ERGs signature predicted the responses to several common chemotherapeutic agents. Particularly, the low-risk group was more sensitive to chemotherapy drugs such as, paclitaxel, docetaxel, and gemcitabine compared with the high-risk group. The lower response in the high-risk group may be due to EMT-induced drug resistance. Paclitaxel and docetaxel, two chemotherapy drugs belonging to the taxene family, have been shown to induce cell cycle arrest in cancer cells by preventing microtubule depolymerization. EMT-induced invasive behavior of cancer cells can cause tolerance towards Paclitaxel and docetaxel. Upstream mediators of EMT, such as ZEb1/2, TGF-β, and microRNA regulate response of cancer cells to Paclitaxel and docetaxel (Ashrafizadeh et al., 2021). During EMT process, the conversion of E-cadherin to N-cadherin reduces the expression of human balanced nucleoside transporter 1 (hENT1), a drug carrier for gemcitabine membrane transport in cancer cells, which triggers gemcitabine resistance in cancer cells (Weadick et al., 2021). In general, we found that patients with low-risk may benefit more from immunotherapy and chemotherapy compared with those with high-risk. This implies that more studies are needed to develop new treatment strategies or multi-drug combinations to improve prognosis of high-risk patients.

Overall, the proposed ERGS signature showed good performance in predicting the prognosis of LUAD. Our results highlight the need to investigate the role of EMT in the progression of LUAD. The constructed prognostic model can also be used to evaluate the tumor immune microenvironment, guide application of individualized therapy, and facilitate the development of targeted therapy. Nevertheless, this study has some limitations. First, this was a retrospective study and independent prospective cohorts are needed to validate the prognostic model developed in the study. The value of six genes as potential targets also needs further investigations. Second, this risk model is highly relied on public databases. As a result of the clinical information downloaded from TCGA and GEO databases is limited or incomplete data, potential prognostic factors, such as personal clinical history and treatment intervention, are missing in our nomogram. It is not clear How the environmental factors such as smoking and exposure to certain toxins might have influence the identified gene signatures. Further investigations need to be undertaken in future clinical researches.

Conclusion

This study constructed a novel prognostic model based on 6 EMT-related genes. The model was established in the training cohort and validated using an external validation cohort, RT-PCR tests, and IHC assays. The results showed that the model was a robust biomarker for predicting the OS in LUAD patients. Furthermore, according to the TME analysis and evaluation of chemotherapy efficacy, the features indirectly demonstrated that patients in the low-risk group based the model had a higher likelihood to benefit from immune therapy and chemotherapy. This study provides new reference findings for further exploration of the mechanisms of EMT and tumor immunity. It also provides insights to guide personalized treatment of LUAD patients.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Medical Association of the Shanxi Province Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors were involved in drafting the article or revising it critically for important content. Q-ZS and FL developed the methodology and acquired the related data. Conception and design of the study: FL and Q-ZS. Acquisition of data: Y-FZ, X-RW, L-MC, and NL Analysis and interpretation of data: L-XZ, S-XZ, X-FZ, Y-FZ, and X-RW. Drafting the article: Y-FZ, X-RW, and L-MC Revising the article critically: L-XZ, S-XZ, and X-FZ. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful for the joint efforts of the members of the research group. We also extend our thanks to all participants of this research.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1008416/full#supplementary-material

References

Akalay, I., Janji, B., Hasmim, M., Noman, M. Z., André, F., De Cremoux, P., et al. (2013). Epithelial-to-mesenchymal transition and autophagy induction in breast carcinoma promote escape from T-cell-mediated lysis. *Cancer Res.* 73, 2418–2427. doi:10.1158/0008-5472.CAN-12-2432

Ashrafizadeh, M., Mirzaei, S., Hashemi, F., Zarrabi, A., Zabolian, A., Saleki, H., et al. (2021). New insight towards development of paclitaxel and docetaxel resistance in cancer cells: EMT as a novel molecular mechanism and therapeutic possibilities. *Biomed. Pharmacother.* 141, 111824. doi:10.1016/j.biopha.2021.111824

Basu, S., Cheriyamundath, S., and Ben-Ze'ev, A. (2018). Cell-cell adhesion: Linking wnt/β-catenin signaling with partial EMT and stemness traits in tumorigenesis. *F1000Res.* 7, F1000 Faculty Rev-1488. doi:10.12688/f1000research.15782.1

Bradshaw, A. D., and Sage, E. H. (2001). SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. *J. Clin. Invest.* 107, 1049–1054. doi:10.1172/JCI12939

Chen, Q., Yang, D., Zong, H., Zhu, L., Wang, L., Wang, X., et al. (2017). Growth-induced stress enhances epithelial-mesenchymal transition induced by IL-6 in clear cell renal cell carcinoma via the Akt/GSK-3β/β-catenin signaling pathway. *Oncogenesis* 6, e375. doi:10.1038/oncsis.2017.74

Chen, Y., Terajima, M., Yang, Y., Sun, L., Ahn, Y. H., Pankova, D., et al. (2015). Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. *J. Clin. Invest.* 125, 1147–1162. doi:10.1172/JCI74725

Cohen, E. N., Gao, H., Anfossi, S., Mego, M., Reddy, N. G., Debeb, B., et al. (2015). Inflammation mediated metastasis: Immune induced epithelial-to-mesenchymal transition in inflammatory breast cancer cells. *PLoS One* 10, e0132710. doi:10.1371/journal.pone.0132710

Cui, G., Cai, F., Ding, Z., and Gao, L. (2019). MMP14 predicts a poor prognosis in patients with colorectal cancer. *Hum. Pathol.* 83, 36–42. doi:10.1016/j.humpath.2018.03.030

Datar, I., and Schalper, K. A. (2016). Epithelial-mesenchymal transition and immune evasion during lung cancer progression: The chicken or the egg? *Clin. Cancer Res.* 22, 3422–3424. doi:10.1158/1078-0432.CCR-16-0336

Devarakonda, S., Morgensztern, D., and Govindan, R. (2015). Genomic alterations in lung adenocarcinoma. *Lancet. Oncol.* 16, e342–51. doi:10.1016/S1470-2045(15)00077-7

Dong, B., Wu, C., Huang, L., and Qi, Y. (2021). Macrophage-related SPP1 as a potential biomarker for early lymph node metastasis in lung adenocarcinoma. *Front. Cell. Dev. Biol.* 9, 739358. doi:10.3389/fcell.2021.739358

Garcia-Lora, A., Algarra, I., and Garrido, F. (2003). MHC class I antigens, immune surveillance, and tumor immune escape. *J. Cell. Physiol.* 195, 346–355. doi:10.1002/jcp.10290

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

Guo, H., Wang, S., Ju, M., Yan, P., Sun, W., Li, Z., et al. (2021). Identification of stemness-related genes for cervical squamous cell carcinoma and endocervical adenocarcinoma by integrated bioinformatics analysis. *Front. Cell. Dev. Biol.* 9, 642724. doi:10.3389/fcell.2021.642724

Gupta, G. P., and Massagué, J. (2006). Cancer metastasis: Building a framework. *Cell.* 127, 679–695. doi:10.1016/j.cell.2006.11.001

Horváth, Z., Kovalszky, I., Fullár, A., Kiss, K., Schaff, Z., Iozzo, R. V., et al. (2014). Decorin deficiency promotes hepatic carcinogenesis. *Matrix Biol.* 35, 194–205. doi:10.1016/j.matbio.2013.11.004

Iwatsuki, M., Mimori, K., Yokobori, T., Ishi, H., Beppu, T., Nakamori, S., et al. (2010). Epithelial-mesenchymal transition in cancer development and its clinical significance. *Cancer Sci.* 101, 293–299. doi:10.1111/j.1349-7006.2009.01419.x

Kim, M. J., Cervantes, C., Jung, Y. S., Zhang, X., Zhang, J., Lee, S. H., et al. (2021). PAF remodels the DREAM complex to bypass cell quiescence and promote lung tumorigenesis. *Mol. Cell.* 81, 1698–1714.e6. e6. doi:10.1016/j.molcel.2021.02.001

Lu, L., Liu, L. P., Zhao, Q. Q., Gui, R., and Zhao, Q. Y. (2021). Identification of a ferroptosis-related LncRNA signature as a novel prognosis model for lung adenocarcinoma. *Front. Oncol.* 11, 675545. doi:10.3389/fonc.2021.675545

Meng, Y., Sun, J., Zhang, G., Yu, T., and Piao, H. (2021). Clinical prognostic value of the PLOD gene family in lung adenocarcinoma. *Front. Mol. Biosci.* 8, 770729. doi:10.3389/fmolb.2021.770729

Papait, A., Stefani, F. R., Cargnoni, A., Magatti, M., Parolini, O., and Silini, A. R. (2020). The multifaceted roles of MSCs in the tumor microenvironment: Interactions with immune cells and exploitation for therapy. *Front. Cell. Dev. Biol.* 8, 447. doi:10.3389/fcell.2020.00447

Peng, D. H., Ungewiss, C., Tong, P., Byers, L. A., Wang, J., Canales, J. R., et al. (2017). ZEB1 induces LOXL2-mediated collagen stabilization and deposition in the extracellular matrix to drive lung cancer invasion and metastasis. *Oncogene* 36, 1925–1938. doi:10.1038/onc.2016.358

Prateep, A., Sumkhemthong, S., Karnsomwan, W., De-Eknamkul, W., Chamni, S., Chanvorachote, P., et al. (2018). Avicequinone B sensitizes anoikis in human lung cancer cells. *J. Biomed. Sci.* 25, 32. doi:10.1186/s12929-018-0435-3

Qiu, D., Hu, J., Hu, J., Yu, A., Othmane, B., He, T., et al. (2021). PMEPA1 is a prognostic biomarker that correlates with cell malignancy and the tumor microenvironment in bladder cancer. *Front. Immunol.* 12, 705086. doi:10.3389/fimmu.2021.705086

Rotow, J., and Bivona, T. G. (2017). Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* 17, 637–658. doi:10.1038/nrc.2017.84

Stawowczyk, M., Wellenstein, M. D., Lee, S. B., Yomtoubian, S., Durrans, A., Choi, H., et al. (2017). Matrix metalloproteinase 14 promotes lung cancer by cleavage of heparin-binding EGF-like growth factor. *Neoplasia* 19, 55–64. doi:10.1016/j.neo.2016.11.005

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tan, B., Chen, Y., Xia, L., Yu, X., Peng, Y., Zhang, X., et al. (2021). PMEPA1 facilitates non-small cell lung cancer progression via activating the JNK signaling pathway. *Cancer Biomark.* 31, 203–210. doi:10.3233/CBM-200966

Terry, S., Savagner, P., Ortiz-Cuaran, S., Mahjoubi, L., Saintigny, P., Thiery, J. P., et al. (2017). New insights into the role of EMT in tumor immune escape. *Mol. Oncol.* 11, 824–846. doi:10.1002/1878-0261.12093

Thiery, J. P., Acloque, H., Huang, R. Y., and Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell.* 139, 871–890. doi:10.1016/j.cell.2009.11.007

Wakiyama, H., Masuda, T., Motomura, Y., Hu, Q., Tobo, T., Eguchi, H., et al. (2018). Cytolytic activity (CYT) score is a prognostic biomarker reflecting host immune status in hepatocellular carcinoma (HCC). *Anticancer Res.* 38, 6631–6638. doi:10.21873/anticanres.13030

Wang, Q., Li, M., Yang, M., Yang, Y., Song, F., Zhang, W., et al. (2020a). Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: Implications for immune checkpoint blockade therapy. *Aging (Albany NY)* 12, 3312–3339. doi:10.18632/aging.102814

Wang, T., Liu, X., Tian, Q., Liang, T., and Chang, P. (2018). Reduced SPOCK1 expression inhibits non-small cell lung cancer cell proliferation and migration through Wnt/β-catenin signaling. *Eur. Rev. Med. Pharmacol. Sci.* 22, 637–644. doi:10.26355/eurrev_201802_14288

Wang, Z., Li, Z., Zhou, K., Wang, C., Jiang, L., Zhang, L., et al. (2021a). Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. *Nat. Commun.* 12, 6500. doi:10.1038/s41467-021-26770-2

Wang, Z., Wang, Y., Yang, T., Xing, H., Wang, Y., Gao, L., et al. (2021b). Prediction of RBP binding sites on circRNAs using an LSTM-based deep sequence learning architecture. *Brief. Bioinform.* 22, bbab342. doi:10.1093/bib/bbab342

Weadick, B., Nayak, D., Persaud, A. K., Hung, S. W., Raj, R., Campbell, M. J., et al. (2021). EMT-induced gemcitabine resistance in pancreatic cancer involves the functional loss of equilibrative nucleoside transporter 1. *Mol. Cancer Ther.* 20, 410–422. doi:10.1158/1535-7163.MCT-20-0316

Yamauchi, M., and Sricholpech, M. (2012). Lysine post-translational modifications of collagen. *Essays Biochem.* 52, 113–133. doi:10.1042/bse0520113

Yan, T., Lin, Z., Jiang, J., Lu, S., Chen, M., Que, H., et al. (2015). MMP14 regulates cell migration and invasion through epithelial-mesenchymal transition in nasopharyngeal carcinoma. *Am. J. Transl. Res.* 7, 950–958.

Yao, J., Chen, X., Liu, X., Li, R., Zhou, X., and Qu, Y. (2021). Characterization of a ferroptosis and iron-metabolism related lncRNA signature in lung adenocarcinoma. *Cancer Cell. Int.* 21, 340. doi:10.1186/s12935-021-02027-2

Yin, Y., Shelke, G. V., Lässer, C., Brismar, H., and Lötvall, J. (2020). Extracellular vesicles from mast cells induce mesenchymal transition in airway epithelial cells. *Respir. Res.* 21, 101. doi:10.1186/s12931-020-01346-8

Zhao, H., Zhang, X., Guo, L., Shi, S., and Lu, C. (2021). A robust seven-gene signature associated with tumor microenvironment to predict survival outcomes of patients with stage III-IV lung adenocarcinoma. *Front. Genet.* 12, 684281. doi:10.3389/fgene.2021.684281

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 30 September 2022
DOI 10.3389/fgene.2022.1005658

OPEN ACCESS

EDITED BY
Yi Yao,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Cuiling Zheng,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Giovanni Brandi,
University of Bologna, Italy

*CORRESPONDENCE
Qinglin Shen,
qinglinshen@whu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 28 July 2022
ACCEPTED 09 September 2022
PUBLISHED 30 September 2022

CITATION
Hu M, Yao W and Shen Q (2022), Advances and challenges of immunocheckpoint inhibitors in the treatment of primary liver cancer. *Front. Genet.* 13:1005658. doi: 10.3389/fgene.2022.1005658

Advances and challenges of immunocheckpoint inhibitors in the treatment of primary liver cancer

Meng Hu[1†], Weirong Yao[1†] and Qinglin Shen[1,2*]

[1]Department of Oncology, Jiangxi Provincial People's Hospital, the First Affiliated Hospital of Nanchang Medical College, Nanchang, China, [2]Institute of Clinical Medicine, Jiangxi Provincial People's Hospital, the First Affiliated Hospital of Nanchang Medical College, Nanchang, China

Primary liver cancer (PLC) is one of the most common malignant tumors, which clinically characterized by occult onset, rapid development, easy recurrence and poor prognosis. With the rapid development of tumor immunotherapy research, tumor immunotherapy has also achieved remarkable clinical efficacy, and jointly promoted the overall improvement of tumor immunology from mechanism research to clinical transformation, from single discipline to multi-disciplinary integration. Immunotherapy has obvious advantages in treatment-related toxicity and efficacy compared with traditional therapy. In hepatocellular carcinoma (HCC), immunotherapy alone or in combination with other therapies may help to control tumor progression, and there are many immune checkpoint inhibitors (ICIs) widely used in clinical or ongoing clinical trials. However, tumor immunology research is still facing many challenges. How to effectively evaluate the efficacy, whether there are related biomarkers, the generation of immune tolerance and the lack of clinical trials to objectively evaluate the efficacy are still urgent problems to be solved, but it also brings new research opportunities for basic and clinical immunology researchers. The study of treatment of ICIs of PLC has become a hot spot in clinical research field. This paper summarizes and prospects the research progress and challenges of ICIs for PLC.

KEYWORDS

primary liver cancer, immune checkpoint inhibitors, immunotherapy, advances, challenges

Introduction

Primary liver cancer (PLC) includes hepatocellular carcinoma (HCC), cholangiocarcinoma (CCA) and mixed hepatocellular carcinoma -cholangiocarcinoma (HCC-CCA). HCC accounts for around 90% of the total number of PLC, with incidence rate of fifth in men and eighth in women. According to the global cancer statistics in 2020, HCC is the sixth largest cancer in the world, with the death rate ranking the fourth in the world (Sung et al., 2021). There are at least 700,000 new cases of HCC raised every year. The number of patients with HCC in China accounts for about 50–55% of the total

number of patients in the world (Jemal et al., 2013), and the mortality accounts for 50% of the total number of patients with HCC in the world. The main pathogenic factors of HCC include: hepatitis B virus (HBV) infection, alcoholic liver disease, diabetes, and some metabolic diseases (Mcglynn et al., 2015). The routine treatment of HCC consists of surgery, chemoradiotherapy, targeted therapy and immunotherapy (De Lorenzo et al., 2018; Tai et al., 2019). Base of the characteristics of PLC and the degree of malignancy, most of the patients with PLC are in the middle and advanced stage when they are initially diagnosed, who would almost lose the chance of surgery treatment, so, the systemic drug therapy is considered as the proper management (Waese et al., 2017; Rizzo et al., 2021b). In December of 2005, the food and Drug Administration (FDA) approved Solafeni for the treatment of first-stage renal cancer. Since that, the period of targeted treatment of PLC was officially initiated, then, the molecular targeted drugs such as Lenvatinib and Regorafenib were licensed subsequently in treatment of HCC. However, the curative effect of tyrosine kinase inhibitors (TKIs) is limited due to the emergence of drug resistance (Saffo and Taddei, 2019; Rizzo et al., 2020).

With the rapid development of molecular biology, studies have found that the liver is important immune organ of the body (Solter and Philip, 2005; Szabo et al., 2018), because there are a large number of macrophages and immune cells in the liver microenvironment, which makes it form a very complex immune tolerance microenvironment (Peterson., 2012). Therefore, immunotherapy for PLC arises at the historic moment. Studies have confirmed that immunotherapy could enhance the body's immune response, break the immune tolerance, activate the body's immune cells to recognize and kill tumor cells, so as to obtain anti-tumor effect (Scheinberg and Pinilla-Ibarz, 2006; Scheinberg and Pinilla-Ibarz, 2009). Immunotherapy for PLC would stimulate the body's immune response to tumor cells and regulate the immune microenvironment of PLC through various ways of immune activation, so as to achieve the anti-tumor effect through the interaction of immune cells and molecules (Cao et al., 2005; Cantor et al., 2013; Rizzo et al., 2021a). Although the immunotherapy of PLC has made gratifying progress, it still faces many problems, e.g., the related immune escape and combined therapy of PLC. This paper here focuses on the summarization of the advances and challenges of ICIs in the treatment of PLC.

Immunosuppression mechanism in microenvironment of PLC

First of all, tumor microenvironment (TME) mainly refers to the internal and external environment of the body where tumor occurs and develops, which plays an important role in the process of tumor development, immune escape, body immunosuppression and anti-tumor (Chen and Hua, 2012). Under physiological conditions, intestinal metabolites and bacteria would enter the liver through the portal system, and act as antigens to stimulate the liver immune system to maintain homeostasis and avoid excessive autoimmune reaction, which is feasible to establish immune tolerance microenvironment. Therefore, PLC is also called immune amnesty organ (Fernández et al., 2019).

China is a traditional country with a large population of hepatitis B. PLC is often derived from chronic hepatitis B cirrhosis and chronic hepatitis patients. Due to the interaction of various inflammatory cells, the liver is immersed in more complex chronic inflammatory microenvironment (Yoon et al., 2010; Altomonte and Ebert, 2014). With the development of molecular biology and the studies in mechanism of malignant tumor, the hepatoma cells are identified in a highly inhibited immune microenvironment. Through down regulation of the main histocompatibility complex-I(MHC-I), secretion of immunosuppressive cytokines and negative co-stimulation signals (Lowe et al., 2014), the host immunosuppression is induced, result in avoiding the autoimmune response. The particularity of PLC leads to the complexity and challenge of its immunotherapy (Yuan et al., 2017; Nguyen et al., 2021). The results show that (Shiraha et al., 2020), tumor related fibroblasts in TME can release a large number of immunosuppressive related molecules, such as prostaglandin E2; In addition, the risk of recurrence of PLC after liver transplantation is correlated with Th1 cells and interferon-γ, and the high expression of these cytokines may be related to the prognosis of tumor. In addition, $CD8^+$ T cell is one of the main immune cells that able to identify and kill tumor cells. In the microenvironment of PLC, the function of $CD8^+$ T cell is inhibited to promote the rapid growth of HCC cells (Du and Wang, 2011). For example, in PLC, bone marrow cells differentiate into more immature myeloid cells, which can affect the immune microenvironment of PLC at a certain stage (Perussia et al., 1984). This cell is also called the marrow derived suppressor cells (MDSCs) due to its immature characteristics and remarkable diversity. Different main immune cells were differentiated in variety of environments and time, including: dendritic cells (DCs), macrophages and neutrophils, which are called tumor related DCs, tumor associated macrophages (TAMs) (Thiem et al., 2021), tumor associated neutrophils (TANs). The microenvironment of tumor also includes cancer associated fibroblasts (CAFs), tumor infiltrating lymphocytes (TILs). Simultaneously, there are many immunosuppressive pathways in the microenvironment of PLC, which exert the progress of PLC and immune escape (Zhu et al., 2019). For instance, in the process of chronic hepatitis caused by long-term hepatitis virus (mainly hepatitis B), many kinds of inhibitory immune factors would be secreted, which promote the occurrence and proliferation of malignant tumor cells (Timperi and Barnaba, 2020); The endogenous cell cycle related kinase (CCRK) can be applied into the liver through EZH2/NF-kB signaling pathway to

TABLE 1 Progress of various immune checkpoint inhibitors in the treatment of primary liver cancer.

Drug name	Characteristics of cohort study	Target (mechanism)	Patient selection	Intervention measures	grouping	Clinical results (ORR, PFS, OS)	Main side effects
Atilizumab + Bevacizumab (A+ T)	Imbravei50 phase III multicenter study	PD-1+ VEGF inhibitor	501 unresectable HCC patients who had not received systematic treatment before	They were randomly assigned into the experimental group and the control group according to 2:1. The experimental group received 1200 mg intravenous infusion of atilizumab, followed by 15 mg/kg intravenous infusion of bevacizumab on the same day, Q3W; The control group was treated with sorafenib 400 mg orally twice a day until the disease progressed or intolerable toxicity appeared	A+ T *vs.* Sorafenib	A+ T *vs.* Sorafenib: mOS 19.2 *vs.* 13.4 m(HR = 0.66, p = 0.0009), mPPS 6.9 *vs.* 4.3 m, ORR 29.8% *vs.* 11.3%	38% of people had serious AE (Grade 3–4), and the most common AEs (in ≥20% of patients) were hypertension, fatigue and proteinuria
Camrelizumab + apatinib	Non-random, open, multi center, phase II project carried out by 25 centers in China	PD-1+VEGFR-2 inhibitor	Phase II clinical study on the first-line and second-line treatment of 190 cases of advanced liver cancer in rescue	Camrelizumab (intravenous, 200 mg [for body weight ≥50 kg] or 3 mg/kg [for body weight <50 kg], Q2W) + Apatinib (250 mg/day, Q4W)	Camrelizumab + Apatinib *vs.* Apatinib	In the first-line treatment group, mOS was 20.3 m, mPFS was 5.7 m, ORR was 34.3%	The frequency of TRAs above Grade 3 was 78.6%
Nivolumab + Ipilimumab (O+ Y)	Checkmate 040 cohort 4 phase I/II global multicenter single arm study	PD-L1+CTLA-4	Second line phase I/II study of 148 cases of advanced liver cancer	The patients were randomly assigned according to the ratio of 1 : 1 : 1. They received 1 mg/kg of Nivolumab combined with 3 mg/kg of Ipilimumab Q3W (4 doses), and then 240 mg of Nivolumab Q2W (group A); Nivolumab 3 mg/kg combined with Ipitumab 1 mg/kg Q3W (4 doses), and then 240 mg Nivolumab Q2W (group B); Or 3 mg/kg of Nivolumab Q2W and 1 mg/kg of Ipilizumab Q6W (Group C)	Group A (Nivolumab 1 mg/kg, Ipilimumab 3 mg/kg, Q3W, followed by Nivolumab 240 mg Q2W after 4 courses of treatment); Group B (Nivolumab 3 mg/kg, Ipilimumab 1 mg/kg, Q3W, followed by Nivolumab 240 mg every 2 weeks after 4 courses of treatment); Group C (Nivolumab 3 mg/kg, Ipilimumab 1 mg/kg, Q6W)	(mOS : 22.8 *vs.* 12.5 m *vs.* 12.7 m, ORR: 32% *vs.* 27% *vs.* 29%)	The TRAEs of double immunosuppressants were slightly higher, 3/4 of the patients in group A had AEs, but they were controllable as a whole

(Continued on following page)

10.3389/fgene.2022.1005658

TABLE 1 (*Continued*) Progress of various immune checkpoint inhibitors in the treatment of primary liver cancer.

Drug name	Characteristics of cohort study	Target (mechanism)	Patient selection	Intervention measures	grouping	Clinical results (ORR, PFS, OS)	Main side effects
Durvalumab + Tremelimumab(T + D)	Study22 VII global multicenter research	PD-L1+CTLA-4	Phase II clinical study on second-line treatment of 322 cases of advanced HCC	T300 + D75 (T 300 mg + D 1500 mg, sequential D 1500 mg after one course of treatment, Q4W)	T300 + D75 (T 300 mg + D 1500mg, sequential D 1500 mg after one course of treatment, Q4W)	mOS: 18.7 *vs.* 11.7 m *vs.* 17.1 *vs.* 11.3 m; ORR: 22.7% *vs.* 9.6% *vs.* 7.2% *vs.* 9.5%	Grade 3–4 AE: 35.1% *vs.* 17.8%*vs.*
				D 104 (D 1500 mg Q4W);	T69 (T 750mg, Q4W in the first 7 cycles and Q12W after 7 cycles);		42.0% *vs.* 24.4%
				T69 (T 750mg, Q4W in the first 7 cycles and Q12W after 7 cycles)	T75 + D 84 (T 75 mg + D 1500mg, sequential D drug 1500 mg after 4 cycles, Q4W)		
				D 104 (D 1500 mg Q4W);			
Nivolumab (O)	Checkmate 459 phase III global multicenter study	PD-1	493 cases of advanced HCC	Checkmate 459 is a randomized, multicenter, phase III clinical study involving 743 patients with advanced liver cancer aged 18 or over who did not receive systematic treatment. 1: 1 after randomization, 371 patients received intravenous 240 mg navulizumab Q2W, and 372 patients took 400 mg sorafenib orally twice a day	Navuliumab(371) *vs.* Sorafenib(372)	OS:16.4 *vs.* 14.7 m; PFS: 3.7 *vs.* 3.8 m; ORR: 15 *vs.* 7%	Grade 3–4 AE: 22 vs. 49%
Pembrolizumab	Keynote 240 phase III global multicenter study	PD-1	Second line treatment of 413 cases of advanced HCC	The patients were randomly assigned to receive Pembrolizumab 200 mg + best supportive treatment vs. Placebo + best supportive treatment according to 2:1, Q3W	Pembrolizumab(278)vs. Placebo(135)	OS:13.9 *vs.* 10.6 m; PFS: 3.3 *vs.* 2.8 m; ORR: 18.3 vs. 4.4%	Grade 3–4 AE: 18.6vs. 7.5%
Camrelizumab	Phase II China multicenter single arm study	PD-1	220 patients were included, of which the proportion of HBV infection was as high as 83.4%	Camrelizumab 3 mg/kg, Group(Q2W)*vs.* Group(Q3W) = 110:110	Q2W(3 mg/kG)vs. Q3W(3 mg/kG)	OS:14.2 vs. 13.2 m; PFS: 2.3 *vs.* 2.0m; ORR: 11.9 *vs.* 17.6%	Grade 3–4 AEs: 22%, mainly reactive skin capillary hyperplasia, and most patients mainly have grade 1–2 AEs

(Continued on following page)

10.3389/fgene.2022.1005658

TABLE 1 (*Continued*) Progress of various immune checkpoint inhibitors in the treatment of primary liver cancer.

Drug name	Characteristics of cohort study	Target (mechanism)	Patient selection	Intervention measures	grouping	Clinical results (ORR, PFS, OS)	Main side effects
Sintilimab + Bevacizumab	Randomized, controlled, open phase III clinical study	Domestic PD-1 + VEGF inhibitor	571 cases of unresectable or metastatic HCC treated with first-line therapy	The patients were randomly divided into groups according to 2:1 and received Sintilimab + Bevacizumab *vs*. Sorafenib	Sintilimab + Bevacizumab *vs*. Sorafenib	mOS:Not reached(NR) *vs*.10.4m; mPFS: 4.6 vs. 2.8m; the risk of death and the risk of disease progression were reduced 43%and 44%, respectively	The incidence of grade 3–4 AEs was similar to that of sorafenib
Toripalimab + Lefatini	Single arm phase II	RTK inhibitor + PD-1 + HIC	36 adult patients with advanced HCC (≥18 years old) had ECoG score of 0–2 and child Pugh class a liver function	Lefatini (8 mg when body weight <60 kg, 12 mg when body weight ≥60kg, oral once a day) was used 3–7 days before the initial HAIC to determine tolerance. They were then treated with lenvatinib for 21 days (one cycle from day 1 to day 21), treprizumab (intravenous infusion of 240 mg on day 1), HAIC (day 1 to day 2) and FOLFOX regimen (oxaliplatin 85 mg/m2, folic acid 400 mg/m2, 5-fluorouracil 400 mg/m2 on day 1 and 5-fluorouracil 2400 mg/m2 for 224 h), Until the disease progresses or intolerable toxicity occurs	Toripalimab + Lefatini *vs*. Lefatini	PFS: 11.1 *vs*. 5.1 m, OS: Not reached(NR) *vs*. 11.0m, ORR:66.7% (CR14.1%),DCR:90.1%	Grade 3–4 TRAs (trae) occurred in 72.2% of patients. The most common were thrombocytopenia (13.9%), elevated aspartate aminotransferase (AST; 13.9%) and hypertension (11.1%)

10.3389/fgene.2022.1005658

upregulates the production of IL-6, inducing MDSCs to gather in TME, which plays a role in stimulating tumor growth (Zhou et al., 2017); Some studies have shown that the proliferation of liver tumor cells will generate local hypoxia to induce the increase of MDSCs, leading to the progress of tumor cell cloning (Chiu et al., 2017). In addition, there are also studies release TANs could raise macrophages and Tregs cells into hepatoma cells, and enhance the growth and progress of HCC (Zhou et al., 2016; Jiang et al., 2017). These immunosuppressive factors play a regulatory role in the occurrence and development of PLC to different degrees.

In short, the occurrence and development of PLC need to be formed through a variety of ways and factors, and the mechanism is relatively complex.

Current status of ICIs for HCC

At present, the studies of immune checkpoint inhibitors (ICIs) is hot in the research of malignant tumor, especially in the immunotherapy of lung cancer, and then spread into the research of a variety of other malignant tumors. The targets of immunosuppressive agents mainly include programmed death-1 (PD-1)/programmed death ligand 1 (PD-L 1) and cytotoxic T lymphocyte associated antigen-4 (CTLA-4), also many other encouraging clinical results have been reported. Progress of various ICIs in the treatment of PLC summarized in Table 1.

Current status of PD - 1/PD—L 1 blocking therapy

Ishida and his colleague (Ishida et al., 1992; Dong et al., 1999; Fitzsimmons and Sadkowsky, 2002) first discovered that PD-1 (CD279) can induce apoptosis in mouse T-cell hybridoma since 1992, there have been more and more studies on PD-1/PD-L1, which has become a “superstar” in cancer research. Among them, PD-1 is a negative costimulatory molecule of CD28 immunoglobulin superfamily of transmembrane receptor. It is a powerful inhibitor of effector T cells response. It is found in a variety of immune cells, such as T cells, B cells and NK cells. It is mainly expressed in $CD8^+$ T cells. It can also be expressed in bone marrow-derived suppressor cells and Treg cells, PD-1 has two kinds of cell membrane protein ligands: PD-L1 (B7-H1/CD274) and PD-L2 (B7-DC/CD273). The process of interaction between PD-1 and PD-L1/PD-L2 is mainly that PD-1 binds with PD-L1/PD-L2, transmits the co inhibitory signal of T cells antigen receptor and inhibits the production of various cytokines by suppressing the activation of T cells, thus assisting tumor immune escape. The higher PD-1 expression level affects the proliferation and differentiation of Treg cells, and further regulates the peripheral immune tolerance response (Jilaveanu et al., 2014).

PD-1 inhibitors

Nivolumab

Nivolumab (BMS936558, MDX-1106, Opdivo) is a completely humanized IgG4 monoclonal antibody targeting PD-1. Since 2014, Nivolumab has been approved by FDA for secondary treatment of metastatic melanoma and NSCLC, and it has been approved by FDA for bladder cancer in February, 2017 and second line treatment of HCC in late stage treated by Sorafenib in September, 2017. The famous Checkmate- 040 (El-Khoueiry et al., 2017) includes two parts: phase I dose climbing study (n = 48) and ll phase queue extension study (n = 214). The results show that the disease control rate (DCR) is 58%, objective response rate (ORR) is 15%, and the overall survival (OS) is 15 months, The OS of 6 and 9 months was 66%. The ORR was 20%, DCR was 64%, and the OS in 6 and 9 months were 83 and 74%, respectively. Compared with the first-line treatment of advanced HCC Checkmate-459 (NCT02576509) (Yau et al., 2019b), the results of the global and multicenter III trials of Sorafenib showed that the median overall survival (mOS) of Nivolumab group and Sorafenib group were 16.4 and 14.7 months (p = 0.0419), respectively. However, the expression of PD-L1, the efficacy of Nivolumab was consistent, and the median progression free survival (mPFS) of both groups were 3.7 and 3.8 months, respectively, there was no significant difference in mPFS, with the ORR of 15 and 7%, respectively. At the same time, the safety of Nivolumab was higher, and 22 and 49% of treatment related adverse reactions (TRAEs) in the Nivolumab group and Sorafenib group were respectively.

Pembrolizumab

Pembrolizumab is a highly selective and humanized IgG 4 monoclonal antibody, which can target to inhibit negative regulation of PD-L signal. The clinical efficacy and safety are similar to that of Nivolumab. In 2018, a non-randomized, multicenter, open phase II single arm clinical study Keynote-224 assessed the safety and efficacy of pembrolizumab in the late stage of HCC. A total of 104 patients with advanced Child-Pugh A who were treated with Sorafenib were enrolled in the study. The study included 104 patients with advanced Child-Pugh A who had been treated with Sorafenib for 2 years, until the disease progresses or an intolerable toxic reaction occurs. The results showed that the DCR was 62%, objective remission rate was 17%, complete remission rate was 1%, mPFS reached 4.9 months, OS was 12.9 months, 12 months OS rate was 54%, the incidence of level 3 related adverse reactions was 16%, and the adverse reactions above grade 4 did not occur, and the adverse reactions were mainly caused by the increase of aspartate transaminase. It is precisely because of the Keynote-224 (Zhu et al., 2018) research results that in 2018, the FDA approved pembrolizumab for the second-line treatment of advanced HCC

patients, and the second PD-1 inhibitor approved by FDA for advanced HCC. In 2019, a randomized controlled phase III clinical trial Keynote-240 (NCT02702401) compared the best support therapy with pembrolizumab for the second-line treatment of advanced HCC. The results showed that pembrolizumab had significant effect. The study included 413 patients with advanced PLC, randomly assigned to the pembrolizumab group and the control group according to 2: 1, and followed up for 13.8 months. The total survival time of pembrolizumab group was prolonged for 3 months (13.9 *vs.* 10.6 m); The results showed that PFS was improved (p = 0.0022) in one side (p = 0.0238). Unfortunately, the difference did not reach the established statistical level. The ORR of pembrolizumab and placebo group was 18.3 and 4.4%, respectively. The efficacy of pembrolizumab group was long-lasting, and the median DOR was 13.8 months. In terms of safety, the treatment group had increased transaminase, bilirubin, diarrhea, rash, *etc.* most of the adverse reactions were 1-2, and 3-4 were rare. Keynote-240 study did not reach the expected results. The reasons for the end point, a value and p value adjustment of the study were found and the clinical research continued to be carried out. The Keynote-394 study, as a second-line treatment, is expected to be used in the randomized, double-blind, international multicenter phase III clinical study in patients with advanced Asian HCC. We believe that the clinical research of PD-1 inhibitor in the treatment of HCC will be more and more profound in the future, and that better clinical data will be obtained.

Camrelizumab

Camrelizumab is the first PD-1 inhibitor independently developed by HENGRUI company of China to be approved as an indication for advanced HCC. In 2019, Professor Qin Shukui led a multicenter, phase II clinical study (Qin et al., 2020) (NCT02989922) on second-line and above treatment of advanced HCC in China. A total of 220 patients were recruited in the study, of which the proportion of HBV infection was as high as 83.4%. They were randomly assigned to Camrelizumab 3 mg/kg, every 2 weeks (Q2W) and 3 mg/kg, Q3W at a ratio of 1:1. The results showed that the overall mPFS was 2.1 months, so the ORR of patients was 14.7% (the ORR of 2W group and 3W group were 11.9 and 17.6%, respectively). The OS of all patients was 13.8 months, the OS of 2W treatment group was 14.2 months, and the OS of 3W group was 13.2 months. The OS rates of all patients at 6 and 12 months were 74.7 and 55.9%, respectively. There was no significant difference in ORR between every 2 weeks and every 3 weeks. The main adverse reactions were reactive cutaneous capillary hyperplasia, and most of the patients had grade 1–2 adverse reactions. The safety was similar to that of Nivolumab and pembrolizumab. In March 2020, it was approved by the National Medical Products Administration (NMPA) as the second-line treatment for advanced HCC.

Tislelizumab

In the clinical research of immunotherapy for advanced HCC, immunotherapy has been unanimously recommended by domestic and foreign HCC guidelines. However, some clinical studies still fail to obtain gratifying results, such as Keynote-240 and Checkmate-459 (Vogel et al., 2020) studies, which did not reach the preset statistical significance. Tislelizumab is a PD-1 monoclonal antibody independently developed in China (Lee and Keam, 2020). It is also a humanized IgG4 anti-PD-1 monoclonal antibody. FC segment structure optimization of Tislelizumab can effectively avoid antibody-dependent cellular phagocytosis (ADCP) effect. It has a good T cell activation effect in PLC cells with immune cell aggregation, and has high affinity and specific binding ability with PD-1. FC segment of TisleliTzumab can also be optimized by genetic engineering technology to make it interact with macrophage FCγ (Zhang et al., 2018) , On 26 December 2019, NMPA approved the market.

The BGB-A317-001 study (Desai et al., 2016; Wu et al., 2019) explored the efficacy, safety and tolerability of Tislelizumab. The study is divided into phase 1A dose climbing and dose exploration. In conclusion, after receiving Tislelizumab monotherapy for more than 12 months, patients can still maintain good tolerance. Regardless of the expression of PD-L1, Tislelizumab monotherapy can produce lasting anti-tumor effect in a variety of solid tumor patients. Currently, BGB-A317-208, BGB-A317-301 and other similar studies have been carried out.

Sintilimab

Sintilimab injection is a monoclonal antibody against human IgG4, which can specifically bind to PD-1 molecule on the surface of T cells, thus blocking PD-1/PD-L1 pathway leading to tumor immune tolerance, starting T cells to kill tumor cells, so as to achieve the purpose of anti-tumor. ESMO conference in 2020 reported a single arm phase II clinical study of Sintilimab combined with Anlotinib in the first-line treatment of patients with advanced HCC (Chen X. et al., 2020). In this study, a total of 14 evaluable patients had ORR as high as 42.9% (RECIST1.1), 5 patients had partial remission (PR), 1 patient had complete remission (CR), and the DCR rate was 92.9%, with good tolerance. Another clinical study, a randomized, controlled, and open phase III clinical study (Ren et al., 2020) (ORIENT-32), explored the comparison between Sintilimab combined with bevacizumab and sorafenib in the first-line treatment of unresectable HCC patients. Nearly 600 Chinese patients were included in the study. The study showed that the ORR of Sintilimab combined with bevacizumab was 5 times that of sorafenib group. The combined group reduced the risk of death and disease progression by 43%. The mOS of the two groups were not reached and 10.4 months, and the mPFS were

4.6 and 2.8 months, respectively. The safety was impressive. The incidence of grade 3–4 TRAE was only 33.7%. Because of the success of this pioneering study, it has become the first phase III clinical study of PD-1 combination therapy for first-line advanced HCC with positive results in the world. At the same time, this research result makes Sintilimab combined with bevacizumab first-line recommended by NCCN guidelines for patients with advanced HCC. Therefore, new schemes and ideas are added for patients with advanced HCC.

Toripalimab

On 17 December 2018, the State Drug Administration approved the first homemade PD-1 monoclonal antibody-toripalimab injection. Toripalimab is an IgG4 type monoclonal antibody with independent intellectual property rights developed by JUNSHI (Jiao Y. et al., 2020), which has a unique binding site. At the same time, a proline (S228P) point mutation was introduced into the serine protein site 228 in the hinge region of the heavy chain to increase the stability of the antibody. It has dual antitumor effects. Its anti-tumor mechanism is mainly to block PD-1/PD-L1 signaling pathway, improve T cell response activity *in vitro* and *in vivo*, and achieve T cell proliferation and anti-tumor effect (Greenwald, 2008); In addition, it mediates the endocytosis of PD-1, reduces the expression of PD-1 membrane, and relieves T cell immunosuppression. A real-world clinical study of domestic PD-1 inhibitor monotherapy for HBV related PLC (Chen J. et al., 2020) showed that the ORR of Toripalimab was 15.4%, the DCR was 53.8%, the total ORR was 17.3%, and the DCR was 72.0%. On 20 March 2021, the second CSCO-JUNSHI biological tumor immunity summit forum announced the results of initial analysis of phase II study of first-line treatment of advanced HCC with Toripalimab combined with bevacizumab, and announced the launch of international multi-center phase III clinical study (JUPITER-10, NCT04723004), A total of 54 patients with unresectable locally advanced or metastatic HCC were enrolled in this phase II study. 87% of the patients had chronic hepatitis B and more than half had extrahepatic metastasis. The ORR was 47.7%. Research data demonstrate that most of the adverse events are mild grade one or two adverse events, and there are no grade four or more serious adverse events. At present, the overall data is not yet mature, the mPFS and mOS have not yet reached, and the research is still in progress, which is worth looking forward to.

Panaprizumab

Panaprizumab is a PD-1 monoclonal antibody developed by a joint venture established by KANGFANG biomedical Co., Ltd. and China Institute of Biopharmaceuticals. It is characterized by complete removal of FC receptor and complement mediated function of Panaprizumab through FC mutation, and slow antigen binding and dissociation rate. These characteristics make it possible for Panaprizumab to become an anti-PD-1 drug with better clinical benefits. At present, in China, the new drug application of Panaprizumab for the treatment of relapsed or refractory (R/R) classical Hodgkin's lymphoma (R/R CHL) has been accepted by Center for Drug Evaluation (CDE) in May 2020 (Song et al., 2019; Mislang et al., 2020).and published in the International Symposium on gastrointestinal cancer (ASCOGI) in 2021. As of November 2020, the confirmed ORR, DCR, mPFS, 6-month PFS, and 6-month OS were 31.0, 82.8, 63.2 and 93.2%, respectively. The incidence of grade 3 and above TRAE associated with Panaprizumab or Anlotinib was 19.4%, and the incidence of serious adverse events associated with Panaprizumab or Anlotinib was 6.5%. Studies suggest that the combination of Panaprizumab and Anlotinib is safe and tolerable, as the first-line treatment for patients with advanced HCC (Kotasek et al., 2019; Shan et al., 2021), showing encouraging antitumor activity. At the same time, the results support the exploration of a phase III clinical trial (NCT04344158) for the first-line treatment of advanced HCC with higher doses of Anlotinib combined with Panaprizumab, and Panaprizumab combined with Anlotinib (10 mg, continuous for two weeks, withdrawal for one week).

PD-L1 inhibitors

PD-L1 inhibitors mainly include Atezolizumab, Avelumab and Durvalumab. There are few clinical studies on PD-L1 inhibitors in the treatment of advanced HCC.

Durvalumab

Durvalumab is a humanized monoclonal antibody (IgG1 K type) against programmed death receptor ligand 1 (PD-L1) expressed in Chinese hamster ovary cells (CHO). In 2017, a phase I-II clinical study on Durvalumab in the treatment of advanced HCC patients who failed to receive sorafenib (Wainberg et al., 2017) included 40 patients with advanced HCC. After treatment with Durvalumab (10 mg/kg), the results showed that ORR was 10%, DCR was 33.3%, mOS was 13.2 months, and the incidence of grade 3–4 adverse reactions was 20%. The effect of single drug treatment was good. At present, there are more and more clinical studies, and the clinical study of combined immunotherapy in lung cancer has a clear curative effect, and the clinical study in PLC is under study.

Atezolizumab

Atezolizumab is a monoclonal antibody that can bind to PD-L1 and block its interaction with PD-1 and B7.1 receptor. These include activation of anti-tumor immune response without inducing antibody dependent cytotoxicity. In the syngeneic mouse tumor model, blocking PD-L1 activity resulted in decreased tumor growth. The phase 1b clinical study was reported at the ESMO meeting in 2019 (Stein et al., 2018). The ORR of 59 patients with advanced HCC treated with Atezolizumab as the first-line treatment was 17%. Compared with Atezolizumab combined with bevacizumab, the mPFS was 3.4 and 5.6 months, respectively. However, more and larger phase III clinical studies are needed to further confirm its efficacy and safety.

CTLA-4 blocker

On 25 March 2011, the United States Food and Drug Administration (FDA) approved the use of CTLA-4 monoclonal antibody (Ipilimumab) in the treatment of advanced melanoma, which has become a major breakthrough in the field of immunotherapy (Trinh and Hwu, 2012). Although the drug research stage has a good survival rate, but combined with a variety of related adverse reactions, it is criticized. In 2015, the US FDA granted the qualification for Tremelimumab, a monoclonal antibody against CTLA-4 inhibitor of AstraZeneca, for the treatment of mesothelioma. Tremelimumab is a human IgG2 monoclonal antibody against CTLA-4. Binding with CTLA-4 can prevent it from binding with B7 ligand, thus inhibiting the decline of T cell activity mediated by B7-CTLA-4. It can stimulate the immune system and anti-tumor effect. CTLA-4 inhibitors can recognize and eliminate tumor cells by enhancing the activity of antigen-presenting cells and T lymphocytes. A phase II clinical study (Sangro et al., 2013) (NCT01008358) included 20 patients with advanced PLC. After Tremelimumab treatment, ORR was 17.6%, DCR was 76.4%, mPFS was 6.48 months, mOS was 8.2 months, and viral load decreased. At the same time, some studies have shown that CTLA-4 inhibitors can bring OS, PFS, DCR benefits to advanced HCC. At the same time, studies have revealed that Tremelimumab has anti hepatitis virus effect. It is believed that in the future, with the continuous exploration of relevant phase III clinical studies, the application prospect of this drug in the treatment of advanced HCC will be expansive, which may bring longer survival hope to more patients with advanced PLC.

Double immunity

The effect of single drug immunotherapy is limited in patients with advanced HCC. Combined immunotherapy has become a research hotspot in the treatment of advanced HCC. Studies have shown that (Hellmann et al., 2016) the combination of ICIs with different mechanisms of action could improve the response rate and anti-tumor effect of patients. The combination of CTLA-4 and PD-1/PD-L1 monoclonal antibody inhibitor would raise the tumor response rate and generate the synergistic effect. In 2017, the phase I clinical study of Tremelimumab combined with Durvalumab in the treatment of patients with advanced HCC (Kanikarla Marie et al., 2021) illustrated that the ORR of Tremelimumab combined with Durvalumab was 8% (25 *vs.* 17%) higher than that of monotherapy, and the DCR of Tremelimumab combined with Durvalumab for 16 weeks was 67.5%. Subsequently, a randomized, open, multicenter phase III trial (Abou-Alfa et al., 2018) (NCT03298451) with more patient samples was conducted to study the efficacy of Durvalumab +/- Tremelimumab compared with sorafenib in the treatment of patients with advanced HCC. The trial expanded the sample number, and 1200 patients were expected to be recruited. The main end point was OS. In 2019, the American Society of Clinical Oncology (ASCO) reported the results of the Checkmate-040 update study (He et al., 2020). In the Checkmate- 040 multi cohort study, 148 patients with advanced HCC who failed to receive sorafenib were included in this study. The results showed that the total ORR was 31% and the *duration of remission*.

(DOR) was 4.6–30.5 months, OS also achieved good results. At the same time, the meeting also reported the results of combined treatment of Nivolumab, Ipilimumab and Cabozantinib. The results indicated that the total ORR of 35 patients was 29%, the OS of 15 months was 70%, and the 3/4 grade of trail was 71%. Of course, a variety of clinical studies on the combination of ICIs in the treatment of patients with advanced HCC are in progress, which is worth to expect.

Immunosuppressant combined with molecular targeted drugs

Molecular targeted drugs for PLC have been proved to have anti -angiogenesis effect, which can affect the immune response of PLC (Tian et al., 2017). In 2018, the ASCO reported a phase IB clinical study of Atezolizumab combined with bevacizumab in the first-line treatment of advanced HCC (Stein et al., 2018). The results showed that 21 cases had an evaluable ORR of 62%, and the effect was obvious. The results of the international multicenter phase III clinical study (Finn et al., 2018) (imbrave150) released at the ESMO meeting in 2019 further confirmed that Atezolizumab combined with bevacizumab is a new first-line regimen for the treatment of advanced HCC superior to sorafenib. Atezolizumab combined with bevacizumab significantly improved the overall survival of advanced HCC. The 6-month mOS rate of the combination group was 85%, the mOS did not reach, and the mPFS was 6.8 months; The 6-month mOS rate of sorafenib group was 72%, the mOS was 13.2 months, the mPFS was 4.3 months, the ORR of combined group was 27%, and the ORR of sorafenib group was 12%. Studies have shown that (Ikeda et al., 2018) Lenvatinib combined with PD-1/PD-L1 monoclonal antibody has a synergistic effect, which can block PD-1 immune escape pathway, and also inhibit monocytes to differentiate into TAM related to immune escape. At the same time, Lenvatinib would also inhibit the growth of tumor cells and TAM by competitively inhibiting the combination of VEGF and VEGFR. A phase IB clinical study was reported at the ASCO meeting in 2019 (Kudo et al., 2021). The results showed that the ORR of Avelumab combined with axitinib in the first-line treatment of 22 patients with advanced HCC was 31.8%, and the mOS was 12.7 months, while the ORR of single drug Avelumab in the treatment of advanced HCC was 13.6%.

ASCO conference in 2020 reported a phase IB clinical study Keynote-524 (Finn et al., 2020). The study showed that the total ORR of Lenvatinib combined with pembrolizumab in the treatment of patients with advanced HCC was 46%, and the mPFS and mOS were 9.3 and 22 months respectively. The main safety aspects were proteinuria and elevated alanine aminotransferase. The results showed that the efficacy and safety of Lenvatinib combined with ICIs were better than monotherapy. It is worth further exploring in the treatment of PLC by combination drug therapy. In 2020 ASCO gastrointestinal conference, a phase IB clinical study of first-line treatment of advanced HCC with Nivolumab combined with Lenvatinib was reported (Kudo et al., 2020). The preliminary results indicated that mPFS was 7.39 months and ORR was 76.7%. The ASCO meeting in 2020 also reported a phase Ib/II clinical study on the first-line treatment of patients with advanced HCC with Panaprizumab (PD-1) combined with Anlotinib (Jiao S. C. et al., 2020). A total of 22 patients were included. The results showed that the total DCR was 84%, the median OS was not reached, and the 6-month OS was 91.6%. The data were satisfactory. At the same time, a phase III clinical study of Lenvatinib combined with pembrolizumab is in progress. In the second-line treatment clinical research, the phase I clinical trial of Camrelizumab combined with Apatinib in the treatment of patients with advanced HCC who failed to be treated by sorafenib (Mei et al., 2021). The experimental group was treated with Camrelizumab combined with Apatinib, Camrelizumab 200 mg, once every 2 weeks, and the control group was treated with single drug Apatinib. The clinical data proved the clinical benefits of ORR, DCR, mPFS, etc. There are 6 related wall reports in 2020 ESMO annual meeting, all of which are phase I/II studies. The drugs are selected in different scenario, including first-line and second-line schemes, but the safety and effectiveness are validated in all. The reported drug regimens include: Apatinib combined with Camrelizumab, Lenvatinib combined with CS1003, Sintilimab combined with Anlotinib, etc (Xu et al., 2019; Xu et al., 2020; Chen et al., 2021). Cabozantinib plus atezolizumab might be a treatment option for select patients with advanced HCC (NCT03755791), mPFS was 6·8 m (99% CI 5·6–8.3) in the combination treatment group *versus* 4·2 m (2·8–7.0) in the sorafenib group (hazard ratio [HR] 0·63, 99% CI 0·44–0.91, $p = 0·0012$). mOS (interim analysis) was 15·4 m (96% CI 13·7–17·7) in the combination treatment group *versus* 15·5 m (12·1-not estimable) in the sorafenib group (HR 0·90, 96% CI 0·69–1.18; $p = 0·44$) (Kelley et al., 2022).

Immunotherapy combined with chemotherapy

In the past, the overall therapeutic effect of chemotherapy for PLC is not satisfactory. Studies have confirmed that chemotherapy boost the exposure of tumor cells to antigens, which is conducive to the immune effect of ICIs and enhance anti-tumor efficacy. Studies have shown that domestic ICIs combined with classical chemotherapy sometimes bring a surprising curative effect. A multi-center phase II study on the first-line treatment of advanced HCC with Camrelizumab combined with FOLFOX4/GEMOX, 34 cases were evaluable, the confirmed ORR was 26.5%, and the mDOR has not yet reached, DCR was 79.4%; mPFS was 5.5 m. At the same time, the tolerance and related adverse reactions of patients were acceptable (Qin et al., 2019), in other words, combined immunochemotherapeutic is an optional choice for advanced PLC patients. In addition, a randomized, open, national multicenter phase II/III clinical trial of first-line treatment of advanced HCC with Treprizumab combined with bevacizumab analogue and FOLFOX4 regimen is expected.

Immunoneoadjuvant therapy

In 2020, the ASCO annual meeting published a research report on neoadjuvant therapy. A total of 30 patients with resectable PLC were enrolled in the study. Three cycles of bevacizumab combined with Ipilimumab or bevacizumab monotherapy were served for HCC before operation. After surgical resection, the pathological complete remission rate was 24% (3 cases in combination group, 2 cases in monotherapy group), and the main pathological remission rate was 16% (necrotic effect, 2 cases in combination group, 1 case in monotherapy group) (Yau et al., 2019a). In 2021, ASCO-GI reported a study on neoadjuvant treatment of borderline resectable or locally advanced HCC with Cabotinib combined with navulizumab (Yau et al., 2020), and the results showed that 12/15 achieved marginal negative resection, 5/12 of the patients who had been operated successfully achieved complete pathological response.

Adjuvant therapy

With the first-line, second-line and neoadjuvant treatment of PLC have achieved gratifying outcome, the application of target therapy in adjuvant treatment of PLC is also constantly in exploration. For example, in June 2020, a multi-center study explored the adjuvant effect of iodine Metoximab after radical resection of PLC. 156 HCC patients with CD147 expression were included in the study. Patients in the treatment group were injected with iodine (Li et al., 2020) rituximab *via* hepatic artery once 4–6 weeks after operation. Results of the 5-year RFs time in the treatment group was significantly increased compared with that in the blank control group (43.4%: 21.7%, $p = 0.0031$), suggesting that the adjuvant regimen can improve the prognosis of patients. The study suggests that patients with PLC should consider on more accurate treatment protocols

according to the classification of biomarker subgroup. At the 2020 ASCO annual meeting, the mid-term analysis results of a multi-center, prospective cohort study (Lance study) on the adjuvant therapy of Lenvatinib combined with transcatheter arterial chemoembolization (TACE) (Gao et al., 2010)for patients with HCC at high risk of recurrence after operation were released. A total of 90 patients with high risk of recurrence after radical surgery were enrolled in this study (macrovascular or bile duct invasion/tumor rupture or invasion of adjacent organs/grade II microvascular invasion with any of the following: tumor number ≥3, maximum tumor diameter ≥ 8cm, unclear tumor margin or incomplete capsule). The results showed that the median disease-free survival (DFS) time of patients in the combination group was significantly longer than that in the TACE group (12.0 *vs.* 8.0 m, $p < 0.05$); HR = 0.5,$p = 0.0359$) (Kudo et al., 2019). These results verify that the adjuvant therapy of Lenvatinib combined with TACE is not only effective and safe, but also prolong the PFS of HCC patients with high recurrence risk.

Transformation therapy

The real-world study of PD-1 combined with TKIs for the treatment of PLC reported by Professor Sun Huichuan in 2020CSCO? in China shows that PD-1 combined with TKIs could be applied into the treatment of PLC. In 60 patients with advanced unresectable PLC, 11 patients were converted to resectable after receiving PD-1 combined with TKIs. When the data was published, 9 patients completed the operation, 5 of them achieved pathological complete remission, and the estimated survival time was more than 1 year (Cheng et al., 2019). 2020 ESMO-Asia reported a prospective, non-randomized, open cohort study (Zhu et al., 2021) of Professor Lu Shichun's team in Beijing 301 Hospital: the study of HCC transformation therapy of TKIs combined with PD-1 inhibitor in the treatment of vascular invasion. As of 20 May 2020, a total of 70 patients were screened and 39 patients were included, of which 35 patients received combined treatment and 30 patients with PVTT, there were 2 cases of venous tumor thrombus and 3 cases of both. The criteria of successful transformation: 1. Child Pugh score <7; 2、ECOG PS≤1; 3. There was no extrahepatic lesion; 4. The hepatic vascular structure was intact and FLR was sufficient. The results showed that the median follow-up time was 7.2 months, no recurrence rate was 60% 6 months after operation, OS and RFs were still not to the end point, and the conversion and resection rate was 30.3%. At the same time, Professor Huang Cheng of Zhongshan Hospital in Shanghai, China, has also made some explorations in this field. As of December 2020, 23 unresectable HCC patients have been enrolled. After TKI combined with PD-1 antibody treatment, the tumor has been transformed and resected.

Immunotherapy combined with local therapy

Previous studies (Marincola et al., 2000) have shown that radiotherapy is closely related to immunotherapy. Radiotherapy can affect the immune microenvironment of PLC cells and stimulate the production of some inflammatory cytokines (Presicce et al., 2009). Immunotherapy can also increase the sensitivity of radiotherapy; At the same time, radiotherapy also has the influence on immunosuppression. Radiotherapy can promote the body to produce immune response, induce the tumor cells to produce immunogenic death, and activate T lymphocytes. It recognizes and kill tumor cells in and out of the radiotherapy field, leading to the occurrence of the so-called "distant effect" after radiotherapy of malignant tumors. Therefore, in theory, radiotherapy combined with immunotherapy would obtain better curative effect. Since the Pacific study confirmed that immunotherapy consolidation therapy can be used as the standard treatment for unresectable non-small cell lung cancer after concurrent chemoradiotherapy, the survival benefit of this kind of tumor patients has been improved to a new height. So, what are the clinical results of immunotherapy combined with radiotherapy in the treatment of advanced HCC? PD-L1 expression level was monitored after local radiotherapy in the experiment. The results showed that radiotherapy combined with PD-L1 group had a significant inhibitory effect on tumor growth. The 7-week survival rates of combined treatment group, radiotherapy group and PD-L1 group were 90, 30 and 0%, respectively.

In the aspect of immunotherapy combined with microwave ablation of PLC, there is also preliminary theoretical evidence that ablation of PLC can produce a large number of inflammatory factors, a variety of immunogenic mediators and chemokines to play an anti-tumor role (Zhang et al., 2017). At the same time, ablation of PLC can increase the expression of HSP70, which may be one of the important reasons for the enhancement of anti-tumor immunity (Nikfarjam et al., 2005). Studies have shown that CTLA-4 monoclonal antibody combined with thermal ablation in the treatment of advanced PLC can significantly reduce the viral load of patients with hepatitis BLC, down regulate Treg cells in TME, increase the infiltration of CD8+T cells in tumor site, and increase the survival rate of patients with advanced PLC, the 6-month and 12-month DFS rates were 57.1 and 33.1%, respectively (Duffy et al., 2017; Kudo and Masatoshi, 2017).

The most important advantage of TACE in the treatment of PLC is to effectively treat PLC and avoid liver injury. After TACE, malignant tumor cells are lysed and necrotic, which can produce a large number of highly immunogenic cell components, thus initiating the related immune response (Zavadil et al., 2019). Tremelimumab combined with TACE (NCT01853618) is considered as an adjuvant therapy for advanced HCC. About 26% of the preliminary results show that it is partially effective. A

phase III clinical study (NCT03778957) explored the efficacy of TACE combined with bevacizumab and Durvalumab in patients with locally advanced HCC. At the same time, another study (NCT0281754) evaluated the efficacy of Durvalumab combined with Tremelimumab (TACE/RAF/cryoablation) in the treatment of advanced HCC. These clinical studies are in progress. In general, the large-scale clinical research of ICIs combined with local therapy for PLC has been or is in progress, which is worth looking forward to.

In recent years, local arterial infusion chemotherapy (HAIC) for PLC has achieved satisfactory results, and has become a "new star" in the research of local treatment of PLC (Verhoef et al., 2008). In 2020, Professor Shi Ming retrospectively analyzed the efficacy and safety of Lenvatinib + Toripalimab + HAIC in 157 patients with advanced PLC (He et al., 2021b). The results showed that the triple therapy group showed longer PFS (11.1 *vs.* 5.1 m, $p < 0.001$); Longer OS (less than: 11 m, $p < 0.001$); And higher ORR (RECIST: 59.2%: 9.3%, $p < 0.001$); mRECIST:67.6%:16.3%, p <0.001); The higher DCR rate (RECIST or mRECIST: 90.1 *vs.* 72.1%, $p = 0.005$) (He et al., 2021a) indicated that HAIC combined with Lenvastinib and Treprizumab could significantly improve the efficacy of treatment on patients with advanced PLC. These studies preliminarily show that HAIC has a remarkable effect in the treatment of advanced PLC.

Challenge and thinking of immunotherapy for liver cancer

ICIs have been widely used in the clinical treatment of advanced HCC, and achieved excellent results. It has changed the new pattern of systemic treatment of HCC. At present, the combination therapy of immunosuppressive agents for HCC is emerging in an endless stream. With the positive results of clinical studies and the accelerated approval of FDA, the combination therapy will become the mainstream of the treatment of HCC in the future. However, with the progress of clinical practice, there are many problems and challenges to be solved in immunotherapy of HCC, such as the requirement of exploration more effective combination modes, optimization of the existing drug treatment sequence, development of new drug and cell combination schemes, the choice of biomarkers, the challenges of drug economy and safety, etc.

Effective biomarkers for predicting efficacy

Although immunosuppressive agents for HCC have shown very encouraging efficacy, they also face the challenge of drug resistance. Therefore, it is very important to find effective biomarkers in the evaluation. PD-L1 is a common biomarker of malignant tumor in clinic, but the clinical guiding value of PD-L1 is different for different malignant tumors. Professor Pinato's study showed that the positive expression rate of PD-L1 in (Pinato et al., 2019) PLC cells was less than 10%, while the correlation between the positive expression of PD-L1 and the efficacy of ICIs was not found in Keynote-224 and Checkmate-040 studies (El-Khoueiry et al., 2017; Zhu et al., 2018). Whether the expression of PD-L1 is related to the efficacy of PD-1/PD-L1 monoclonal antibody in the treatment of advanced HCC is still uncertain. Some clinical studies (Carbone et al., 2018; Klempner et al., 2020) have shown that TMB is an independent predictor for evaluating the therapeutic effect of ICI for a variety of tumors, and TMB is positively correlated with the efficacy of immunotherapy (Klempner et al., 2020). The results of the clinical trial Keynote - 158 make the tumor mutation load the second tumor associated diagnostic marker approved for clinical application (Marabelle et al., 2020). However, some studies have also shown that compared with other malignant tumors, TMB is the second tumor associated diagnostic marker approved for clinical application, the expression level of tumor mutation load in HCC is not significant, and its predictive value of curative effect is limited (Yarchoan et al., 2017). Therefore, the predictive effect of TMB on the curative effect and the determination of the cut-off value still need to be further explored. In addition, mismatch repair defects and microsatellite instability only occur in 2–3% of PLC patients, and their application value in PLC is very limited. In 2019, ESMO reported the predictive value of neutrophil to lymphocyte ratio (NLR) and platelet lymphocyte ratio (PLR) in the treatment of PLC with Nivolumab. The results show that the degree of lymphocyte infiltration in TME is closely related to the heterogeneity of HCC.

There is a correlation regions with high tumor heterogeneity have higher degree of immune infiltration and better response to immunotherapy (Losic et al., 2020). Harding et al. (Harding et al., 2019). Showed that patients with CTNNB1 mutation had poor response to PD-1 inhibitors, which may indicate the predictive role of CTNNB1 mutation in curative effect. Therefore, the clinical value of PD-1, TMB, microsatellite instability and other commonly used biomarkers in the treatment of PLC is limited, and more research is needed to find more effective biomarkers.

How to improve the combination therapy strategy

In addition, the data of local treatment combined with systemic treatment is still insufficient. All kinds of phase I/II combination therapy show promising efficacy, but there is still a lack of confirmed results of phase III study. Immunotherapy combined with targeted therapy or double immunotherapy has a high ORR, which also provides the possibility of transformation

therapy for unresectable or borderline resectable PLC (Tarantino and Curigliano, 2020). However, at present, there is no clinical recommendation of the most effective combination mode. With the continuous improvement of evidence-based medicine, systematic treatment may participate in the whole process of advanced PLC treatment. The specific combination therapy, interventional therapy timing and dose course still need to be further standardized and refined.

Real-world research status of ICIs

At present, many large-scale clinical studies of ICIs in the treatment of malignant tumors have achieved satisfactory results (Mazdak et al., 2021), but in the real world, ICIs in the treatment of HCC do not have such a convincing effect (Shintaro et al., 2008). Randomized controlled trials (RCTs) are generally accepted in the medical community to evaluate the safety and effectiveness of drugs, but the inclusion and exclusion criteria of RCTs are too strict. The results of the study are not completely consistent with the real-world research, and it is also a difficult problem to constantly select the dominant population for refinement.

Progress of immunization

The research of ICI has brought a breakthrough in the field of tumor research. Immunosuppresses can effectively inhibit tumor growth and progress of disease e.g.HCC. However, immuno hyperprogress is a tough topic to face. Champiat et al. (Champiat et al., 2017) proposed that the definition of the super progress is that the time to treatment failure (TTF) is less than 2 months, the tumor load increases by more than 50%, and the progress pace (PP) increases more than twice as much as the progress. The incidence of immune hyperevolution is about 4–30%, which may be related to the types of tumor. The pathogenesis of immune hyperprogress is not clear. Some studies have reported the potential clinical and biological predictors of hyperprogress (Demets, 2013). The possible predictors are over 65 years old, two or more metastasis sites, gender, low expression of PD-L1, etc. Of course, the diagnosis criteria of the super progress caused by immunotherapy needs the involvement of histopathology.

How to transform “cold tumor” into “hot tumor"

HCC is a special high immune type cancer, most of the patients belong to the characteristics of “cold tumor immune cycle”. Cold tumor is the tumor with no or little immune cell infiltration around it (Galon and Bruni, 2019). Cold tumor with poor prognosis, is often the most difficult tumor to treat while “hot tumor” is the opposite. The transformation of cold tumor into thermal tumor is usually realized through the combination therapy. Most of the cold tumors can be transformed into thermotumor by direct infusion of activated immune cells *in vitro* through the adoptive cell therapy, and the immune effect ability can be enhanced. It can also be transformed into thermotumor by combining radiotherapy and chemotherapy and local treatment (Gabriele et al.). At present, the understanding of “cold and hot tumor” is not comprehensive, and the research on the transformation of “cold tumor” is still a big challenge in immunotherapy.

Security challenges

With the wide application of immunotherapy and the acceptance of immunotherapy drugs in medical insurance, the price of immunotherapy drugs has decreased significantly, and the incidence of related adverse reactions has further increased, such as immune myocarditis, immune pneumonia, immune hepatitis, enteritis, etc. How to reduce the incidence of immune related adverse reactions would be a giant challenge in the near future (Lee et al., 2021).

Summary and prospect

At present, the treatment of PLC has entered the era of immune 3.0, especially the application of ICIs has become the most promising drug for the treatment of PLC. Although the effect of single immunotherapy is accountable, the combined immunotherapy shows us the new hope in the treatment of PLC. In a number of studies, we have seen that the combined immunotherapy has achieved the consequence of $1 + 1 > 2$. For example, the combination of Atezolizumab and bevacizumab has achieved good results in both ORR and PFS (Wallin et al., 2016). In the future, we will continue to explore a variety of combination models based on immunotherapy, such as the combination of immunity and immunity, immunity combined with radiotherapy, immunity combined with ablation, immunity combined with chemotherapy and targeted drugs, immunity combined with intervention, etc. Meanwhile, we also need to think about how to combine treatment for patients with different types and stages, which is a raising problem we are facing. In the future, the drug resistance of immunotherapy is bound to affect the curative effect of PLC treatment. To study the drug resistance mechanism of PLC and the corresponding new drugs needs our continuous efforts and innovation, and further clinical research is required to provide evidence-based medicine to choose the

optimal scheme for the individualized treatment of PLC patients.

Author contributions

QS and WY conceived the project. All authors collected and analyzed the data. QS and MH prepared the tables. QS wrote the manuscript. All authors edited and commented on the manuscript.

Funding

This work was supported by a grant from the Science and technology project of Jiangxi Health Commission (Grant No. 202130003).

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

Abou-Alfa, G. K., Chan, S. L., Furuse, J., Galle, P. R., Sangro, B., Qin, S., et al. (2018). A randomized, multicenter phase 3 study of durvalumab (D) and tremelimumab (T) as first-line treatment in patients with unresectable hepatocellular carcinoma (HCC): HIMALAYA study. *J. Clin. Oncol.* 36 (15), TPS4144. doi:10.1200/jco.2018.36.15_suppl.tps4144

Altomonte, J., and Ebert, O. (2014). Sorting out pandora's box: Discerning the dynamic roles of liver microenvironment in oncolytic virus therapy for hepatocellular carcinoma. *Front. Oncol.* 4, 85. doi:10.3389/fonc.2014.00085

Cantor, H., Kim, H. J., and Lu, L. (2013). *Discovery of regulatory T cells programmed to suppress an immune response*.

Cao, Y., Hegewisch-Becker, S., Blum, I., Bartels, K., Atan Ac Kovic, D., Leuwer, R., et al. (2005). A local enrichment of regulatory T cells within the tumor tissue might suppress an effective anti-tumor T cell response in patients with head and neck cancer. *J. Clin. Oncol.* 23 (16), 9666. doi:10.1200/jco.2005.23.16_suppl.9666

Carbone, D. P., Sharpnack, M., and He, K. (2018). "Abstract IA20: Immunotherapy: Biomarkers and checkpoint blockade in NSCLC," in *Abstracts: Fifth AACR-IASLC international joint conference: Lung cancer translational science from the bench to the clinic* (San Diego, CA)

Champiat, S., Dercle, L., Ammari, S., Massard, C., Hollebecque, A., Postel-Vinay, S., et al. (2017). Hyperprogressive disease is a new pattern of progression in cancer patients treated by anti-PD-1/PD-L1. *Clin. Cancer Res.* 23 (8), 1920–1928. doi:10.1158/1078-0432.Ccr-16-1741

Chen, C. H., and Hua, B. J. (2012). *The role of inflammation and inflammatory microenvironment in Initiation,Progress and metastasis of lung cancer*. Medical Recapitulate.

Chen, J., Hu, X., Li, Q., Dai, W., Yuan, G., Huang, W., et al. (2020a). Effectiveness and safety of toripalimab, camrelizumab, and sintilimab in a real-world cohort of Hepatitis B virus associated hepatocellular carcinoma patients. *Ann. Transl. Med.* 8 (18), 1187. doi:10.21037/atm-20-6063

Chen, X., Li, W., Wu, X., Zhao, F., Shu, Y., Gu, Y., et al. (2020b). 170P Sintilimab plus anlotinib as first-line therapy in patients (pts) with advanced hepatocellular carcinoma (aHCC). *Ann. Oncol.* 31, S1305. doi:10.1016/j.annonc.2020.10.191

Chen, X., Li, W., Wu, X., Zhao, F., Shu, Y., Wu, H., et al. (2021). Sintilimab plus anlotinib as first-line therapy in patients (pts) with advanced hepatocellular carcinoma (aHCC). *J. Clin. Oncol.* 39 (15), e16146. doi:10.1200/jco.2021.39.15_suppl.e16146

Cheng, H., Sun, G., Chen, H., Li, Y., Han, Z., Li, Y., et al. (2019). Trends in the treatment of advanced hepatocellular carcinoma: Immune checkpoint blockade immunotherapy and related combination therapies. *Am. J. Cancer Res.* 9 (8), 1536

Chiu, K. C., Tse, P. W., Xu, M. J., Cui, J. D., Lai, K. H., Li, L. L., et al. (2017). Hypoxia inducible factor HIF-1 promotes myeloid-derived suppressor cells accumulation through ENTPD2/CD39L1 in hepatocellular carcinoma. *Nat. Commun.* 8 (1), 517. doi:10.1038/s41467-017-00530-7

De Lorenzo, S., Tovoli, F., Barbera, M. A., Garuti, F., Palloni, A., Frega, G., et al. (2018). Metronomic capecitabine *vs.* best supportive care in child-pugh B hepatocellular carcinoma: A proof of concept. *Sci. Rep.* 8 (1), 9997. doi:10.1038/s41598-018-28337-6

Demets, D. L. (2013). *The role and potential of surrogate outcomes in clinical trials: Have we made any progress in the past decade*. New York: Springer.

Desai, J., Markman, B., Sandhu, S. K., Gan, H. K., and Millward, M. (2016). "A phase I dose-escalation study of BGB-A317, an anti-programmed death-1 (PD-1) mAb in patients with advanced solid tumors," in *ASCO Annual meeting*. Boston.

Dong, H., Zhu, G., Tamada, K., and Chen, L. (1999). B7-H1, a third member of the B7 family, co-stimulates T-cell proliferation and interleukin-10 secretion. *Nat. Med.* 5 (12), 1365–1369. doi:10.1038/70932

Du, C., and Wang, Y. (2011). The immunoregulatory mechanisms of carcinoma for its survival and development. *J. Exp. Clin. Cancer Res.* 30 (1), 12. doi:10.1186/1756-9966-30-12

Duffy, A. G., Ulahannan, S. V., Makorovarusher, O., Rahma, O., Wedemeyer, H., Pratt, D., et al. (2017). Tremelimumab in combination with ablation in patients with advanced hepatocellular carcinoma. *J. Hepatol.* 66 (3), 545–551. doi:10.1016/j.jhep.2016.10.029

El-Khoueiry, A. B., Sangro, B., Yau, T., Crocenzi, T. S., Kudo, M., Hsu, C., et al. (2017). Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): An open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 389, 2492–2502. doi:10.1016/S0140-6736(17)31046-2

Fernández, J., Luddy, K., Harmon, C., and O'Farrelly, C. (2019). Hepatic tumor microenvironments and effects on NK cell phenotype and function. *Int. J. Mol. Sci.* 20 (17), 4131. doi:10.3390/ijms20174131

Finn, R. S., Ducreux, M., Qin, S., Galle, P. R., Cheng, A. L., Ikeda, M., et al. (2018). IMbrave150: A randomized phase III study of 1L atezolizumab plus bevacizumab *vs.* sorafenib in locally advanced or metastatic hepatocellular carcinoma. *J. Clin. Oncol.* 36 (15), TPS4141. doi:10.1200/jco.2018.36.15_suppl.tps4141

Finn, R. S., Ikeda, M., Zhu, A. X., Sung, M. W., Baron, A. D., Kudo, M., et al. (2020). Phase Ib study of lenvatinib plus pembrolizumab in patients with unresectable hepatocellular carcinoma. *J. Clin. Oncol.* 38 (26), 2960–2970. doi:10.1200/jco.20.00808

Fitzsimmons, G. J., and Sadkowsky, K. R. (2002). The Australian Institute of Health and welfare. *Commun. Dis. Intell. Q. Rep.* 26 (4), 605

Gabriele, P., D., Flemmens, M., S., Robertson, J., H., and Hogan, A. *Article and method for focused delivery of therapeutic and/or diagnostic materials*. Us

Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat. Rev. Drug Discov.* 18 (3), 197–218. doi:10.1038/s41573-018-0007-y

Gao, B. Y., Liu, Y., Xia, L. P., Zheng, W. P., and Chen, G. P. (2010). *Intrahepatic arterial infusion of endostatin combined with transcatheter arterial*

chemoembolization for the treatment of advanced hepatocellular carcinoma. Journal of Hainan Medical University.

Greenwald, R. (2008). Progressive counting: A new trauma resolution method. *J. Child. Adolesc. Trauma* 1 (3), 249–262. doi:10.1080/19361520802313619

Harding, J. J., Nandakumar, S., Armenia, J., Khalil, D. N., Albano, M., Ly, M., et al. (2019). Prospective genotyping of hepatocellular carcinoma: Clinical implications of next-generation sequencing for matching patients to targeted and immune therapies. *Clin. Cancer Res.* 25 (7), 2116–2126. doi:10.1158/1078-0432.Ccr-18-2293

He, A. R., Yau, T., Hsu, C., Kang, Y. K., El-Khoueiry, A. B., Santoro, A., et al. (2020). Nivolumab (NIVO) + ipilimumab (IPI) combination therapy in patients (pts) with advanced hepatocellular carcinoma (aHCC): Subgroup analyses from CheckMate 040. *J. Clin. Oncol.* 38 (4), 512. doi:10.1200/jco.2020.38.4_suppl.512

He, M. K., Liang, R. B., Zhao, Y., Xu, Y. J., Shi, M., Zhou, Y. M., et al. (2021a). Lenvatinib, toripalimab, plus hepatic arterial infusion chemotherapy versus lenvatinib alone for advanced hepatocellular carcinoma. *Ther. Adv. Med. Oncol.* 13, 17588359211002720. doi:10.1177/17588359211002720

He, M. K., Ming, S., Lai, Z., and Li, Q. J. (2021b). A phase II trial of lenvatinib plus toripalimab and hepatic arterial infusion chemotherapy as a first-line treatment for advanced hepatocellular carcinoma (LTHAIC study). *J. Clin. Oncol.* 39 (15), 4083. doi:10.1200/jco.2021.39.15_suppl.4083

Hellmann, M., Rizvi, N. A., Goldman, J. W., Gettinger, S. N., Antonia, S. J., Brahmer, J. R., et al. (2016). Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): Results of an open-label, phase 1, multicohort study. *Lancet. Oncol.* 18 (1), 31–41. doi:10.1016/S1470-2045(16)30624-6

Ikeda, M., Sung, M. W., Kudo, M., Kobayashi, M., Okusaka, T., Finn, R. S., et al. (2018). A phase 1b trial of lenvatinib (LEN) plus pembrolizumab (PEM) in patients (pts) with unresectable hepatocellular carcinoma (uHCC). *J. Clin. Oncol.* 36 (15), 4076. doi:10.1200/jco.2018.36.15_suppl.4076

Ishida, Y., Agata, Y., Shibahara, K., and Honjo, T. (1992). Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *Embo J.* 11 (11), 3887–3895. doi:10.1002/j.1460-2075.1992.tb05481.x

Jemal, A., Bray, F., Center, M. M., Ferlay, J., and Lortet-Tieulent, J. (2013). Global cancer statistics, 2012. *Ca. Cancer J. Clin.* 65 (2), 87–108. doi:10.3322/caac.21262

Jiang, R., Tang, J., Chen, Y., Deng, L., Ji, J., Xie, Y., et al. (2017). The long noncoding RNA lnc-EGFR stimulates T-regulatory cells differentiation thus promoting hepatocellular carcinoma immune evasion. *Nat. Commun.* 8, 15129. doi:10.1038/ncomms15129

Jiao, S. C., Bai, L., Dong, J., Bai, C., Xia, Y., Shen, L., et al. (2020a). Clinical activity and safety of penpulimab (Anti-PD-1) with anlotinib as first-line therapy for advanced hepatocellular carcinoma (HCC). *J. Clin. Oncol.* 38 (15), 4592. doi:10.1200/jco.2020.38.15_suppl.4592

Jiao, Y., Liu, M., Luo, N., Guo, H., and Li, J. (2020b). Successful treatment of advanced pulmonary sarcomatoid carcinoma with the PD-1 inhibitor toripalimab: A case report. *Oral Oncol.* 112, 104992. doi:10.1016/j.oraloncology.2020.104992

Jilaveanu, L. B., Shuch, B., Zito, C. R., Parisi, F., Ba Rr, M., Kluger, Y., et al. (2014). PD-L1 expression in clear cell renal cell carcinoma: An analysis of nephrectomy and sites of metastases. *J. Cancer* 5 (3), 166–172. doi:10.7150/jca.8167

Kanikarla Marie, P., Haymaker, C., Parra, E. R., Kim, Y. U., Lazcano, R., Gite, S., et al. (2021). Pilot clinical trial of perioperative durvalumab and tremelimumab in the treatment of resectable colorectal cancer liver metastases. *Clin. Cancer Res.* 27 (11), 3039–3049. doi:10.1158/1078-0432.Ccr-21-0163

Kelley, R. K., Rimassa, L., Cheng, A. L., Kaseb, A., Qin, S., Zhu, A. X., et al. (2022). Cabozantinib plus atezolizumab versus sorafenib for advanced hepatocellular carcinoma (COSMIC-312): A multicentre, open-label, randomised, phase 3 trial. *Lancet. Oncol.* 23 (8), 995–1008. doi:10.1016/s1470-2045(22)00326-6

Klempner, S. J., Fabrizio, D., Bane, S., Reinhart, M., Peoples, T., Ali, S. M., et al. (2020). Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: A review of current evidence. *Oncologist* 25 (1), e147–e159. doi:10.1634/theoncologist.2019-0244

Kotasek, D., Coward, J., Souza, P., Underhill, C., Prawira, A., Li, B., et al. (2019). A phase I dose escalation and dose expansion study of the anti-programmed cell death-1 (PD-1) antibody AK105. *J. Clin. Oncol.* 37 (15), e14006. doi:10.1200/jco.2019.37.15_suppl.e14006

Kudo, M., (2017). Immuno-oncology in hepatocellular carcinoma: 2017 update. *Oncology* 93 (11), 147–159. doi:10.1159/000481245

Kudo, M., Ikeda, M., Motomura, K., Okusaka, T., Kobayashi, M., Dutcus, C. E., et al. (2020). A phase Ib study of lenvatinib (LEN) plus nivolumab (NIV) in patients (pts) with unresectable hepatocellular carcinoma (uHCC): Study 117. *J. Clin. Oncol.* 38 (4), 513. doi:10.1200/jco.2020.38.4_suppl.513

Kudo, M., Motomura, K., Wada, Y., Inaba, Y., and Furuse, J. (2021). Avelumab in combination with axitinib as first-line treatment in patients with advanced hepatocellular carcinoma: Results from the phase 1b VEGF liver 100 trial. *Liver Cancer*10 (3), 249–259. doi:10.1159/000514420

Kudo, M., Ueshima, K., Chan, S., Minami, T., Nishida, N., Aoki, T., et al. (2019). Lenvatinib as an initial treatment in patients with intermediate-stage hepatocellular carcinoma beyond up-to-seven criteria and child–pugh A liver function: A proof-of-concept study. *Cancers* 11 (8), 1084. doi:10.3390/cancers11081084

Lee, A., and Keam, S. J. (2020). Tislelizumab: First approval. *Drugs* 80 (6), 617–624. doi:10.1007/s40265-020-01286-z

Lee, D. J., Lee, H. J., Jr., Farmer, J. R., and Reynolds, K. L. (2021). Mechanisms driving immune-related adverse events in cancer patients treated with immune checkpoint inhibitors. *Curr. Cardiol. Rep.* 23 (8), 98. doi:10.1007/s11886-021-01530-2

Li, J., Xing, J., Yang, Y., Liu, J., Wang, W., Xia, Y., et al. (2020). Adjuvant (131)I-metuximab for hepatocellular carcinoma after liver resection: A randomised, controlled, multicentre, open-label, phase 2 trial. *Lancet. Gastroenterol. Hepatol.* 5 (6), 548–560. doi:10.1016/s2468-1253(19)30422-4

Losic, B., Craig, A. J., Villacorta-Martin, C., Martins-Filho, S. N., Villanueva, A., Chen, X., et al. (2020). Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat. Commun.* 11 (1), 291. doi:10.1038/s41467-019-14050-z

Lowe, M. M., Mold, J. E., Kanwar, B., Huang, Y., Louie, A., Pollastri, M. P., et al. (2014). Identification of cinnabarinic acid as a novel endogenous aryl hydrocarbon receptor ligand that drives IL-22 production. *PLoS One* 9 (2), e87877. doi:10.1371/journal.pone.0087877

Marabelle, A., Fakih, M., Lopez, J., Shah, M., Shapira-Frommer, R., Nakagawa, K., et al. (2020). Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: Prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet. Oncol.* 21 (10), 1353–1365. doi:10.1016/s1470-2045(20)30445-9

Marincola, F. M., Jaffee, E. M., Hicklin, D. J., and Ferrone, S. (2000). Escape of human solid tumors from T-cell recognition: Molecular mechanisms and functional significance. *Adv. Immunol.* 74, 181–273. doi:10.1016/s0065-2776(08)60911-6

Mazdak, M., Ringlstetter, R., Tabrizi, P. F., Akkoyun, M., Tezval, H., Schmitz, J., et al. (2021). Comparison of PD-L1 scores in primary kidney tumors versus accompanying venous tumor thrombi: Retrospective, comparative, monocentric study in treatment-naive patients. *Adv. Ther.* 38 (6), 3373–3388. doi:10.1007/s12325-021-01737-3

Mcglynn, K. A., Petrick, J. L., and London, W. T. (2015). Global epidemiology of hepatocellular carcinoma: An emphasis on demographic and regional variability. *Clin. Liver Dis.* 19 (2), 223–238. doi:10.1016/j.cld.2015.01.001

Mei, K., Qin, S., Chen, Z., Liu, Y., and Zou, J. (2021). Camrelizumab in combination with apatinib in second-line or above therapy for advanced primary liver cancer: cohort A report in a multicenter phase Ib/II trial. *J. Immunother. Cancer* 9 (3), e002191. doi:10.1136/jitc-2020-002191

Mislang, A., Coward, J., Cooper, A., Underhill, C. R., Zheng, Y., Xu, N., et al. (2020). 157P Efficacy and safety of penpulimab (AK105), a new generation anti-programmed cell death-1 (PD-1) antibody, in upper gastrointestinal cancers. *Ann. Oncol.* 31, S1300–S1301. doi:10.1016/j.annonc.2020.10.178

Nguyen, L. N., Nguyen, L. N. T., Zhao, J., Schank, M., Dang, X., Cao, D., et al. (2021). Immune activation induces telomeric DNA damage and promotes short-lived effector T cell differentiation in chronic HCV infection. *Hepatology* 74 (5), 2380–2394. doi:10.1002/hep.32008

Nikfarjam, M., Muralidharan, V., Su, K., Malcontenti-Wilson, C., and Christophi, C. (2005). Patterns of heat shock protein (HSP70) expression and Kupffer cell activity following thermal ablation of liver and colorectal liver metastases. *Int. J. Hyperth.* 21 (4), 319–332. doi:10.1080/02656730500133736

Perussia, B., Dayton, E. T., Fanning, V., Thiagarajan, P., Hoxie, J., and Trinchieri, G. (1984). Immune interferon and leukocyte-conditioned medium induce normal and leukemic myeloid cells to differentiate along the monocytic pathway. *J. Exp. Med.* 158 (6), 2058–2080. doi:10.1084/jem.158.6.2058

Peterson, R., A. (2012). Regulatory T-cells: Diverse phenotypes integral to immune homeostasis and suppression. *Toxicol. Pathol.* 40 (2), 186–204. doi:10.1177/0192623311430693

Pinato, D. J., Mauri, F. A., Spina, P., Cain, O., Siddique, A., Goldin, R., et al. (2019). Clinical implications of heterogeneity in PD-L1 immunohistochemical detection in hepatocellular carcinoma: The blueprint-HCC study. *Br. J. Cancer* 120 (11), 1033–1036. doi:10.1038/s41416-019-0466-x

Presicce, P., Giannelli, S., Taddeo, A., Villa, M. L., and Bella, S. D. (2009). Human defensins activate monocyte-derived dendritic cells, promote the production of proinflammatory cytokines, and up-regulate the surface expression of CD91. *J. Leukoc. Biol.* 86 (4), 941–948. doi:10.1189/jlb.0708412

Qin, S., Chen, Z., Liu, Y., Xiong, J., Zou, J., Meng, Z., et al. (2019). A phase II study of anti–PD-1 antibody camrelizumab plus FOLFOX4 or GEMOX systemic chemotherapy as first-line therapy for advanced hepatocellular carcinoma or biliary tract cancer. *J. Clin. Oncol.* 37 (15), 4074. doi:10.1200/jco.2019.37.15_suppl.4074

Qin, S., Ren, Z., Meng, Z., Chen, Z., Chai, X., Xiong, J., et al. (2020). Camrelizumab in patients with previously treated advanced hepatocellular carcinoma: A multicentre, open-label, parallel-group, randomised, phase 2 trial. *Lancet. Oncol.* 21 (4), 571–580. doi:10.1016/s1470-2045(20)30011-5

Ren, Z., Fan, J., Xu, J., Bai, Y., Xu, A., Cang, S., et al. (2020). LBA2 Sintilimab plus bevacizumab biosimilar *vs.* sorafenib as first-line treatment for advanced hepatocellular carcinoma (ORIENT-32)2 - ScienceDirect. *Ann. Oncol.* 31. doi:10.1016/j.annonc.2020.10.134

Rizzo, A., Ricci, A. D., and Brandi, G. (2021a). Atezolizumab in advanced hepatocellular carcinoma: Good things come to those who wait. *Immunotherapy* 13 (8), 637–644. doi:10.2217/imt-2021-0026

Rizzo, A., Ricci, A. D., and Brandi, G. (2020). Systemic adjuvant treatment in hepatocellular carcinoma: Tempted to do something rather than nothing. *Future Oncol.* 16 (32), 2587–2589. doi:10.2217/fon-2020-0669

Rizzo, A., Ricci, A. D., Gadaleta-Caldarola, G., and Brandi, G. (2021b). First-line immune checkpoint inhibitor-based combinations in unresectable hepatocellular carcinoma: Current management and future challenges. *Expert Rev. Gastroenterol. Hepatol.* 15 (11), 1245–1251. doi:10.1080/17474124.2021.1973431

Saffo, S., and Taddei, T. H. (2019). Systemic management for advanced hepatocellular carcinoma: A review of the molecular pathways of carcinogenesis, current and emerging therapies, and novel treatment strategies. *Dig. Dis. Sci.* 64 (4), 1016–1029. doi:10.1007/s10620-019-05582-x

Sangro, B., Gomez-Martin, C., Manuel, D., I?Arrairaegui, M., Garralda, E., Barrera, P., et al. (2013). A clinical trial of CTLA-4 blockade with tremelimumab in patients with hepatocellular carcinoma and chronic hepatitis C. *J. Hepatol.* 59 (1), 81–88. doi:10.1016/j.jhep.2013.02.022

Scheinberg, D. A., and Pinilla-Ibarz, J. (2006). *Synthetic HLA binding WT-1 peptide analogues and uses thereof*. US.

Scheinberg, D., and Pinilla-Ibarz, J. (2009). *Synthetic HLA binding peptide analogues and uses thereof*. Tampa.

Shan, Y., Zhong, C., Ni, Q., Zhang, M., and Zhou, F. (2021). Anlotinib enhanced penpulimab efficacy through remodeling of tumor vascular architecture and immune microenvironment in hPD-L1/hPD-1 humanized mouse model. *J. Clin. Oncol.* 39 (15), 2581. doi:10.1200/jco.2021.39.15_suppl.2581

Yamazaki, S., and Takayama, T. (2008). Surgical treatment of hepatocellular carcinoma: Evidence-based outcomes. *World J. Gastroenterol.* 14 (5), 685–692. doi:10.3748/wjg.14.685

Shiraha, H., Iwamuro, M., and Okada, H. (2020). Hepatic stellate cells in liver tumor. *Tumor Microenviron.*

Solter, P. F., (2005). Clinical pathology approaches to hepatic injury. *Toxicol. Pathol.* 33 (1), 9–16. doi:10.1080/01926230590522086

Song, Y., Zhu, J., Lin, N., Zhang, C., Xia, Y., Xu, S., et al. (2019). A phase I/II study of the anti-programmed cell death-1 (PD-1) antibody AK105 in patients with relapsed or refractory classic Hodgkin lymphoma (cHL). *J. Clin. Oncol.* 37 (15), e19017. doi:10.1200/jco.2019.37.15_suppl.e19017

Stein, S., Pishvaian, M. J., Lee, M. S., Lee, K. H., Ryoo, B. Y., Kwan, A., et al. (2018). Safety and clinical activity of 1L atezolizumab + bevacizumab in a phase Ib study in hepatocellular carcinoma (HCC). *J. Clin. Oncol.* 36 (15), 4074. doi:10.1200/jco.2018.36.15_suppl.4074

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Szabo, G., Saha, B., and Ambade, A. (2018). *The liver as an immune organ*. Seventh Edition. Zakim and Boyer's Hepatology, 66–76.

Tai, D., Choo, S. P., and Chew, V. (2019). Rationale of immunotherapy in hepatocellular carcinoma and its potential biomarkers. *Cancers (Basel)* 11 (12), E1926. doi:10.3390/cancers11121926

Tarantino, P., Curigliano, G., and Azoulay, D. (2020). Atezolizumab and bevacizumab in hepatocellular carcinoma. *N. Engl. J. Med.* 383 (7), 693–694. doi:10.1056/NEJMc2021840

Thiem, K., Keating, S. T., Netea, M. G., Riksen, N. P., Stienstra, R., van Diepen, J., et al. (2021). Hyperglycemic memory of innate immune cells promotes *in vitro* proinflammatory responses of human monocytes and murine macrophages. *J. Immunol.* 206 (4), 807–813. doi:10.4049/jimmunol.1901348

Tian, L., Goldstein, A., Wang, H., Ching Lo, H., Sun Kim, I., Welte, T., et al. (2017). Mutual regulation of tumour vessel normalization and immunostimulatory reprogramming. *Nature* 544 (7649), 250–254. doi:10.1038/nature21724

Timperi, E., and Barnaba, V. (2020). Viral hepatitides, inflammation and tumour microenvironment. *Adv Exp Med Biol.* 1263, 25–43.

Trinh, V. A., and Hwu, W. J. (2012). Ipilimumab in the treatment of melanoma. *Expert Opin. Biol. Ther.* 12 (6), 773–782. doi:10.1517/14712598.2012.675325

Verhoef, C., Wilt, J., Brunstein, F., Marinelli, A., Etten, B. V., Vermaas, M., et al. (2008). Isolated hypoxic hepatic perfusion with retrograde outflow in patients with irresectable liver metastases; A new simplified technique in isolated hepatic perfusion. *Ann. Surg. Oncol.* 15 (5), 1367–1374. doi:10.1245/s10434-007-9714-z

Vogel, A., Rimassa, L., Sun, H. C., Abou-Alfa, G. K., Merle, P., Pinato, D. J., et al. (2020). Clinical value of atezolizumab + bevacizumab for first-line unresectable hepatocellular carcinoma (HCC): A network meta-analysis. *J. Clin. Oncol.* 38 (15), 4585. doi:10.1200/jco.2020.38.15_suppl.4585

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., et al. (2017). ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell.* 00073, 1806–1821. doi:10.1105/tpc.17.00073

Wainberg, Z. A., Segal, N. H., Jaeger, D., Lee, K. H., Massard, C., Antonia, S. J., et al. (2017). Safety and clinical activity of durvalumab monotherapy in patients with hepatocellular carcinoma (HCC). *J. Clin. Oncol.* 35 (15), 4071. doi:10.1200/jco.2017.35.15_suppl.4071

Wallin, J. J., Bendell, J. C., Funke, R., Sznol, M., Korski, K., Jones, S., et al. (2016). Atezolizumab in combination with bevacizumab enhances antigen-specific T-cell migration in metastatic renal cell carcinoma. *Nat. Commun.* 7, 12624. doi:10.1038/ncomms12624

Wu, C. Y., Budha, N., Gao, Y., Castro, H., Nkobena, A., Ben, Y., et al. (2019). Tislelizumab exposure-response analyses of efficacy and safety in patients with advanced tumors. *Ann. Oncol.* 30, v182–v183. doi:10.1093/annonc/mdz244.044

Xu, J., Shen, J., Gu, S., Zhang, Y., Wang, Q. R., Wu, J., et al. (2020). Camrelizumab in combination with apatinib in patients with advanced hepatocellular carcinoma (rescue): A nonrandomized, open-label, phase II trial. *Clin. Cancer Res.* 27 (4), 1003–1011. doi:10.1158/1078-0432.CCR-20-2571

Xu, J., Zhang, Y., Jia, R., Yue, C., Chang, L., Liu, R., et al. (2019). Anti-PD-1 antibody SHR-1210 combined with apatinib for advanced hepatocellular carcinoma, gastric, or esophagogastric junction cancer: An open-label, dose escalation and expansion study. *Clin. Cancer Res.* 25 (2), 515–523. doi:10.1158/1078-0432.Ccr-18-2484

Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* 377 (25), 2500–2501. doi:10.1056/NEJMc1713444

Yau, T., Kang, Y. K., Kim, T. Y., El-Khoueiry, A. B., Hsu, C., Sangro, B., et al. (2019a). Nivolumab (NIVO) + ipilimumab (IPI) combination therapy in patients (pts) with advanced hepatocellular carcinoma (aHCC): Results from CheckMate 040. *J. Clin. Oncol.* 37 (15), 4012. doi:10.1200/jco.2019.37.15_suppl.4012

Yau, T., Park, J. W., Finn, R. S., Cheng, A. L., Mathurin, P., Edeline, J., et al. (2019b). CheckMate 459: A randomized, multi-center phase III study of nivolumab (NIVO) vs sorafenib (sor) as first-line (1L) treatment in patients (pts) with advanced hepatocellular carcinoma (aHCC). *Ann. Oncol.* 30, v874–v875. doi:10.1093/annonc/mdz394.029

Yau, T., Zagonel, V., Santoro, A., Acosta-Rivera, M., Piscaglia, F., Matilla, A., et al. (2020). Nivolumab (NIVO) + ipilimumab (IPI) + cabozantinib (CABO) combination therapy in patients (pts) with advanced hepatocellular carcinoma (aHCC): Results from CheckMate 040. *J. Clin. Oncol.* 38 (4), 478. doi:10.1200/jco.2020.38.4_suppl.478

Yoon, Y. S., Han, H. S., Cho, J. Y., and Ahn, K. S. (2010). Total laparoscopic liver resection for hepatocellular carcinoma located in all segments of the liver. *Surg. Endosc.* 24 (7), 1630–1637. doi:10.1007/s00464-009-0823-6

Yuan, F., Zhang, W., Di, M., and Gong, J. (2017). Kupffer cells in immune activation and tolerance toward HBV/HCV infection. *Adv. Clin. Exp. Med.* 26 (4), 739–745. doi:10.17219/acem/62759

Zavadil, J., Juráček, J., Čechová, B., Andrašina, T., Slabý, O., and Goldberg, N. (2019). Dynamic changes in circulating MicroRNA levels in liver cancer patients undergoing thermal ablation and transarterial chemoembolization. *Klin. Onkol.* 32 (1), 164

Zhang, H., Hou, X., Cai, H., and Zhuang, X. (2017). Effects of microwave ablation on T-cell subsets and cytokines of patients with hepatocellular carcinoma. *Minim. Invasive Ther. Allied Technol.* 26 (4), 207–211. doi:10.1080/13645706.2017.1286356

Zhang, T., Song, X., Xu, L., Ma, J., Zhang, Y., Gong, W., et al. (2018). The binding of an anti-PD-1 antibody to FcγRI has a profound impact on its biological functions. *Cancer Immunol. Immunother.* 67 (7), 1079–1090. doi:10.1007/s00262-018-2160-x

Zhou, J., Liu, M., Sun, H., Feng, Y., Xu, L., Chan, A., et al. (2017). Hepatoma-intrinsic CCRK inhibition diminishes myeloid-derived suppressor cell immunosuppression and enhances immune-checkpoint blockade efficacy. *Gut* 67, 931–944. doi:10.1136/gutjnl-2017-314032

Zhou, S. L., Zhou, Z. J., Hu, Z. Q., Huang, X. W., Wang, Z., Chen, E. B., et al. (2016). Tumor-associated neutrophils recruit macrophages and T-regulatory cells to promote progression of hepatocellular carcinoma and resistance to sorafenib. *Gastroenterology* 150 (7), 1646–1658. doi:10.1053/j.gastro.2016.02.040

Zhu, A. X., Finn, R. S., Edeline, J., Cattan, S., Ogasawara, S., Palmer, D., et al. (2018). Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): A non-randomised, open-label phase 2 trial. *Lancet. Oncol.* 19 (7), 940–952. doi:10.1016/s1470-2045(18)30351-6

Zhu, X. D., Huang, C., Shen, Y. H., Ji, Y., and Sun, H. C. (2021). Downstaging and resection of initially unresectable hepatocellular carcinoma with tyrosine kinase inhibitor and anti-PD-1 antibody combinations. *Liver Cancer* 10 (4), 320–329. doi:10.1159/000514313

Zhu, Y., Yang, J., Xu, D., Gao, X. M., Zhang, Z., Hsu, J. L., et al. (2019). Disruption of tumour-associated macrophage trafficking by the osteopontin-induced colony-stimulating factor-1 signalling sensitises hepatocellular carcinoma to anti-PD-L1 blockade. *Gut* 68 (9), 1653–1666. doi:10.1136/gutjnl-2019-318419

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 October 2022
DOI 10.3389/fgene.2022.1011390

OPEN ACCESS

EDITED BY
Minglun Li,
LMU Munich University Hospital, Germany

REVIEWED BY
Cheng-Hao Wang,
The Second Affiliated Hospital of Harbin Medical University, China.
Jimin Gao,
Wenzhou Medical University, China
Xu Chen,
Sun Yat-sen Memorial Hospital, China

*CORRESPONDENCE
Kai Yao,
yaokai_sysucc@163.com

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

RECEIVED 04 August 2022
ACCEPTED 21 September 2022
PUBLISHED 04 October 2022

CITATION
Zou Y, Yuan G, Tan X, Luo S, Yang C, Tang Y, Wang Y and Yao K (2022), Immune-related gene risk score predicting the effect of immunotherapy and prognosis in bladder cancer patients.
Front. Genet. 13:1011390.
doi: 10.3389/fgene.2022.1011390

Immune-related gene risk score predicting the effect of immunotherapy and prognosis in bladder cancer patients

Yuantao Zou[1,2,3], Gangjun Yuan[4], Xingliang Tan[1,2,3], Sihao Luo[1,2,3], Cong Yang[1,2,3], Yi Tang[1,2,3], Yanjun Wang[1,2,3] and Kai Yao[1,2,3]*

[1]Department of Urology, Sun Yat-sen University Cancer Center, Guangzhou, China, [2]State Key Laboratory of Oncology in Southern China, Guangzhou, China, [3]Collaborative Innovation Center of Cancer Medicine, Guangzhou, China, [4]Department of Urology Oncological Surgery, Chongqing University Cancer Hospital, Chongqing, China

Background: Immune checkpoint inhibitor therapy has changed the treatment model of metastatic bladder cancer. However, only approximately 20% of patients benefit from this therapy, and robust biomarkers to predict the effect of immunotherapy are still lacking. In this study, we aimed to investigate whether immune-related genes could be indicators for the prognosis of bladder cancer patients and the effect of immunotherapy.

Methods: Based on bladder cancer dataset from the Cancer Genome Atlas (TCGA) and GSE48075, 22 immune microenvironment-related cells were identified by CIBERSORT. After performing a series of bioinformatic and machine learning approaches, we identified distinct tumor microenvironment clusters and three bladder cancer specific immune-related genes (EGFR, OAS1 and MST1R). Then, we constructed immune-related gene risk score (IRGRS) by using the Cox regression method and validated it with the IMvigor210 dataset.

Results: IRGRS-high patients had a worse overall survival than IRGRS-low patients, which was consistent with the result in the IMvigor210 dataset. Comprehensive analysis shows that patients with high IRGRS scores are mainly enriched in basal/squamous type (Ba/Sq), and tumor metabolism-related pathways are more Active, with higher TP53 and RB1 gene mutation rates, lower CD4+/CD8+ T cell infiltration, higher M0 macrophage infiltration, and lower immunotherapy efficacy. In contrast, Patients with low IRGRS scores are mainly enriched in the luminal papillary type (LumP), which is associated

Abbreviations: AUC, Area under curve; BLCA, Bladder cancer; CDF, Cumulative distribution function; CIBERSORT, Cell-type Identification by Estimating Relative Subsets of RNA Transcripts; DEGs, Diferentially expressed genes; GSEA, Gene set enrichment analysis; GSVA, Gene set variation analysis; GO, Gene oncology; GEO, Gene Expression Omnibus; GS, Gene significance; IRGRS, Immune related gene risk score; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, Least absolute shrinkage and selection operator; MM, Module membership; OS, Overall survival; ROC, Area under the curve; ssGSEA, single sample gene set enrichment analysis; TMB, Tumour mutation burden; TCGA, The Cancer Genome Atlas; TPM, Transcripts per million; TOM, Topological overlap matrix; WGCNA, Weighted gene coexpression network analysis.

with the activation of IL-17 and TNF signaling pathways, higher mutation rates of FGFR3 and CDKN1A genes, higher CD4+/CD8+ T cell infiltration content, and The level of M0 macrophage infiltration was relatively low, and the immunotherapy was more probably effective.

Conclusion: Our study constructed an IRGRS for bladder cancer and clarified the immune and molecular characteristics of IRGRS-defined subgroups of bladder cancer to investigate the association between IRGRS and its potential implications for prognosis and immunotherapy.

KEYWORDS

IRGRS, bladder cancer, immunotherapy, microenvironemnt, prognosis

Introduction

Bladder cancer (BLCA) is one of the most prevalent urinary tract malignancies, with an estimated 430,000 new cases and 165,000 deaths worldwide (Lenis et al., 2020). Immunotherapies such as anti-PD-1/PD-L1 inhibitors have demonstrated substantial antitumour activity in advanced and metastatic BLCA, although cisplatin-based chemotherapy and radical cystectomy are still the first-line treatments for muscle-invasive BLCA (Jordan and Meeks, 2019; Patel et al., 2020). However, patients with advanced or metastatic BLCA ineligible for cisplatin only showed an objective remission rate (ORR) of 23%, and the median OS was 15.9 months after receiving the PD-L1 inhibitor atezolizumab as treatment in a phase II trial (Balar et al., 2017). Although some advanced DNA methylation based urinary assay could detect the early stage bladder cancer leading to early treatment, the prognosis of bladder is still unsatisfied (Chen et al., 2020). Besides, how to screen out patients suitable for immunotherapy is still an urgent problem to be solved. At present, the standard biomarkers for clinicians to select patients who are eligible for immunotherapy are immunohistochemistry assays for PD-L1 protein and tumour mutation burden (TMB), but some studies have found conflicting results when using the two biomarkers to predict immunotherapy response or overall survival. Furthermore, many patients whose tumours have low or no detectable PD-L1 expression can also benefit from immunotherapy (Rosenberg et al., 2016). There was no significant association between high TMB and the efficacy of immunotherapy in BLCA (Necchi et al., 2018; Powles et al., 2019). Therefore, it is crucial to develop robust predictive biomarkers to predict the effect of immunotherapy and the prognosis of BLCA patients. Although there have been some studies on the development of molecular markers for the efficacy of immunotherapy (Zhang et al., 2021a; Cao et al., 2021), they did not elucidate the mechanisms behind the molecular markers.

In this study, we analysed three BLCA transcriptomic datasets from patient cohorts (GSE48075, TCGA-BLCA, and IMvigor210). We used the GSE48075 and TCGA-BLCA datasets as training sets to identify the hub genes related to the immune microenvironment. Two computational algorithms, namely, CIBERSORT and ESTIMATE, were used to analyse the expression levels of 22 immune cell types and cancer-related fibroblasts to profile the immune landscape of bladder cancer. Then, we divided patients into different subgroups and examined the correlations of the subgroups with corresponding genomic characteristics and clinical features. Finally, we constructed IRGRS based on the expression of three immune-related genes. The IRGRS was verified to be a robust prognostic biomarker to predict the response to immune checkpoint inhibitors and prognosis.

Materials and methods

Dataset and processing

The Bladder Cancer Dataset from TCGA was used in this study. BLCA samples ($n = 412$) with both RNA sequencing (RNA-seq) data and detailed follow-up information were included for further analysis. RNA-seq data of 270 bladder samples (GSE48075) and corresponding survival information were downloaded from the Gene Expression Omnibus (GEO). IMvigor210 was a cohort in which 195 muscle invasive bladder cancer (MIBC) patients were treated with an anti-PD-L1 agent (atezolizumab) to evaluate the effect of immunotherapy in locally advanced or metastatic urothelial bladder cancer (Mariathasan et al., 2018). Genome, transcriptomic, and clinical data can be downloaded from http://research-pub.gene.com/IMvigor210CoreBiologies. We removed samples whose survival data were not available and then carried out logarithmic processing by the "voom" function of the R package "Limma" (Ritchie et al., 2015). All the RNA-seq datasets in the form of fragments per kilobase of transcript per million mapped reads (FPKM) values were converted into transcripts per million (TPM) to make samples from TCGA and GEO more comparable.

Inference of immune infiltrating cells in the tumour environment

To calculate the composition ratio of 22 tumour-infiltrating immune cells in each cancer sample, CIBERSORT was utilized based on the preset 547 barcode genes of the gene expression

matrix (Newman et al., 2015). CIBERSORT is a deconvolution algorithm to estimate immune cell type (including B cells, T cells, natural killer cells, macrophages, DCs, and myeloid subsets) proportions in data from tumour tissues with mixed cell types.

Unsupervised consensus clustering of 22 tumour-infiltrating immune cells

Unsupervised clustering methods were applied to identify distinct immune patterns and to classify tumour samples for further analysis based on 22 tumour-infiltrating immune cell expression matrices. The R package "ConsensusClusterPlus" was used to perform the above procedure, and 1000 rounds were repeated to guarantee the robustness of classification (Wilkerson and Hayes, 2010). A consensus heatmap was mapped for each sequence of cluster numbers (k = 2, 3, 4, 5, ...), and a progression graph and corresponding cumulative distribution function (CDF) were generated to determine the optimal cluster number.

Identification of differentially expressed genes associated with immune subtypes

We classified patients into four distinct immune patterns by unsupervised consensus clustering to identify immune-related genes. The R package "Limma" was utilized to determine DEGs among the 4 immune subtypes (Ritchie et al., 2015). The criterion for selecting significant DEGs was an adjusted p value < 0.01.

Construction of immune-related genes score

DEGs among all immune clusters were identified, and a union set of genes was extracted. First, we adopted an unsupervised clustering method based on all DEGs to classify patients into several groups for deeper analysis. Then, we defined the optimal number of gene clusters to perform weighted gene coexpression network analysis (WGCNA) to select the related modules of the gene cluster (Langfelder and Horvath, 2008). The "WGCNA" package in R software was used to construct an adjacency matrix with a soft threshold of $\beta = 5$, which was then transformed into a topological overlap matrix (TOM). The corresponding dissimilarity (1-TOM) was calculated as the distance to cluster genes. Then, we built a dynamic pruning tree to identify the related modules. Five modules were identified after setting the merging threshold function at 0.25. Gene significance (GS) and module membership (MM) were calculated for intramodular analysis to select the hub genes. GS is an absolute value to quantify the correlation between a specific gene and its phenotypic trait. MM shows the correlation between the gene and a given module. Hub genes were screened out by setting the cut-off criteria of GS > 0.01 and MM > 0.01. Then, we conducted K–M survival analysis to choose the genes associated with overall survival based on the expression value and clinical data of the hub gene. Then, a univariate Cox regression model was used to perform the prognostic analysis for genes selected after survival analysis. We utilized the least absolute shrinkage and selection operator (LASSO) to precisely predict the outcome of hub genes in BLCA patients. The IRGRS was then constructed by using the coefficients obtained by the LASSO–Cox algorithm, and the IRGRS was calculated by the sum of all gene expression levels multiplied by their corresponding coefficients.

Immune characteristics and molecular biological differences between the high-IRGRS and low-IRGRS groups

According to the median value of the IRGRS in the training dataset, we separated the samples into two groups: the high IRGRS group and the low IRGRS group. To elucidate the underlying biological mechanism in different IRGRS groups, we used gene set enrichment analysis (GSEA), gene ontology (GO), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) method with the clusterProfiler package of R (p < 0.05 and FDR<0.25) (Yu et al., 2012). Then we performed single sample GSEA (ssGSEA) analysis on several representative gene sets with the GSVA (Gene Set Variation Analysis) package of R (Hänzelmann et al., 2013). In addition, to further identify the differences in biological pathways between the high-IRGRS and low-IRGRS groups, GSVA enrichment analysis was conducted by using the "GSVA" package. GSVA is a method based on a nonparametric and unsupervised method to estimate the variation in pathway and biological process activity in the samples. We downloaded the gene sets of "c2.cp. kegg.v6.2.symbols" from the MsigDB database for GSVA. An adjusted p value less than 0.05 was regarded as statistically significant.

Statistical analysis

The statistical significance of the mean value of variables between two groups was estimated by unpaired Student's t tests. Correlation coefficients were computed using Spearman's and distance correlation analyses. Spearman and distance correlation analyses were used to compute the correlation coefficients between each kind of TME infiltrating immune cell. Difference comparisons of three or more groups were conducted by one-way ANOVA and Kruskal–Wallis tests (Robertson et al., 2017). To determine the correlation between the IRGRS and patient survival, we divided patients

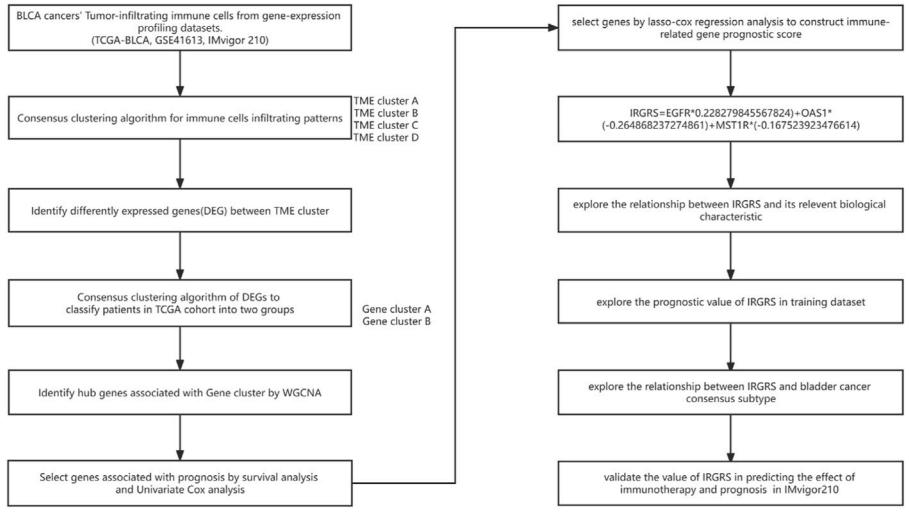


FIGURE 1
Overview of workflow about the study design.

into high- and low-IRGRS groups based on the median IRGRS value in the training group. The Kaplan-Meier method and log-rank tests were utilized to identify the significance of differences in the survival curves for the prognostic analysis. A univariate Cox regression model was adopted to compute the hazard ratios (HRs) in the process of selecting the hub genes. A multivariable Cox regression model was constructed to ascertain the independent prognostic factors. We assessed the specificity and sensitivity of the IRGRS by receiver operating characteristic (ROC) curve analysis and quantified the area under the curve (AUC) with the time ROC package. We used the maftools package to present the mutation landscape in patients with high and low IRGRS subtypes in the TCGA-BLCA cohort. All statistical p values were two-sided, with $p < 0.05$ indicating statistical significance. All data processing was done with R 4.0.2 software.

Results

Landscape of the tumour microenvironment of BLCA

The workflow of how we constructed TME cell-infiltrating patterns and the IRGRS was systematically evaluated (Figure 1). The R package "ConsensusClusterPlus" was used to classify patients with different immune microenvironment patterns based on the amount of 22 tumour-infiltrated immune cells, and four distinct patterns termed TME clusters A, B, C, and D were recognized as the optimal cluster number after we evaluated clustering stability (Supplementary Figure S1C). The 22 tumour-infiltrated immune cell networks portrayed a comprehensive landscape of interactions and their impacts on the overall survival of patients with bladder cancer (Figure 2A; Supplementary Table S1). TME cluster B revealed a particularly prominent survival advantage, and TME cluster C showed the worst prognosis compared with that of the other three TME clusters (log-rank test, $p < 0.01$; Figure 2B). Taken together, we can conclude that crosstalk plays roles among different immune cells in the process of classifying distinct patterns. Then, we visualized the immune microenvironment of the four subtypes in a heatmap (Figure 2C), from which we could see that TME cluster A was characterized by high expression of CD4 memory activated T cells. TME cluster B was characterized by high expression of $CD8^+$ T cells and $CD4^+$ memory activated T cells. TME cluster C was characterized by high expression of resting mast cells. TME cluster D was characterized by high expression of M0 macrophages. A violin plot (Figure 2D) showed that TME cluster B had significantly higher PD-L1 expression than that of the other three TME clusters, and TME cluster C had the lowest PD-L1 expression among the four TME clusters. Except for TME clusters A and D, there were significant differences in the expression of PD-L1 between any two other groups.

Construction of the TME signature and functional annotation

To investigate the potential biological characteristics of each immune subtype, unsupervised analysis of DEGs gathered between each pattern was used to identify optimal genomic

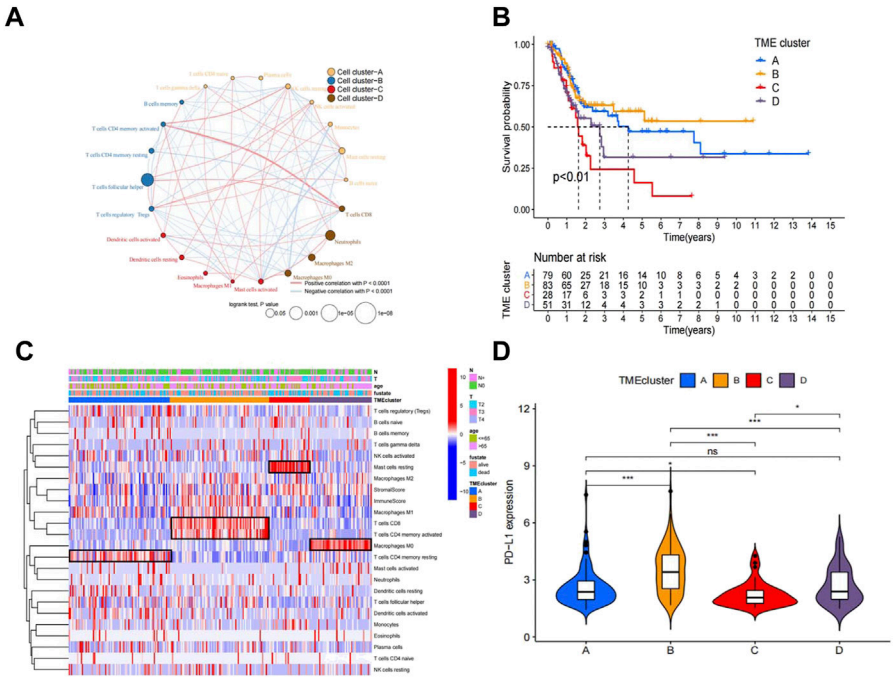


FIGURE 2
Landscape of the TME in bladder cancer and characteristics of TME subtypes. **(A)** The interaction between TME cell type in bladder cancer. Cell cluster A, orange; cell cluster B, blue; cell cluster C, red; cell cluster D, brown.The size of each cell represents survival impact of each TME cell type, calculation used the formula log10 (log-rank test *p* values indicated) respectively. The lines linking TME cells showed their interactions, and thickness represented the correlation estimated by Spearman correlation analysis and strength between regulators. Negative correlation was indicated with blue and positive correlation with red. **(B)**. Kaplan–Meier curves for overall survival (OS) of 241 bladder cancer patients from training cohorts (TCGA + GSE48075) with the TME infifiltration clusters. The numbers of patients in TMEcluster-A, -B, -C,and-D phenotypes are $n = 79$, $n = 83$, $n = 28$, $n = 51$ respectively. Log-rank test shows overall $p < 0.01$. **(C)**. Heatmap of 22 TME cells and ImmuneScore for 249 patients in the training cohort. **(D)**. Comparison of the PD-L1 among four TME subtypes in the training cohort.*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

subtypes. Two gene clusters were recognized as the most suitable method to separate the training cohort population into 2 distant patient clusters (Supplementary Figure S2A), termed gene cluster A and gene cluster B. To obtain the gene cluster-related hub genes, WGCNA was carried out on all genes in gene clusters. Module membership (MM) is an index to measure the correlation between the gene and a given module (Langfelder and Horvath, 2008). Gene significance (GS) represents the correlation between the specific gene and gene cluster. Selected genes and their corresponding modules are shown in a heatmap (Figure 3A). We used a topological overlap matrix (TOM) to cluster all selected genes by dissimilarity measure based on the dynamic tree cut algorithm to divide the tree into five modules (Figure 3B) labelled with different colours. The results showed that the highest positive correlation coefficient between GS for the gene cluster and MM was in the green module (correlation coefficient = 0.77, *p* value = 9e-9), and the lowest negative correlation was in the blue module (correlation coefficient = −0.61, *p* value = 2e-56) (Figure 3B). The criteria for selecting the hub gene were MM > 0.01 and GS > 0.01. Among them, a total of 86 hub genes were identified in the green module (Supplementary Table S2), and 48 hub genes were identified in the blue module (Supplementary Table S3). To determine the independent prognostic genes, univariate Cox regression analysis and K-M survival analysis for OS were performed among the 134 hub genes in the blue and green modules. Twenty-one genes were determined by the selection criteria of Cox *p* value < 0.05 and K-M value < 0.05 (Supplementary Table S4). In order to solve the problem of overfitting of variables, we performed lasso-cox regression to remove 8 genes causing multicollinearity, and obtained 14 genes for subsequent analysis (Figures 3C,D). According to the results of the multivariate Cox hazard model, EGFR ($p < 0.05$), OAS1 ($p < 0.01$), and MST1R ($p < 0.01$) were significantly related to overall survival in BLCA patients (Figure 3E). Then, we constructed a prognostic index for all cancer samples calculated by the formula IRGRS = expression level of EGFR*0.228279845567824 + expression level of OAS1*(-0.264868237274861)+expression level of MST1R*(-0.167523923476614). We used the median IRGRS as the cut-off value, and high-IRGRS patients had a worse OS than low-IRGRS patients ($p < 0.0001$, log-rank test; Figure 3F).

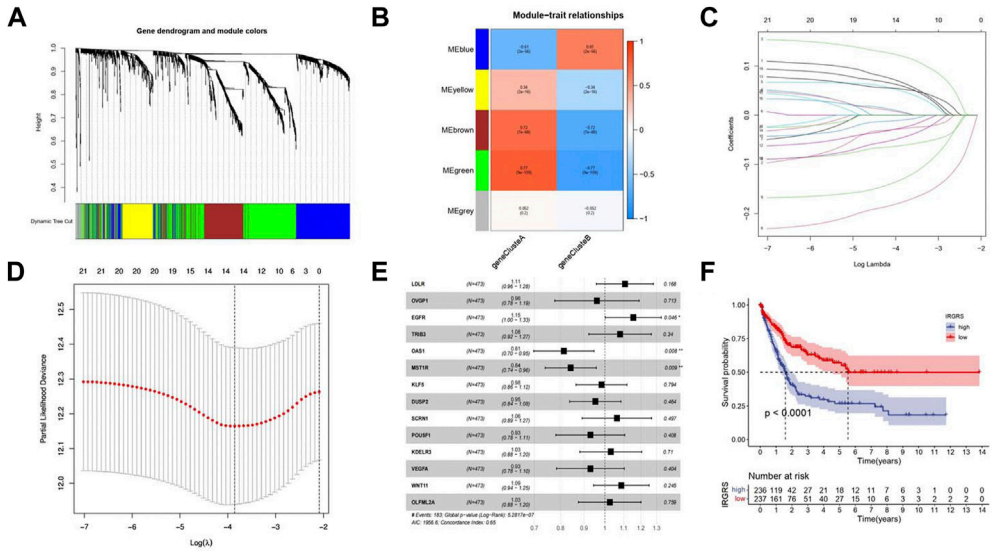


FIGURE 3
Methods about how to filter the hub genes to construct IRGRS system. **(A)**. Clustering dendrograms shows the relationshiop between gene and its corresponding module. **(B)**. Heatmap by WGCNA suggests Module-trait associations. Each row corresponds to a ME and column corresponds to genecluster. The number in the rectangle is the correlation coefficient, and the number in brackets is the corresponding *p* value. **(C)**. Least absolute shrinkage and selection operator (LASSO) coefficient profiles of 21 genes. **(D)**. Partial likelihood deviance for LASSO coefficient profiles. The vertical dotted line is shown at the optimal values, The red dots represent the partial likelihood values, the gray lines represent the standard error (SE). **(E)**. Forest plots of the multivariate Cox hazard model for overall survival. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Unadjusted HRs are shown with 95% confidence intervals. **(F)**. Survival analyses for high and low IRGRS patient groups in training cohort using Kaplan-Meier curves ($p <$ 0.0001, Log-rank test).

Molecular characteristics and functional enrichment analysis in different IRGRS subgroups

Then we investigated somatic mutation differences between the high- and low-IRGRS groups to further elucidate the biological mechanism of IRGRS. The two groups'mutation landscapes are depicted in Figure 4A and Figure 4B. We listed the top 20 genes with the highest mutation rates in the IRGRS subgroups and found that missense variations were the most common mutation type in both types. The mutation rates of TTN, TP53, MUC16, ARID1A, PIK3CA, KMT2D, NOTCH1, and SYNE1 were higher than 15% in both groups. The mutation rate of the TP53 and RB1 genes in the high-IRGRS subgroup was higher than that in the low-IRGRS subgroup. We determined the DEGs to further investigate the underlying biological behaviour of IRGRS by using the limma package (Ritchie et al., 2015), GO, KEGG, GSEA, GSVA, and ssGSEA were performed by The clusterProfiler package (Yu et al., 2012) and GSVA package (Huang et al., 2021) for the DEGs between the high-IRGRS group and the low-IRGRS group. We conducted GO and KEGG pathway enrichment analyses to explore the functional characteristics of the DEGs. In the GO functional enrichment analysis, the top 10 enriched biological processes were "extracellular matrix organization", "extracellular structure organization", "axonogenesis', "connective tissue development", "ossification", "cell–substrate adhesion", "skeletal system morphogenesis", "cartilage development", "neuron projection guidance" and "sulfur compound metabolic process" (Figure 4C). Based on DEGs from the IRGRS, we performed GSVA to explore the biological behaviour differences between the IRGRS subgroups. We found that the low IRGRS subgroup was markedly enriched in drug metabolism, steroid hormone biosynthesis, and retinol metabolism. The high-IRGRS subgroup presented enrichment in the cell cycle, gap junctions and regulation of the actin cytoskeleton (Figure 4D).

We performed GSEA to identify the corresponding gene sets enriched in different IRGRS subgroups. The top five significantly enriched pathways in the high- and low-IRGRS groups are shown in Figures 5A,B. Genes in the low IRGRS groups were mostly enriched in "cell cycle", "ECM-receptor interaction", "IL-17 signalling pathway", "protein digestion and absorption" and "TNF signalling pathway". These factors are tightly associated with the immune response. Genes in the high-IRGRS group were mostly enriched in pathways related to chemical carcinogenesis and metabolism. Detailed results of the GSEA are listed in Supplementary Table S5. To evaluate how IRGRS reflects the cell type in the tumour immune microenvironment, the ESTIMATE and CIBERSORT algorithms were applied to

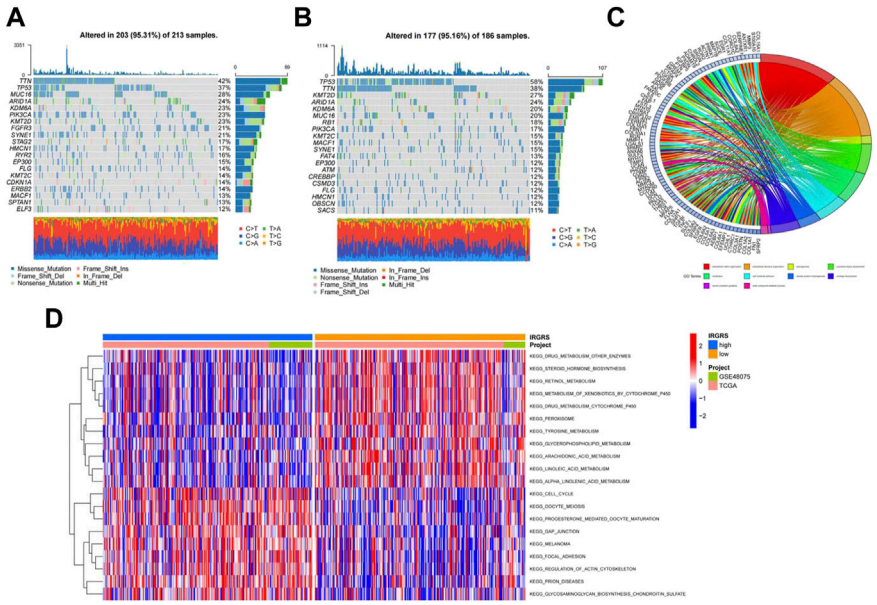


FIGURE 4
Molecular characteristic and functional enrichment analysis between high and low IRGRS groups. **(A)**. The waterfall plot of tumor somatic mutation established by those with low IRGRS groups, Mutated genes (rows, top 20) are ordered by mutation rate; Each column represented individual patients. The upper barplot showed the overall number of mutations. The right barplot showed the percentage of each variant type and the mutation frequency of each gene. The color coding indicates the mutation type. **(B)**. The waterfall plot of tumor somatic mutation established by those with high IRGRS group. **(C)**. The GO terms are defined as indicated color bars at the bottom and shown on the right of chord diagram, the involved genes are listed on the left. The genes associated ten significant signaling pathways. **(D)**. Heatmap by GSVA analysis between high and low IRGRS group.The upper barplot showed the IRGRS defined subgroups (high-IRGRS and low-IRGRS) and the origin of dataset (TCGA and IMvigor210). The rows of the heatmap showed the activation of corresponding pathways.

compute the infiltration of immune cells in BLCA. Differentially infiltrated cells between the low- and high-IRGRS groups are presented in Figures 5C–F. We observed that the high-IRGRS group had lower levels of immune cells (including $CD8^+$ T cell infiltration, $CD4^+$ memory-activated T cells, and follicular helper T cells). Conversely, the level of M0 macrophages was higher in the high-IRGRS group than that in the low-IRGRS group. In the ssGSEA analysis(Figure 5G), many immune-related cells(including activated B cells, activited $CD4^+$ T cells, immature B cell and so on) showed a higher amount in high-IRGRS group than that in low-IRGRS group. KEGG pathway analysis revealed the significant pathways between high and low IRGRS groups (Figure 5H).

Relationship between IRGRS grouping and other immune and molecular subtypes

A consensus molecular classification subtype can describe the landscape of bladder cancer according to the RNA-sequence data and can be summarized as six molecular subtypes, namely, Ba/Sq, LumNS, LumP, LumU, stroma-rich, and NE-like (Kamoun et al., 2020), which is a classification system based on six published classification systems. Then, we focused on the distribution of different molecular subtypes in the IRGRS groups. Because the number of NE-like samples was below 10, we did not include it in our analysis. In our study, the low-IRGRS subgroup comprised 26% Ba/Sq samples, 2% LumNS samples, 48% LumP samples, 14% LumU samples, and 9% stroma-rich subtype samples, while the high-IRGRS subgroup comprised 51% Ba/Sq samples, 8% LumNS samples, 16% LumP samples, 12% LumU samples, and 13% stroma-rich subtype samples (Figure 6A). We found that the Ba/Sq and LumP subtypes accounted for a large proportion of all samples from the TCGA database. There were more LumP samples in the low-IRGRS subgroup than in the high-IRGRS subgroup ($p < 0.01$). The violin plot (Figure 6B) shows the different molecular classifications and their corresponding IRGRS. The Ba/Sq subgroup was markedly associated with a higher IRGRS than the LumP and LumU subtypes. The LumP subgroup was associated with a lower IRGRS than the LumU and stroma-rich subtypes. Several genes (such as FGFR3, TP53, and RB1) have been identified as being vital for the characterization of each consensus class (Kamoun et al., 2020). Therefore, we analysed the relationship between the IRGRS and the mutation status of the three genes mentioned (Figures 6C–E). The IRGRS of p53-mutated samples was higher than that of p53 wild-type samples ($p < 0.01$). Conversely, in the FGFR3 gene, FGFR3-mutated samples had lower IRGRS values than FGFR3 wild-type

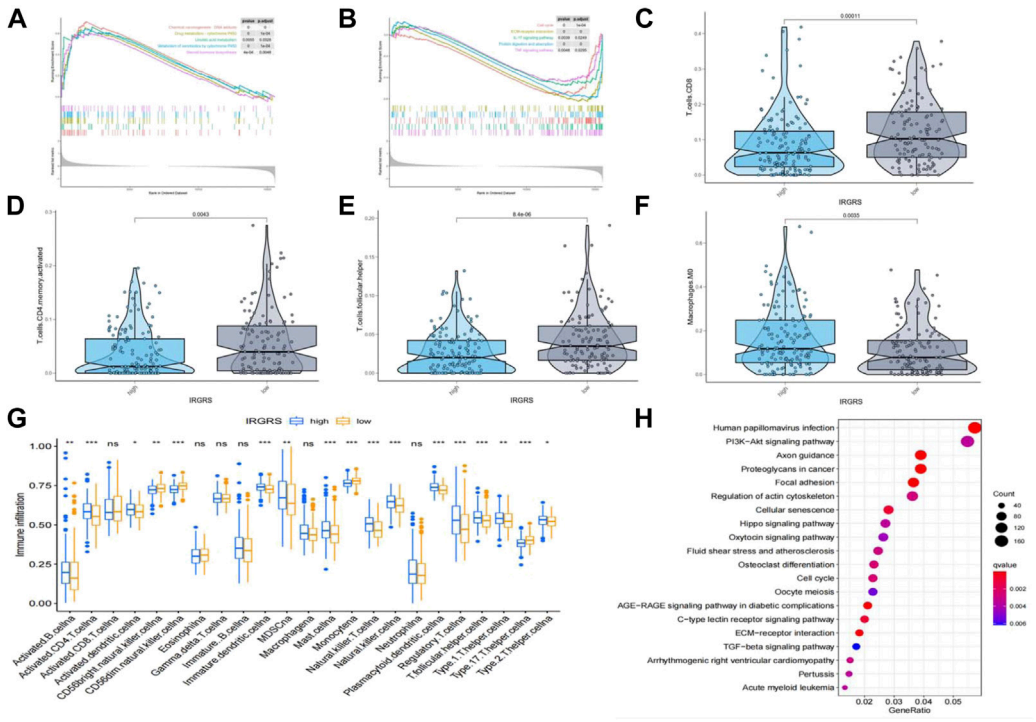


FIGURE 5
Difference in tumor infiltrated immune cells between high and low IRGRS groups. **(A)**. high score pathways enriched by GSEA analysis between high and low IRGRS groups. **(B)**. low score pathways enriched by GSEA analysis between high and low IRGRS groups. **(C–F)**. different kinds of T cells in tumor environment expressed diffrently between high and low IRGRS groups. **(G)**. tumor infiltrated immune cells analysed by ssGSEA analysis between high and low IRGRS groups. **(H)**. KEGG analysis of DEGs between high and low IRGRS groups.

samples. However, there was no significant difference in IRGRS levels between RB1-mutated samples and wild-type samples. We evaluated the survival prediction ability of IRGRS in TCGA-BLCA datasets by using time-dependent ROC analysis. We found that the AUCs for the IRGRS were 0.66, 0.71, and 0.69 at 1, 3 and 5 years, respectively (Figure 6F).

The benefit of immunotherapy in different IRGRS subgroups

We then explored the potential clinical efficacy of IRGRS in predicting the effect of immunotherapy based on the IMvigor210 dataset. All samples were classified into immune-desert, immune-excluded and immune-inflamed phenotypes. The immune-desert phenotype was characterized by the suppression of immunity. The immune-excluded phenotype was characterized by innate immune cell infiltration and stromal activation. The immune-inflamed phenotype was characterized by adaptive immune cell infiltration and immune activation. In our results, the immune-excluded phenotype had a higher IRGRS than those of the other two subgroups, implying that high-IRGRS patients could benefit less from immunotherapy than low-IRGRS patients (Figure 7A). Then, we evaluated whether the IRGRS could predict patients' response to immune checkpoint blockade therapy based on IMvigor210 cohorts. Survival analysis (Figure 7B) showed that high-IRGRS patients had worse OS than low-IRGRS patients, which was consistent with the results of the training datasets. We included Ba/Sq, LumNS, LumP, LumU four subgroups into our survival analysis. Patients with a low IRGRS exhibited a greater clinical response to anti-PD-1/L1 immunotherapy than those with a high IRGRS (Figure 7C). We could find from Figure 7D that there were more Ba/Sq samples and fewer Lump samples in the high-IRGRS subgroup than in the low-IRGRS subgroup ($p < 0.001$, x2 test). The result of which was consistent with training dataset from TCGA + GSE48075. Given that the immune cell(IC) level, tumor cell(TC) level, immune phenotype, consensus subtype had been shown to be highly predictive of the response to immune therapy (Tsao et al., 2018; Kamoun et al., 2020; Hornburg et al., 2021), we speculated that they might function as synergistic factors in predicting the response to immunotherapy. Therefore, a nomogram was developed to include all factors above to offer a quantitative approach for predicting the effect of immunotherapy. The nomogram was constructed in the IMvigor210 cohort and the corresponding

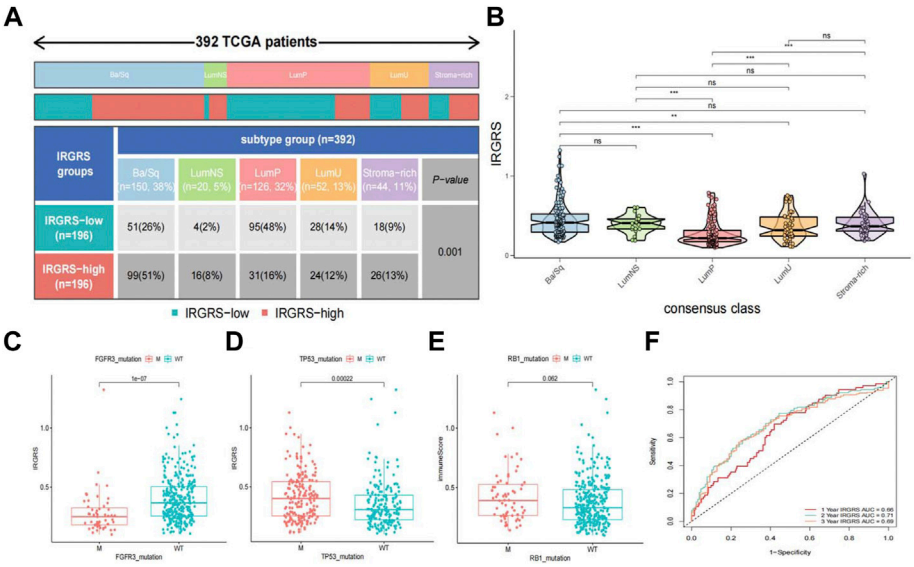


FIGURE 6
Relationship between molecular classification and IRGRS **(A)**. Heatmap and table showing the distribution of bladder cancer consensus molecular subtypes between the IRGRS and subgroups in TCGA dataset. **(B)**. Five bladder cancer molecular subtypes and their corresponding IRGRS. **(C)**. Comparison of IRGRS between the FGFR3 mutated groups and FGFR3 wild type groups.M = mutated, WT = wild type. **(D)**. Comparison of IRGRS between the TP53 mutated groups and TP53 wild type groups. **(E)** Comparison of IRGRS between the RB1 mutated groups and RB1 wild type groups. **(F)**. Time dependent ROC curve analysis of survival prediction by the IRGRS.

calibration curve was constructed (Figure 7E). To find a relationship between IRGRS and bladder cancer immune landscape, we portray the IRGRS-defined subgroups to contrast the IRGRS-defined subgroups with IC, TC and immunotherapy responsein (Figure 7F). Consistent with the importance of TMB, we observed that the low-IRGRS subtype was significantly enriched for response of immunotherapy.

Discussion

Increasing evidence suggests that the tumour immune microenvironment plays an important role in innate immunity as well as antitumour effects through interactions between immune cells and tumour cells (Binnewies et al., 2018). Based on the mechanism of immune evasion, immunotherapy has proven to be effective for patients unsuitable or recurrent after cisplatin-based treatment. However, only a few patients can benefit from immunotherapy (Ott et al., 2020). Biomarkers including PD-1 expression, PD-L1 expression, tumour mutation burden and MSI status are not efficient for predicting the benefits of immune checkpoint blockade (Subrahmanyam et al., 2018; Ganesh et al., 2019; Jardim et al., 2021). In addition, the clinical prognosis heterogeneity of BLCA reveals that immune-relevant subtypes may exist between BLCA samples in the same clinical stage. This situation highlights the urgent need to develop a robust biomarker and subgroup analysis for guiding immunotherapy in BLCA.

In our study, based on 22 tumour-infiltrated immune cell lines, we identified four distinct tumour microenvironment patterns. These four patterns had significantly different tumour-related immune cell characteristics. Cluster A was characterized by a low expression level of $CD8^+$ T cells and a high level of resting memory $CD4^+$ T cells. In contrast, cluster B displayed more $CD8^+$ T cells and less resting memory $CD4^+$ T cells. Cluster C showed a higher resting mast cell quantity than those of the other clusters. Cluster D was characterized by a high level of M0 macrophages. Each TME cluster showed unique features with respect to the tumour-infiltrated immune microenvironment. In many previous studies, only the results from the transcriptome profile and enriched pathways associated with immunity were considered. However, in our study, to identify the underlying mechanism and hub genes connected with the TME clusters, we conducted several computational algorithms to construct an IRGRS system. The IRGRS is proven to be a robust biomarker for guiding the immunotherapy of bladder cancer, with better survival in low-IRGRS patients and worse survival in high-IRGRS patients in both training and validation cohorts.

The IRGRS consists of three genes: EGFR, OAS1, and MST1R. Epidermal growth factor receptor (EGFR) is widely recognized because it is of great importance in many kinds of

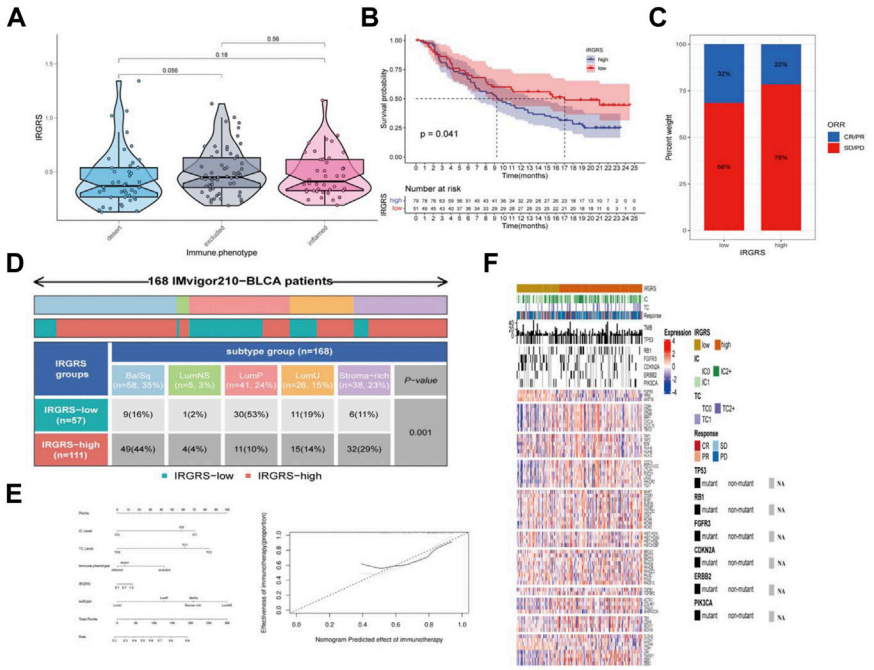


FIGURE 7
The role of IRGRS in predicting the effect of immunotherapy of bladder cancer. **(A)**. Comparison of the IRGRS of different immune phenotype in bladder cancer. **(B)**. Survival analysis of immunotherapy gene set in diferent IRGRS groups (high-IRGRS and low-IRGRS). **(C)**. The proportions of clinical response (CR/PR, SD/PD) after accepting immunotherapy in the high-IRGRS and low-IRGRS groups in IMgivor210 dataset **(D)**. Heatmap and table showing the distribution of bladder cancer consensus molecular subtypes between the IRGRS and subgroups in IMgivor210 dataset. **(E)** Nomogram and corresponding calibration curve for predicting survival probability in the validation cohort. **(F)**. Heatmap representing evaluated patients first sorted based on a IRGRS-based subtyping scheme, Immune cell and tumour cell PD-L1 status are given. Then by response to atezolizumab. In addition, TMB and mutation status (black, mutated; grey, patients without mutation data) for a few genes of interest are shown. The rows of the heatmap show expression (Z scores) of genes of interest, grouped into the biologies or pathways.

cancers (Ganesh et al., 2019; Zeng et al., 2020). Mutations and amplification in its exon region have been identified to be driving events in many cancer types. The protein encoded by EGFR is a receptor for members of the epidermal growth factor family. Many research and drug development efforts have been prompted by its role in non-small-cell lung cancer (Harrison et al., 2020), basal-like breast cancers (Gonzalez-Conchas et al., 2018) and glioblastoma (Eskilsson et al., 2018). Tyrosine kinase inhibitors such as gefitinib and erlotinib have shown efficacy in tumours with EGFR exon amplification. However, some studies revealed that patients diagnosed with EGFR-mutated non-small-cell lung cancer could draw limited benefit from immunotherapy (Proto et al., 2019). These results suggest that the EGFR gene is a vital factor influencing whether immunotherapy can exert a positive effect in patients. In addition, EGFR has been identified as an oncogenetic mechanism in the basal/squamous (Ba/Sq) subtype among the six molecular classification subtypes (Kamoun et al., 2020). Oligodenylated synthase 1 (OAS1) is a protein encoded by OAS1 that results in RNA degradation and the inhibition of viral replication; it has been included in several prognostic signatures and has been found to be a robust biomarker to predict the effect of immunotherapy (Luo et al., 2020; Jin et al., 2021). Macrophage stimulating 1 receptor (MST1R) is a gene that encodes a cell surface receptor for macrophage-stimulating protein with tyrosine kinase activity. Studies have found that suppression of MST1R expression results in reduced pancreatic tumour size, changes in macrophage polarization and enhanced T-cell infiltration (Braun et al., 2018; Tan et al., 2019).

Then, we studied the gene mutations of different IRGRS subgroups to uncover the underlying immunologic mechanism. The most common gene mutations in both the high-IRGRS and low-IRGRS samples were missense variations. However, for some other mutation types, such as nonsense mutations and frameshift mutations, there was quite a difference between the different IRGRS groups. TP53 mutation was the most differentially expressed gene in the top 20 mutated genes between high-IRGRS and low-IRGRS samples. TP53 mutation is one the most common mutation types in many kinds of cancer and can lead to poor

outcomes (Vousden and Prives, 2005; Olivier et al., 2006). TP53 can regulate the p53/TGFβ signalling pathway, which has an influence on tumour cell proliferation by the cell cycle. In addition, there was a higher rate of RB1 mutation in the high-IRGRS subgroup than in the low-IRGRS subgroup. RB1 was the first tumour suppressor gene found, and the protein encoded by RB1 is a negative regulator of the cell cycle. Therefore, high-IRGRS patients with high TP53 and RB1 mutation burdens have a worse outcome than low-IRGRS patients with low TP53 and RB1 mutation burdens, in agreement with our survival results. GO, GSVA, and GSEA analyses between the high and low IRGRS groups suggest that apart from activated immune-related pathways, there are also many other mechanisms in the tumour immune microenvironment.

Several molecular classifications have been reported in the development of a more precise patient stratification (Choi et al., 2014; Robertson et al., 2018; Tan et al., 2019). However, even the consensus classification system of subtypes has not translated universally into clinical trials or clinical applications (Kamoun et al., 2020). Thus, we analysed the association between the IRGRS and the consensus classification system (LumP, LumU, stroma-rich, LumNS, and Ba/Sq). Each consensus class has distinct differentiation patterns, oncogenic mechanisms, tumour microenvironments, and histological and clinical associations.For example, the tumor driving mechanism of the Lump subtype is mainly related to the overexpression of FGFR3, and the Ba/Sq subtype is mainly related to the overexpression of EGFR.In addition, the mutation spectrum of different molecular subtypes is also different. For example, the mutation rate of the RB1 gene in the Ba/Sq subtype is significantly higher than that of other subtypes, and the KDM6A gene has the highest mutation rate in the Lump subtype. We found that more than half of the high-IRGRS samples were distributed in the Ba/Sq classification, and nearly half of the low-IRGRS samples were enriched in the LumP classification. The Ba/Sq subtype was identified to be more sensitive to immunotherapy than the other subtypes (Kamoun et al., 2020), which was consistent with our results. FGFR3 mutation has been recognized as one of the oncogenic mechanisms in the development of the LumP subtype of MIBC (Robertson et al., 2017). We also revealed that the IRGRS correlated with FGFR3 mutation in our study. FGFR3-targeted therapy may be an encouraging choice for low-IRGRS tumours, especially in the LumP subtype of MIBC. Molecular classification of bladder cancer showed tumour biological heterogeneity, which could provide an innovative approach to improving therapeutic effectiveness. When we combined the IRGRS with a molecular classification system, we could classify MIBC subgroups and guide personalized antitumour therapies more precisely. Prospective clinical trials need to be performed to certify the therapy-related predictive value of the IRGRS, and certain therapies need more investigation through *in vitro* or *in vivo* experiments.

Then, we confirmed the effect of the IRGRS in predicting the efficacy of immunotherapy based on the IMvigor210 dataset. We found different immune microenvironment-related cells between the high and low IRGRS groups, which might partly explain the different responses to immunotherapy between the two groups. Integration with 3 immune-related subtypes (immune desert, immune exclusion, and immune inflamed) allowed IRGRS grouping to distinguish different immune subtypes of BLCA. Unfortunately, there were no significant differences between the immune-desert and immune-excluded groups, which may be because the number of samples in the IMvigor210 dataset was not large enough. It has been recognized that the effective rate of immunotherapy for PD-L1 positive bladder cancer patients is only about 20%, which suggests the limitation of PD-L1 as an indicator. While in patients with low IRGRS, the effectiveness of immunotherapy can reach 32%. This further demonstrates the superiority of the IRGRS. More importantly, our study has developed several new insights for bladder cancer immunotherapy that target the IRGRS phenotype and immune phenotype. By combining the IRGRS and molecular classification, we might select patients who are suitable for immunotherapy more accurately. Further reversing the adverse TME cell infiltration may contribute to exploiting the development of novel drug combination strategies or novel immunotherapeutic agents in the future. Moreover, the patients in the high-IRGRS group had a shorter follow-up time than those in the low-IRGRS group. Several limitations of this study should be considered. First, Recent studies suggested that OICR-9429 and HSF1 played important roles in regulating the tumor microenvironment of bladder cancer. They conducted in-depth study about how the two genes work (Zhang et al., 2021b; Huang et al., 2022). Although we have reviewed the roles of the three genes that construct IRGRS in tumors, the underlying molecular mechanisms require further exploration of *in vivo* and *in vitro* functional experiments. Second, our study is a bioinformatics analysis based on public databases and lacks validation of independent clinical cohorts.

Conclusion

By applying a series of bioinformatics methods, we constructed IRGRS that could accurately predict the effect of immunotherapy and prognosis in bladder cancer. In addition, when we conbined IRGRS with bladder cancer consensus classification system, we could improve the robustness of prediction. However, further prospective clinical studies are needed to verify the absoluteness of our conclusion.

Data availability statement

All the data used in this study are from public datasets (TCGA, GEO, IMvigor210) and can be accessed without restriction.

Author contributions

YZ performed the bioinformatics analysis and wrote the manuscript. XT and YW conducted the statistical analysis. SL, CY and YT were responsible for the data processing. GY and KY designed and supported the study.All authors read and approved the final manuscript

Funding

This study was supported by the Fundamental Research Funds for the Central Universities (No. 19ykpy178 to KY), the Natural Science Foundation of Guangdong Province (No. 2019A1515010197 to KY), the Sun Yat-sen University Cancer Center Medical Scientist Training Program (No. 14zxqk08 to KY), the Natural Science Foundation of Chongqing (No. cstc2019jcyjmsxmX0420 to GY).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1011390/full#supplementary-material

References

Balar, A. V., Galsky, M. D., Rosenberg, J. E., Powles, T., Petrylak, D. P., Bellmunt, J., et al. (2017). Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: A single-arm, multicentre, phase 2 trial. *Lancet (London, Engl.* 389, 67–76. doi:10.1016/S0140-6736(16)32455-2

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24, 541–550. doi:10.1038/s41591-018-0014-x

Braun, S., Enculescu, M., Setty, S. T., Cortés-López, M., De Almeida, B. P., Sutandy, F. X. R., et al. (2018). Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. *Nat. Commun.* 9, 3315. doi:10.1038/s41467-018-05748-7

Cao, R., Yuan, L., Ma, B., Wang, G., and Tian, Y. (2021). Tumour microenvironment (TME) characterization identified prognosis and immunotherapy response in muscle-invasive bladder cancer (MIBC). *Cancer Immunol. Immunother.* 70, 1–18. doi:10.1007/s00262-020-02649-x

Chen, X., Zhang, J., Ruan, W., Huang, M., Wang, C., Wang, H., et al. (2020). Urine DNA methylation assay enables early detection and recurrence monitoring for bladder cancer. *J. Clin. Invest.* 130, 6278–6289. doi:10.1172/JCI139597

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E. R., Hoffman-Censits, J., et al. (2014). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 25, 152–165. doi:10.1016/j.ccr.2014.01.009

Eskilsson, E., Røsland, G. V., Solecki, G., Wang, Q., Harter, P. N., Graziani, G., et al. (2018). EGFR heterogeneity and implications for therapeutic intervention in glioblastoma. *Neuro. Oncol.* 20, 743–752. doi:10.1093/neuonc/nox191

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in colorectal cancer: Rationale, challenges and potential. *Nat. Rev. Gastroenterol. Hepatol.* 16, 361–375. doi:10.1038/s41575-019-0126-x

Gonzalez-Conchas, G. A., Rodriguez-Romo, L., Hernandez-Barajas, D., Gonzalez-Guerrero, J. F., Rodriguez-Fernandez, I. A., Verdines-Perez, A., et al. (2018). Epidermal growth factor receptor overexpression and outcomes in early breast cancer: A systematic review and a meta-analysis. *Cancer Treat. Rev.* 62, 1–8. doi:10.1016/j.ctrv.2017.10.008

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Harrison, P. T., Vyse, S., and Huang, P. H. (2020). Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. *Semin. Cancer Biol.* 61, 167–179. doi:10.1016/j.semcancer.2019.09.015

Hornburg, M., Desbois, M., Lu, S., Guan, Y., Lo, A. A., Kaufman, S., et al. (2021). Single-cell dissection of cellular components and interactions shaping the tumor immune phenotypes in ovarian cancer. *Cancer Cell* 39, 928–944.e6. e6. doi:10.1016/j.ccell.2021.04.004

Huang, M., Dong, W., Xie, R., Wu, J., Su, Q., Li, W., et al. (2022). HSF1 facilitates the multistep process of lymphatic metastasis in bladder cancer via a novel PRMT5-WDR5-dependent transcriptional program. *Cancer Commun. Lond. Engl.* 42, 447–470. doi:10.1002/cac2.12284

Huang, R., Wang, S., Zhu, R., Xian, S., Huang, Z., Cheng, L., et al. (2021). Identification of key eRNAs for spinal cord injury by integrated multinomial bioinformatics analysis. *Front. Cell Dev. Biol.* 9, 728242. doi:10.3389/fcell.2021.728242

Jardim, D. L., Goodman, A., De Melo Gagliato, D., and Kurzrock, R. (2021). The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* 39, 154–173. doi:10.1016/j.ccell.2020.10.001

Jin, K., Qiu, S., Jin, D., Zhou, X., Zheng, X., Li, J., et al. (2021). Development of prognostic signature based on immune-related genes in muscle-invasive bladder cancer: Bioinformatics analysis of TCGA database. *Aging* 13, 1859–1871. doi:10.18632/aging.103787

Jordan, B., and Meeks, J. J. (2019). T1 bladder cancer: Current considerations for diagnosis and management. *Nat. Rev. Urol.* 16, 23–34. doi:10.1038/s41585-018-0105-y

Kamoun, A., de Reyniès, A., Allory, Y., Sjödahl, G., Robertson, A. G., Seiler, R., et al. (2020). A consensus molecular classification of muscle-invasive bladder cancer. *Eur. Urol.* 77, 420–433. doi:10.1016/j.eururo.2019.09.006

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Lenis, A. T., Lec, P. M., Chamie, K., and Mshs, M. D. (2020). Bladder cancer: A review. *Jama* 324, 1980–1991. doi:10.1001/jama.2020.17598

Luo, Y., Chen, L., Zhou, Q., Xiong, Y., Wang, G., Liu, X., et al. (2020). Identification of a prognostic gene signature based on an immunogenomic landscape analysis of bladder cancer. *J. Cell. Mol. Med.* 24, 13370–13382. doi:10.1111/jcmm.15960

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554, 544–548. doi:10.1038/nature25501

Necchi, A., Anichini, A., Raggi, D., Briganti, A., Massa, S., Lucianò, R., et al. (2018). Pembrolizumab as neoadjuvant therapy before radical cystectomy in patients with muscle-invasive urothelial bladder carcinoma (PURE-01): An open-label, single-arm, phase II study. *J. Clin. Oncol.* 36, 3353–3360. doi:10.1200/JCO.18.01148

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Olivier, M., Langerød, A., Carrieri, P., Bergh, J., Klaar, S., Eyfjord, J., et al. (2006). The clinical value of somatic TP53 gene mutations in 1, 794 patients with breast cancer. *Clin. Cancer Res.* 12, 1157–1167. doi:10.1158/1078-0432.CCR-05-1029

Ott, P. A., Hu-Lieskovan, S., Chmielowski, B., Govindan, R., Naing, A., Bhardwaj, N., et al. (2020). A phase Ib trial of personalized neoantigen therapy plus anti-PD-1 in patients with advanced melanoma, non-small cell lung cancer, or bladder cancer. *Cell* 183, 347–362. doi:10.1016/j.cell.2020.08.053

Patel, V. G., Oh, W. K., and Galsky, M. D. (2020). Treatment of muscle-invasive and advanced bladder cancer in 2020. *Ca. Cancer J. Clin.* 70, 404–423. doi:10.3322/caac.21631

Powles, T., Kockx, M., Rodriguez-Vida, A., Duran, I., Crabb, S. J., Van Der Heijden, M. S., et al. (2019). Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. *Nat. Med.* 25, 1706–1714. doi:10.1038/s41591-019-0628-7

Proto, C., Ferrara, R., Signorelli, D., Lo Russo, G., Galli, G., Imbimbo, M., et al. (2019). Choosing wisely first line immunotherapy in non-small cell lung cancer (NSCLC): What to add and what to leave out. *Cancer Treat. Rev.* 75, 39–51. doi:10.1016/j.ctrv.2019.03.004

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

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., et al. (2017). Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 171, 540–556. e25. doi:10.1016/j.cell.2017.09.007

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., et al. (2018). Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 174, 1033. doi:10.1016/j.cell.2018.07.036

Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., et al. (2016). Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: A single-arm, multicentre, phase 2 trial. *Lancet (London, Engl.* 387, 1909–1920. doi:10.1016/S0140-6736(16)00561-4

Subrahmanyam, P. B., Dong, Z., Gusenleitner, D., Giobbie-Hurder, A., Severgnini, M., Zhou, J., et al. (2018). Distinct predictive biomarker candidates for response to anti-CTLA-4 and anti-PD-1 immunotherapy in melanoma patients. *J. Immunother. Cancer* 6, 18. doi:10.1186/s40425-018-0328-8

Tan, T. Z., Rouanne, M., Tan, K. T., Huang, R. Y., and Thiery, J. P. (2019). Molecular subtypes of urothelial bladder cancer: Results from a meta-cohort analysis of 2411 tumors. *Eur. Urol.* 75, 423–432. doi:10.1016/j.eururo.2018.08.027

Tsao, M. S., Kerr, K. M., Kockx, M., Beasley, M. B., Borczuk, A. C., Botling, J., et al. (2018). PD-L1 immunohistochemistry comparability study in real-life clinical samples: Results of blueprint phase 2 project. *J. Thorac. Oncol.* 13, 1302–1311. doi:10.1016/j.jtho.2018.05.013

Vousden, K. H., and Prives, C. (2005). P53 and prognosis: New insights and further complexity. *Cell* 120, 7–10. doi:10.1016/j.cell.2004.12.027

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinforma. Oxf. Engl.* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics a J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zeng, H., Zhou, Q., Wang, Z., Zhang, H., Liu, Z., Huang, Q., et al. (2020). Stromal LAG-3$^+$ cells infiltration defines poor prognosis subtype muscle-invasive bladder cancer with immunoevasive contexture. *J. Immunother. Cancer* 8, e000651. doi:10.1136/jitc-2020-000651

Zhang, H., Song, J., Dong, J., Liu, Z., Lin, L., Wang, B., et al. (2021). Tumor microenvironment analysis identified subtypes associated with the prognosis and the tumor response to immunotherapy in bladder cancer. *Front. Genet.* 12, 551605. doi:10.3389/fgene.2021.551605

Zhang, J., Zhou, Q., Xie, K., Cheng, L., Peng, S., Xie, R., et al. (2021). Targeting WD repeat domain 5 enhances chemosensitivity and inhibits proliferation and programmed death-ligand 1 expression in bladder cancer. *J. Exp. Clin. Cancer Res.* 40, 203. doi:10.1186/s13046-021-01989-5

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 October 2022
DOI 10.3389/fgene.2022.1002673

OPEN ACCESS

EDITED BY
Yi Yao,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Kazi Ferdous Mahin,
United International University, Bangladesh
Xiaobin Gu,
First Affiliated Hospital of Zhengzhou University, China

*CORRESPONDENCE
Lianyuan Tao,
tly2007tly@hotmail.com
Deyu Li,
lideyu19@hotmail.com

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

RECEIVED 25 July 2022
ACCEPTED 16 September 2022
PUBLISHED 04 October 2022

CITATION
Ge J, Mu S, Xiao E, Tian G, Tao L and Li D (2022), Expression, oncological and immunological characterizations of BZW1/2 in pancreatic adenocarcinoma. *Front. Genet.* 13:1002673. doi: 10.3389/fgene.2022.1002673

Expression, oncological and immunological characterizations of BZW1/2 in pancreatic adenocarcinoma

Jiachen Ge, Senmao Mu, Erwei Xiao, Guangjin Tian, Lianyuan Tao* and Deyu Li*

Department of Hepatobiliary Surgery, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, China

Background: Despite the progress in early diagnosis and treatment, prognosis of pancreatic adenocarcinoma (PAAD) is still poor. Basic leucine zipper and W2 domain-containing protein 1 (BZW1) and protein 2 (BZW2) are attached to the basic leucine zipper (bZIP) superfamily. Recently, BZW1 was identified as an important role in glycolysis of PAAD. However, the comprehensive reports about BZW1/2 in PAAD are not sufficient.

Methods: RNA-seq data in the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were retrospectively analyzed. We explored the expression of BZW1/2 in PAAD tissues and the associations between BZW1/2 and prognosis. In addition, the potential roles of BZW1/2 in tumor microenvironment (TME) of PAAD were analyzed. Finally, clinicopathological data of 49 patients with PAAD in our institution were collected. Immunohistochemistry was used to determine the expression of BZW1/2 in PAAD samples.

Results: BZW1 and BZW2 were upregulated in PAAD tissues compared to normal tissues ($p < 0.05$). The expression of BZW1/2 were not significantly correlated with gender, grade and stage of PAAD ($p > 0.05$). High expression of BZW2 was an independent predictor for poor prognosis of PAAD (HR 1.834, 95%CI 1.303–2.581, $p = 0.001$). And a nomogram to predict overall survival (OS) of PAAD was established with a C-index of 0.685. BZW1 and BZW2 expression were positively associated with T cell mediated immune response to tumor cell and Th2 cells in xCell database. Tumor Immune Single-Cell Hub (TISCH) analyses indicated that BZW1 and BZW2 were mainly expressed in B cells and malignant cells. External cohort furtherly validated that high expression of BZW1 and BZW2 were predictors for poor prognosis of PAAD.

Conclusion: We found that BZW1 and BZW2 are highly expressed in malignant cells and B cells in the TME of PAAD. BZW2 is an independent predictor for OS of PAAD. BZW1 and BZW2 expression are positively associated with T cell mediated immune response to tumor cell and Th2 cells in PAAD.

KEYWORDS

pancreatic adenocarcinoma, BZW1, BZW2, tumor microenvironment, prognosis, tumor infiltrating immune cell

Introduction

Current evidence suggests that the incidence of pancreatic adenocarcinoma (PAAD) has increased apparently in recent years. In the USA, PAAD has become the 10th newest cancer in men and the eighth in women (Siegel et al., 2022). Despite the progress in early diagnosis and standard treatment, the prognosis of PAAD remains poor, with a 5-year overall survival rate of about 10% (Mizrahi et al., 2020). This may be due to resistance to standard therapies, including surgery, chemotherapy and radiotherapy. With the development of precise and targeted therapy, the emerging immunotherapies might successfully improve the prognosis of PAAD (Yin et al., 2022). Given the intricacy of PAAD tumor microenvironment (TME) characterized by fibrosis and poor vascularization, the efficiency of anti-tumor therapy is probably disturbed (Truong and Pauklin, 2021).

Basic leucine zipper and W2 domain-containing protein 1 (BZW1, BZAP45) and protein 2 (BZW2, 5MP1) are important members of the basic leucine zipper (bZIP) superfamily (Mitra et al., 2001; Singh et al., 2011). BZW1 and BZW2 genes encode a 45 kDa protein containing an N-terminal bZIP domain and a C-terminal W2 domain, and are highly expressed in bronchial epithelial cells and placenta, respectively (Hiraishi et al., 2014). As a transcription factor, BZW1 is identified as a conserved regulator during the G1/S transition (Mitra et al., 2001). Moreover, BZW2 is associated with tumorigenesis maintenance and cell-cell adhesion *via* translation initiation and cadherin binding, respectively (Guo et al., 2014). Currently, the roles of BZW1/2 in PAAD have been largely understudied. As a paralog of BZW1, we hypothesized that BZW2 might play a similar role in the progression of PAAD. In this study, we tried to analyze the expression profiles of BZW1/2 and their immunological characteristics in PAAD samples *via* public database. Furthermore, PAAD tissue samples were collected and protein expression of BZW1/2 and prognosis of PAAD were analyzed in an external validation cohort.

Methods

Sample collection and pan-cancer analysis

Clinical, pathological and gene-expression data were acquired from the Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) and Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/) databases. Patients who were diagnosed as PAAD with pathological evidence were included in this study. Samples from patients with pancreatic neuroendocrine or metastatic neoplasms were excluded from this study. Normalized pan-cancer datasets, TCGA TARGET GTEx (PANCAN, N = 19131, G = 60499), were downloaded from the UCSC database (https://xenabrowser.net/). Samples derived from solid tissue normal, primary solid tumor, normal tissue, primary blood derived cancer-bone marrow and primary blood derived cancer-peripheral blood were selected, and gene-expression data of BZW1 (ENSG00000082153) and BZW2 (ENSG00000136261) in each sample were extracted. Carcinoma sample number less than 3 was eliminated. As a result, 34 species of carcinoma were incorporated into the pan-cancer analysis.

Protein levels of BZW1/2 in PAAD samples

Immunohistochemical staining images of PAAD and normal tissues were collected from the Human Protein Atlas (HPA, https://www.proteinatlas.org/) (Colwill and Gräslund, 2011). The HPA could provide normal and pathological human tissue images stained by immunohistochemical sections. Protein levels of BZW1 and BZW2 in normal and PAAD tissues were explored in the HPA. Next, the differential expression of BZW1 and BZW2 proteins between normal and PAAD tissues was analyzed by clinical proteomic tumor or analysis consortium (CPTAC, http://ualcan.path.uab.edu/analysis-prot.html) database (Chandrashekar et al., 2017).

Roles of BZW1/2 in PAAD characteristics and prognosis

The mRNA expression data of BZW1 and BZW2 were downloaded from TCGA database. The mRNA profiles were standardized by $\log_2$ (x+0.001) transformation for further analysis. Follow-up was conducted from the data of tumor diagnosis, and the end was data of all-cause death. The recorded age, gender, grade, American Joint Committee on cancer (AJCC) stage, TNM stage, overall survival (OS) time and status of the samples were collected. Patients were grouped based on gender, grade, AJCC stage and TNM stage. The differences in BZW1 and BZW2 expression among groups were analyzed. Survival differences between the groups were also analyzed. Clinicopathological variables that were significantly correlated with prognosis were used to establish a prognostic model. An external cohort, GSE85916,

was used to validate the effects of BZW1 and BZW2 on the prognosis of PAAD patients.

Similar gene detection analysis

Similar gene detection module on the Gene Expression Profiling Interactive Analysis (GEPIA) website (http://gepia.cancer-pku.cn/detail.php) was used to explore genes with similar expression patterns to BZW1 and BZW2 of PAAD in TCGA dataset (Tang et al., 2017). The top 100 correlated genes of BZW1 and BZW2 were listed in descending order according to the Pearson correlation coefficient (PCC) values, which reflected the correlation intensity between variables (0.0–0.2, extremely poor; 0.2–0.4, poor; 0.4–0.6, moderate; 0.6–0.8, strong; >0.8, extremely strong).

Functional enrichment analysis

To better understand the potential functions of BZW1 and BZW2, Gene ontology (GO) annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations were performed by the Database for Annotation, Visualization, and Integrated Discovery (DAVID 6.8, https://david.ncifcrf.gov/). Cellular component (CC), molecular function (MF) and biological process (BP) categories were explored in GO analyses. Statistical significance was set as $p < 0.05$. The p values of the top five pathways were sorted in ascending order and displayed.

Gene set variation analysis (GSVA)

Gene lists of pancreatic cells and immune processes were acquired from Molecular Signatures Database (MSigDB v7.5.1, http://www.gsea-msigdb.org/gsea/msigdb/). We calculated the functional enrichment score of each PAAD sample. PCCs were used to evaluate the correlation between BZW1/2 expression and pancreatic cells or immune processes.

Establishment of the protein-protein interaction (PPI) network

It is well-established that the STRING database could be used to collect, score and integrate publicly available sources of PPI information (von Mering et al., 2005). The top 100 correlated genes were selected and incorporated to establish PPI networks. The network was constructed with a medium confidence at 0.400.

Tumor immune single-cell hub analysis

Tumor Immune Single-Cell Hub (TISCH, http://tisch.comp-genomics.org/home/) could provide detailed cell-type annotation at the single-cell level, and enabled the exploration of TME about different cancer types. The PAAD_CRA001160 dataset was selected to compare BZW1 and BZW2 expression in three main cell types, including immune cells, stromal cells and malignant cells. PAAD_CRA001160 dataset in TISCH contained single-cell RNA-seq of 57443 individual pancreatic cells from 35 primary PAAD patients (Peng et al., 2019).

ESTIMATE-stromal-immune score analysis

We further analyzed the roles of BZW1 and BZW2 in the TME. The Estimation of Stromal and Immune Cells in Malignant Tumors Tissues using Expression Data (ESTIMATE) algorithm was used to detect the fractions of infiltrating stromal and immune cells. And immune and stromal scores could be inferred by ESTIMATE algorithm. Higher immune or stromal components were associated with higher immune or stromal scores. The associations of BZW1/2 genes and ESTIMATE-Stromal-Immune score were explored.

Tumor-infiltrating immune cell abundance analysis

The correlations of BZW1/2 genes and tumor-infiltrating immune cell (TIIC) in PAAD samples were analyzed. The PCCs were calculated by CIBERSORT (https://cibersort.stanford.edu/), xCell (https://github.com/dviraran/xCell), MCP counter (https://github.com/ebecht/MCPcounter), Estimate the Proportion of Immune and Cancer cells (EPIC, https://gfellerlab.shinyapps.io/EPIC_1-1/), Tumor Immune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/), quanTIseq (http://icbi.i-med.ac.at/software/quantiseq/doc/index.html) and immunophenoscore (IPS, https://tcia.at/home). The computational methods mentioned were applied to estimate TIIC abundance in PAAD samples. The associations between BZW1/2 expression and TIIC abundance were analyzed.

Cell lines and culture

The human pancreatic ductal epithelium cell line HPDE6-C7 and two human pancreatic cancer cell lines, including MiaPaCa-2 and Panc-1, were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). The cell culture

medium was Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS) (Invitrogen). Cells (10^5) were grown in a humidified 5% CO_2 incubator at 37°C.

Real-time quantitative polymerase chain reaction (RT-qPCR)

Total RNA was extracted from cell lines using TRIzol reagent (Invitrogen, USA)and reverse transcribed into cDNA *via* the PrimeScript RT Reagent Kit with gDNA Eraser (Takara, Japan). RT-qPCR was carried out *via* TB Green Premix Ex Taq II (Takara). In this respect, the RNA template and its primer were incubated at 70°C for 10min, and cooled on ice for 5min. The complex mixture was incubated at 42°C for 50min, and then heated at 70°C for 15min. The mixed reaction mixtures were amplified for 40 cycles with the following procedure, 95°C for 10s, 60°C for 10s and 72°C for 20s. Expression of mRNA was calculated by $2^{-\Delta\Delta Ct}$ method. GAPDH was used to normalize the results of RT-qPCR. The primer sequences were as following: GAPDH, GGAGCGAGATCCCTCCAAAAT(F), GGCTGTTGTCA TACTTCTCATGG(R); BZW1, AAGAGAGGTTTGACCCTACTC AG(F), CTGCATATCGACGGTAATCAAGT(R); BZW2, CTA ACAGGCCAGCGGTTCAAA(F), GGACAAGTGTATCCCTG AAGACT(R).

Immunohistochemistry of BZW1 and BZW2 in PAAD samples

To further clarify the expression of BZW1 and BZW2 in PAAD samples, we retrospectively collected the clinicopathological data of 49 patients who underwent needle biopsy or radical surgery in the Department of Hepatobiliary Surgery at Henan Provincial People's Hospital from June 2018 to May 2021. Patients were asked to receive systematic chemotherapy post-operation if circumstances permitted. After discharge, they were requested to periodic follow-up. Immunohistochemistry staining was performed to analyze the expression of BZW1 and BZW2 in tumor specimens. BZW1-(#ab85090) and BZW2-(#ab254772) antibodies were purchased from Abcam company. The EnVision two-step method was used to stain the specimen. The expression of BZW1 and BZW2 were recorded by light microscopy and evaluated by two pathologists independently. The correspondence rules of staining intensity and scoring standard were as following: uncolored, 0; light yellow, 1; yellow, 2; brown, 3. Positive expression region percentage criteria were as following: 0–30%, 0; 30–60%, 1; >60%, 2. After calculating the total points, samples were grouped into low-expression (0–2 points) and high-expression (three to five points) groups. The Pearson correlation coefficient between the two pathologists was 0.87.

Statistical analysis

All statistical analyses in this study were conducted by R software (version 4.1.0). Comparison analyses among expression quantities of BZW1 and BZW2 were performed by student's t test, Mann-Whitney *U* test or variance analysis. The Kaplan-Meier method was used to analyze survival difference between low- and high-expression level groups. Significance of the correlation between the two groups was tested by Pearson correlation analysis. OS was calculated by the Kaplan-Meier method, and the survival differences were compared by the Log-Rank test. Univariate and multivariate cox proportional hazards regression methods (backward selection) were used to identify clinicopathologic variables significantly associated with OS. The optimal cut-off values between low- and high-expression groups were determined by X-tile software (version 3.6.1). The PPI network was visible *via* Cytoscape software (version 3.9.1). R package "regplot" was used to construct the nomogram. A two-tailed *p* value less than 0.05 was statistically significant unless otherwise mentioned.

Results

BZW1/2 were upregulated in PAAD

Expression data about 34 types of carcinomas was acquired from the UCSC database. Mann-Whitney *U* test was used to make differential expression analyses. A significant expression difference of BZW1 between normal and tumor samples was observed in 28 species. Among them, up-regulated expression of BZW1 in tumor tissue was detected among 22 types of carcinomas, including GBM, GBMLGG, LGG, BRCA, CESC, LUAD, ESCA, STES, COAD, COADREAD, PRAD, STAD, LUSC, LIHC, WT, SKCM, THCA, OV, PAAD, UCS, ALL and LAML (*P* all <0.001). And down-regulated expression of BZW2 was detected among 6 types of carcinomas, including KIRP, KIPAN, TGCT, PCPG, ACC and KICH (*Ps* < 0.05, Figure 1A, Supplementary Table S1). Highly expressed BZW2 was observed in 30 species of carcinomas, including GBM, GBMLGG, LGG, UCEC, BRCA, CESC, LUAD, ESCA, STES, KIRP, KIPAN, COAD, COADREAD, PRAD, STAD, HNSC, LUSC, LIHC, WT, SKCM, BLCA, READ, OV, PAAD, TGCT, UCS, ALL, LAML, ACC and CHOL (*Ps* < 0.05). Moreover, down-regulated expression of BZW2 in tumor tissue was detected among KIRC, THCA and KICH (*Ps*, Figure 1B, Supplementary Table S1). We acquired expression data of two cohorts from the GEO database, GSE28735 and GSE62452. And the results indicated

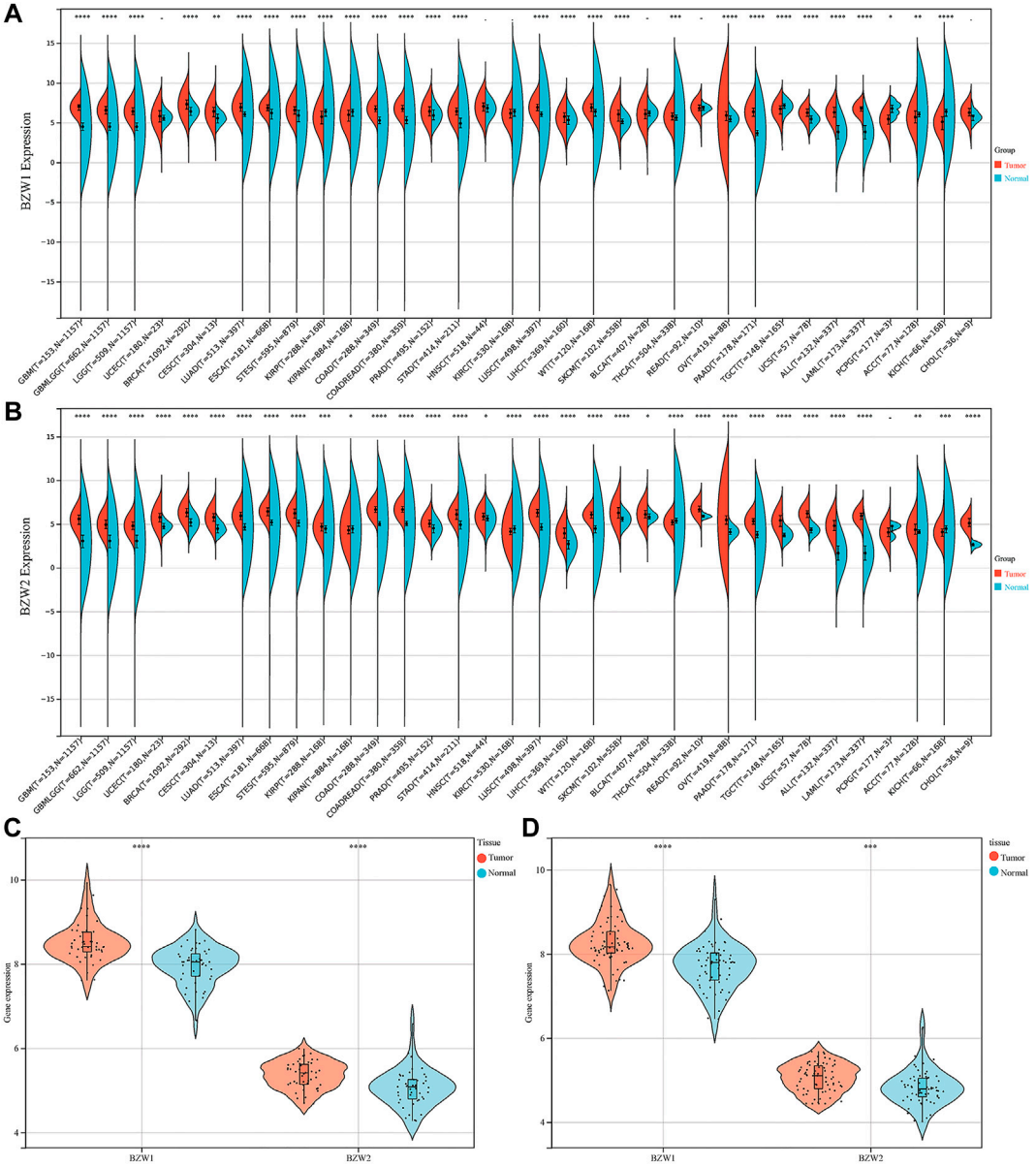


FIGURE 1
BZW1 **(A)** and BZW2 **(B)** expression in pan-cancer. Expression of BZW1 and BZW2 was upregulated in tumor tissues in GSE28735 **(C)** and GSE62452 **(D)** cohorts. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

that BZW1 and BZW2 were up-regulated in PAAD compared to normal tissues (*P* all <0.001, Figures 1C,D, Supplementary Table S2). Immunohistochemistry images of BZW1 and BZW2 were obtained from the HPA website. Compared to normal tissues, BZW1 and BZW2 proteins were highly expressed in PAAD tissues using the same antibody (BZW1 antibody: HPA053272; BZW2 antibody: HPA022813) (Figures 2A–D). In the CPTAC database, expression levels of BZW1 and BZW2 in PAAD samples were significantly higher than in normal samples (Figures 2E,F).

BZW2 was an independent predictor for OS of PAAD patients

The relationships between BZW1/2 expression and clinicopathological parameters were investigated. The results indicated that patients with T3 stage PAAD exhibited higher expression level of BZW2 than T2 and T4 stage (*Ps* < 0.05, Figure 3J, Supplementary Table S3). In addition, BZW1 and BZW2 expression were not significantly associated with gender, grade, AJCC stage and TNM stage of

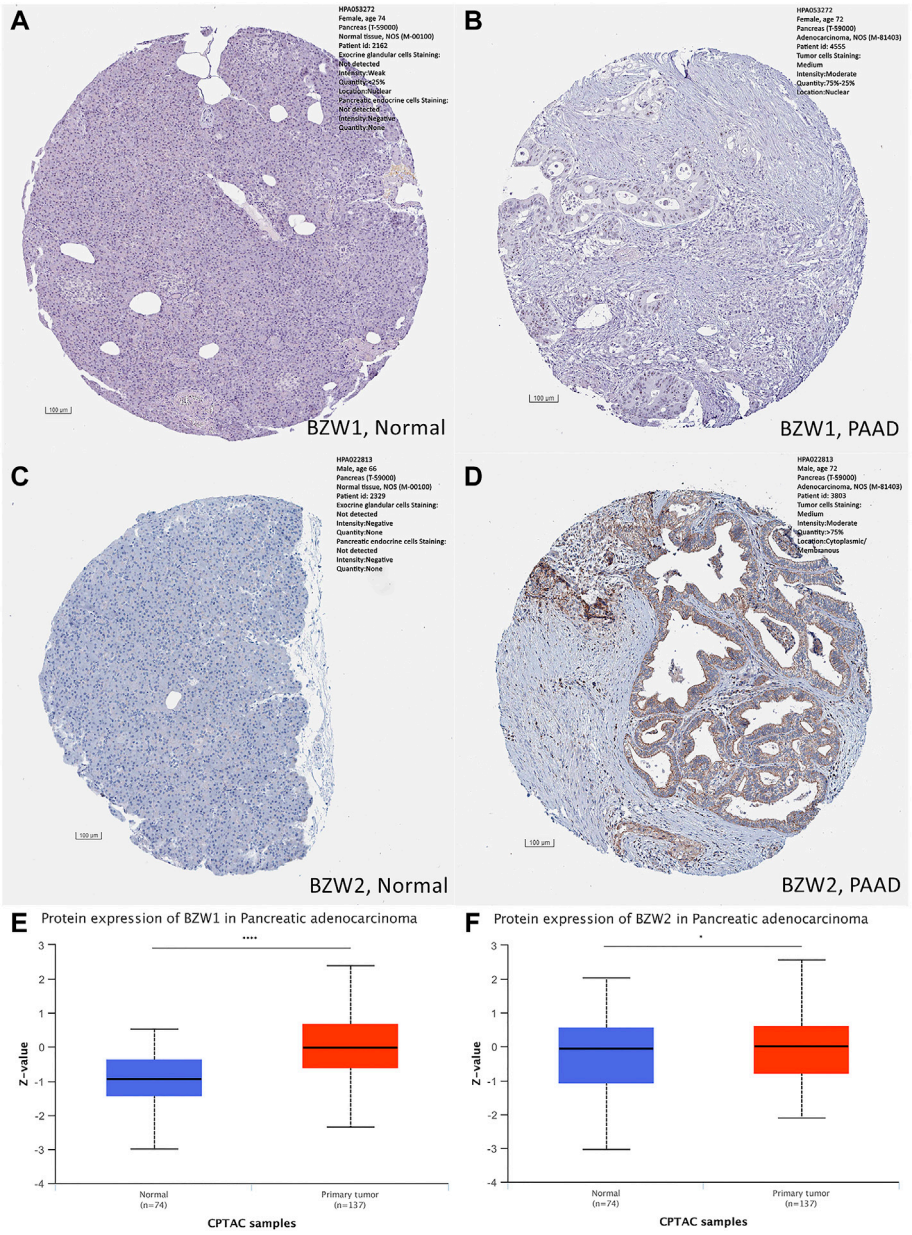


FIGURE 2
Immunohistochemistry staining in normal and PAAD tissues, respectively. Compared to normal tissues, the expression levels of BZW1 and BZW2 were upregulated in PAAD tissues **(A,B,C,D)**. Compared to normal tissues, expression levels of BZW1 **(E)** and BZW2 **(F)** proteins were significantly higher in PAAD tissues based on CATAC database.

PAAD patients (*Ps* > 0.05, Figure 3A-I, K, L, Supplementary Table S3).

Kaplan-Meier survival analysis indicated that PAAD patients with high BZW1 and BZW2 expression had poorer OS than those with low BZW1 and BZW2 levels ($p < 0.001$ for BZW1, $p = 0.004$ for BZW2) (Figures 4A,B). GSE85916 cohort was used to validate the influences of BZW1 and BZW2 expression on prognosis. Survival analyses showed that high expression of BZW1 and BZW2 were significantly associated with short OS for PAAD patients ($p = 0.040$ for BZW1 probe 11757867_s_at, $p = 0.046$ for BZW1 probe 11744775_x_at and $p = 0.030$ for BZW2 probe 11747677_a_at) (Figures 4D–F). Univariate cox regression analysis indicated that age, grade, AJCC stage, N stage, chemotherapy, radiotherapy, BZW1 and BZW2 expression were

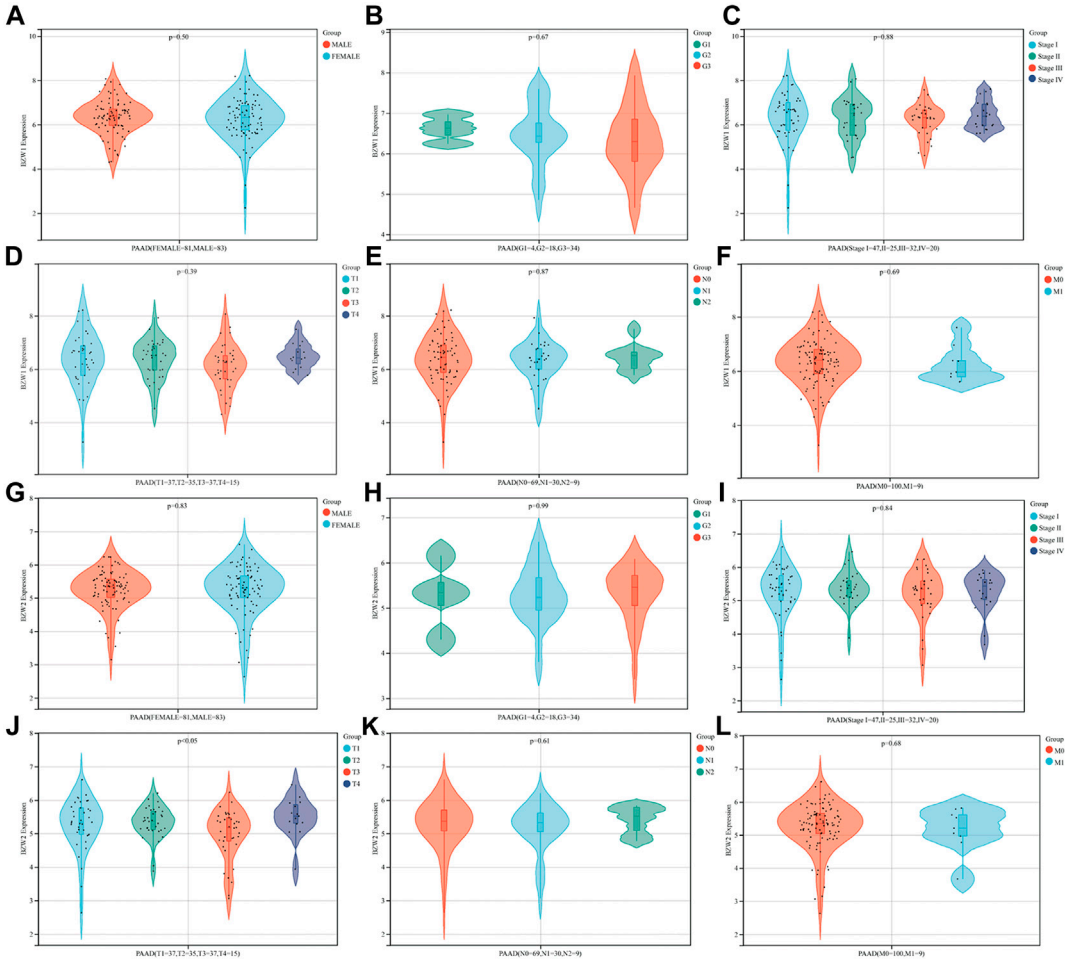


FIGURE 3
Relationships between BZW1/2 expression and clinicopathological parameters of PAAD. The expression of BZW2 in patients with T3 stage disease was higher than in T2 and T4 stage **(J)**. BZW1 and BZW2 expression were not significantly associated with gender **(A,G)**, grade **(B,H)**, AJCC stage **(C,I)** and TNM stage **(D,E,F,K,L)** of PAAD patients.

independent predictors for OS (Figure 5A). Multivariate cox proportional regression showed that older age (HR 1.020, 95%CI 1.001–1.040, $p = 0.037$), N1 stage (HR 2.100, 95%CI 1.273–3.462, $p = 0.004$), no chemotherapy (HR 2.860, 95%CI 1.846–4.433, $p < 0.001$) and higher BZW2 expression (HR 1.834, 95%CI 1.303–2.581, $p = 0.001$) were independently associated with poor OS (Figure 5B). A nomogram for OS of patients with PAAD was established based on these independent predictors. After adding the score corresponding to each factor, the risk score of each patient and the probability of OS longer than 180-, 365- and 1095-day could be calculated (Figure 5C). Survival probability of patients with higher risk score was significantly lower than those with lower risk score ($p = 0.009$, Figure 4C). This nomogram demonstrated a C index of 0.685 (95%CI 0.657–0.713). Heatmap and scatter plot indicated that the survival rate decreased with an increased risk score (Figure 5D). The receiver operator characteristic curves (ROCs) of 180-,365- and 1095-day survival were plotted in Figure 5E.

Similar gene analyses of BZW1/2

Similar genes module in the GEPIA webtool was used to explore similar genes of BZW1 and BZW2 genes. The top similar 100 genes of BZW1 and BZW2 were listed in Supplementary Table S3. The top five highly correlated genes with BZW1 were BZW1P2, NAA50, KPNA4, G2E3 and RAB10 according to the PCC values. And the top five highly correlated genes with BZW2 were CBX3, NIFK, AVL9, PPIAP22 and PPIA. The

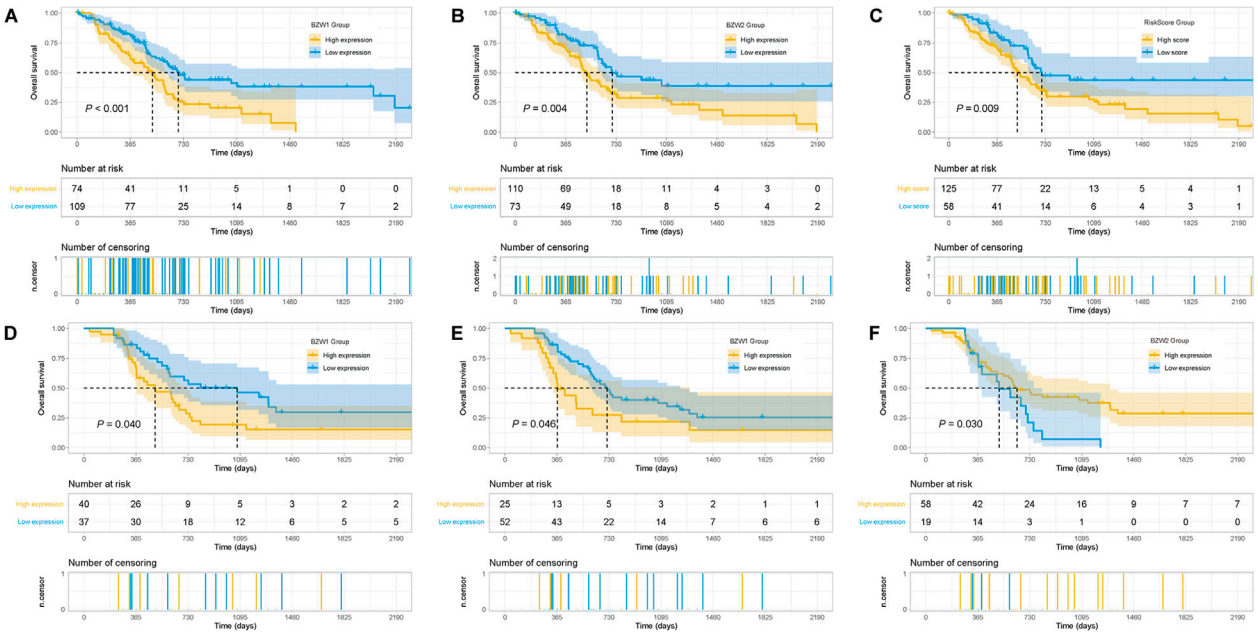


FIGURE 4
Kaplan-Meier survival curves of different BZW1 and BZW2 expression for TCGA-PAAD patients **(A,B)**. Survival curve of low- and high-RiskScore group for TCGA-PAAD patients **(C)**. Survival curves indicated that patients with high expression of BZW1 (11757867_s_at and 11744775_x_at) and BZW2 (11747677_a_at) suffered poor prognosis in GSE85916 cohort **(D)**: 11757867_s_at, **(E)** 11744775_x_at and **(F)** 11747677_a_at).

heatmap in Supplementary Figure S1 indicated the clustering and correlation of similar genes.

Biological function exploration

To further explore the potential biological functions of BZW1 and BZW2, we performed GO and KEGG analyses based on BZW1/2 and their similar genes. The top five enriched BP clusters of BZW1 included RNA localization, intracellular transport, establishment of RNA localization, intracellular protein transport and nucleobase containing compound transport (Figure 6A). The top five enriched MF clusters of BZW1 included RNA binding, ribonucleoprotein complex binding, arrestin family protein binding, hydrolase activity acting on acid anhydride and peptide alpha-N-acetyltransferase activity (Figure 6B). The top five enriched CC clusters of BZW1 included nuclear pore, phosphatase complex, nuclear envelop, NatA complex and nuclear protein containing complex (Figure 6C). Significantly enriched KEGG pathways involved in variations of BZW1 expression included RNA transport, protein processing in the endoplasmic reticulum, AMPK signaling pathway and mRNA surveillance pathway (Figure 6D). The top five enriched BP clusters of BZW2 included ribonucleoprotein complex biogenesis, ribosome biogenesis, peptide biosynthetic process, RRNA metabolic process and peptide metabolic process (Figure 6E). The top five enriched CC clusters of BZW2 included ribonucleoprotein complex, envelope, pre-ribosome, mitochondrion and organelle inner membrane (Figure 6F). The top five enriched MF clusters of BZW2 included RNA binding, translation initiation factor activity, translation factor activity RNA binding, translation regulator activity and translation regulator activity nucleic acid binding (Figure 6G). The enriched KEGG pathways involved in variations of BZW2 expression included ribosome biogenesis in eukaryotes and spliceosome (Figure 6H).

BZW1/2 expression correlated with pancreatic cells and immune processes

During the progression and evolution of the TME, it was highly likely that interactions prevail between PAAD and normal pancreatic cell or immune processes. Hence, we tried to explore the effects of BZW1/2 genes on pancreatic cells and immune processes. GSVA was used to determine the enrichment scores

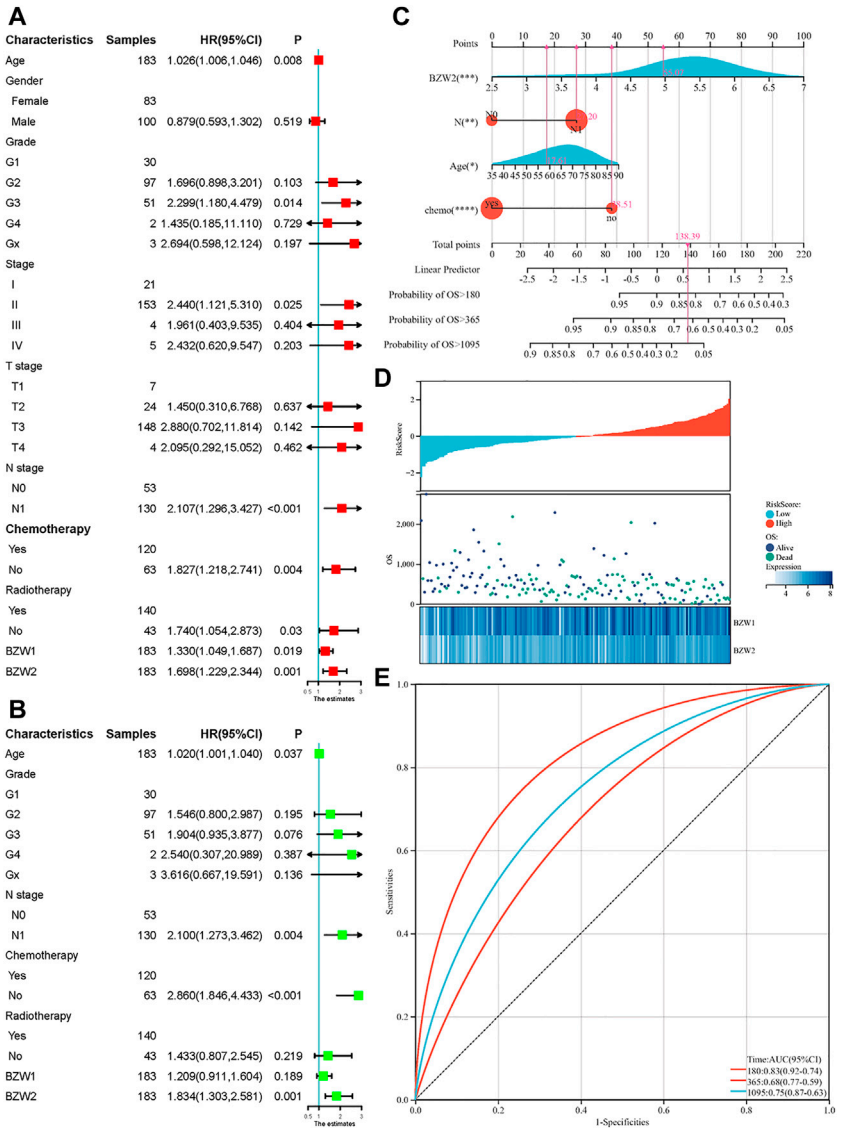


FIGURE 5
Univariate (A) and multivariate (B) cox regression for OS of PAAD patients. A nomogram was established to predict the OS of PAAD patients based on the independent predictors (C). Heatmap and scatter plot of risk score, survival time and status of all samples (D). ROC curves of the nomogram to predict 180-, 365- and 1095-day survival (E).

of pancreatic cells and immune processes in TCGA samples. Correlation analysis indicated that BZW1 and BZW2 expression was negatively associated with pancreas alpha cell, pancreatic polypeptide cell and epsilon cell. Moreover, BZW1 and BZW2 expression was positively associated with pancreas ductal cells. In addition, the results of correlation analysis suggested that BZW1 and BZW2 expression was negatively associated with neutrophil activation involved in the immune response but positively correlated with T cell mediated immune response to tumor cells ($Ps < 0.05$, Figures 7A,B).

Associations between BZW1/2 expression and the TME of PAAD

It was widely acknowledged that the ESTIMATE-Stromal-Immune score could be used to evaluate the composition of immune cells and stromal cells in the TME. In the present study, stromal score, immune score and ESTIMATE score of each sample were calculated. The results indicated that BZW1 expression was positively associated with stromal score and ESTIMATE score. However, BZW2 expression was be not

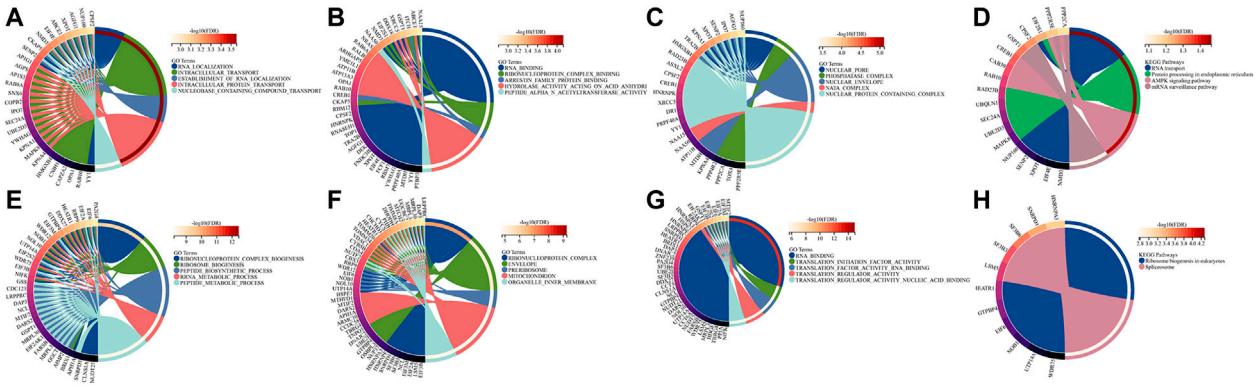


FIGURE 6
Circos plots displayed the results of GO and KEGG analyses for BZW1, BZW2 and their similar genes. **(A)** BZW1 GO BP. **(B)** BZW1 GO MF. **(C)** BZW1 GO CC. **(D)** BZW1 KEGG. **(E)** BZW2 GO BP. **(F)** BZW2 GO CC. **(G)** BZW2 GO MF. **(H)** BZW2 KEGG.

significantly correlated with stromal score, immune score and ESTIMATE score (Table 1 and Supplementary Figure S2).

Correlation between BZW1/2 and immune infiltration in PAAD

We further assessed the correlations between BZW1/2 expression and the types of TIICs. TIICs were evaluated by CIBERSORT, xCELL, MCPcounter, EPIC, TIMER, quanTIseq and IPS methods. These public microarray datasets were used to calculate immune infiltration score of each sample. After filtering the results using $p > 1.0E\text{-}5$, we found that BZW1 expression was negatively associated with the infiltration of basophils, $CD4^+$ Tcm, MEP, NKT and Th1 cells, and positively associated with the presence of CLP, smooth muscle, Th2 cells, neutrophils, T cell CD8, DC, macrophages M1. Moreover, BZW2 expression was negatively correlated with hepatocytes, neurons and Tgd cells, and positively correlated with epithelial cells, keratinocytes, sebocytes, Th2 cells and macrophages M1 (Table 2, Supplementary Figure S3). These results indicated that BZW1 and BZW2 expression influenced TIICs through severalpathways.

PPI network construction and TISCH analysis

To clarify the co-expression and interactions among of proteins, the PPI network was constructed based on the STRING database (Figure 8A). 18 and 11 proteins were proven to be interacted with BZW1 and BZW2, respectively. The interaction in the PPI network were shown in Supplementary Table S5. AGR3, eIF1, eIF2S2, eIF5 and SNX13 genes were interacted with both BZW1 and BZW2 at the same time. The PAAD_CRA001160 cohort in the TISCH database was used to investigate the relationships between BZW1/2 expression levels and tumor stromal cell infiltrations. Due to the heterogeneity of TME, we investigated the expression of BZW1 and BZW2 in stromal cell components. BZW1 was mainly expressed in B cells, fibroblasts and malignant cells, and BZW2 was upregulated in B cells, endothelial and malignant cells (Figures 8B,C).

Expression of BZW1/2 in pancreatic cell lines

BZW1 and BZW2 expressions were further verified through RT-qPCR in HPDE6-C7, MiaPaCa-2 and Panc-1 cell lines. The mRNA levels of BZW1 and BZW2 were significantly higher in Panc-1 and MiaPaCa-2 cell lines than in the HPDE6-C7 cell line (Figures 9A,F).

External cohort validation confirmed that BZW1/2 expression could predict OS in PAAD

49 samples were acquired as an external cohort, and all patients were successfully followed up. After immunohistochemical staining accomplished, the expression of BZW1 and BZW2 were classified as low- and high-expression based on scores of samples (Figure 9B-E, G-J). Among them, 13 and 36 patients were recognized as

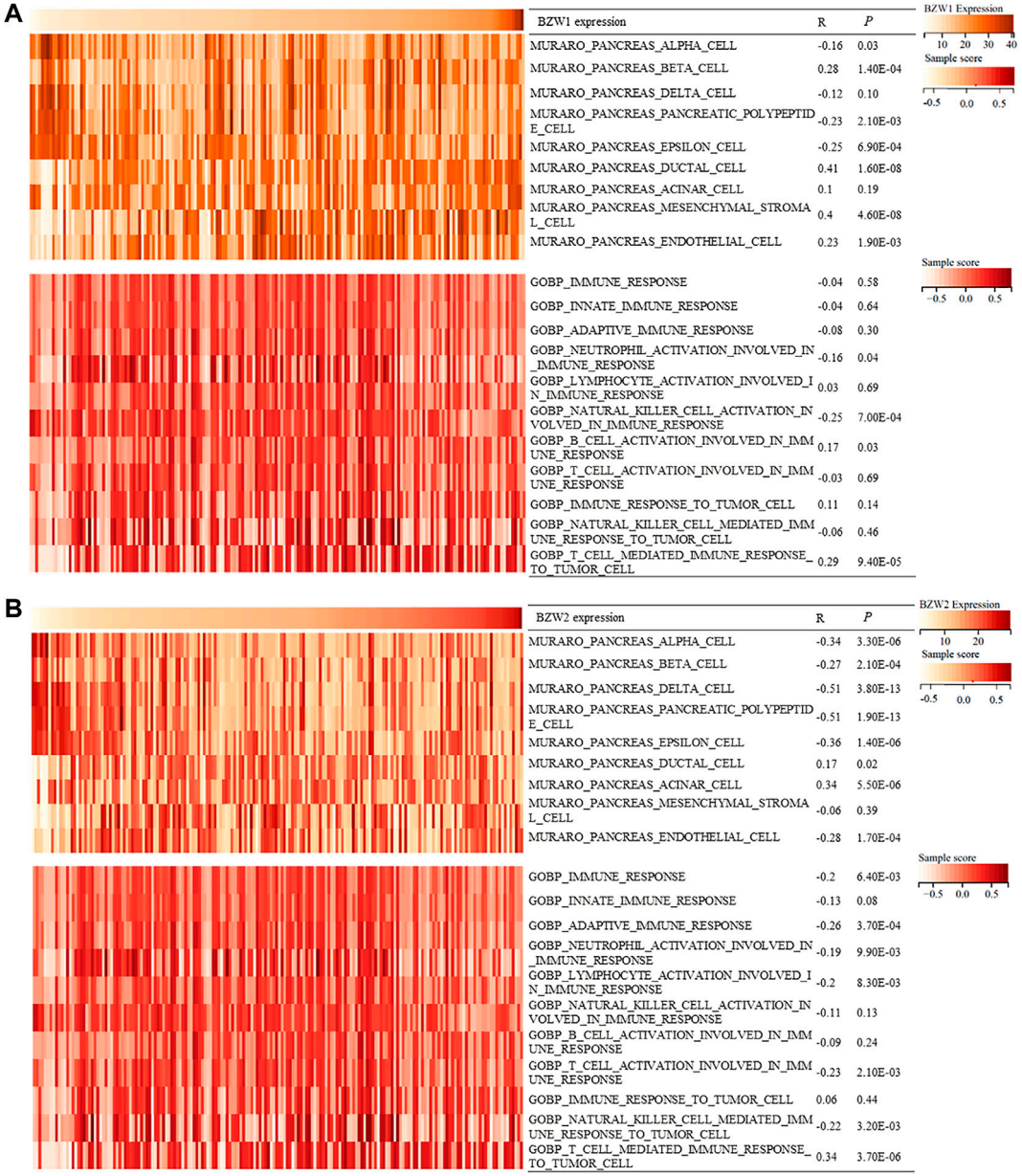


FIGURE 7
Heatmap showed BZW1/2 expression and enrichment scores of pancreatic cells and immune responses about each sample in each TCGA sample. **(A)** Correlation analyses between BZW1 expression and pancreatic cell or immune response enrichment scores. **(B)** Correlation analyses between BZW2 expression and pancreatic cell or immune response enrichment scores.

TABLE 1 Correlation of BZW1/2 expression and ESTIMATE-Stromal-Immune score (Pearson).

Gene symbol	Stromal score		Immune score		ESTIMATE score	
	R value	*p* value	R value	*p* value	R value	*p* value
BZW1	0.253	6.57*E*-4	0.078	0.301	0.173	0.021
BZW2	-0.030	0.689	-0.029	0.698	-0.031	0.678

TABLE 2 Correlations between BZW1/2 expression and TIICs in PAAD (Pearson).

	Negative correlation			Positive correlation		
	Cell type (dataset)	*p* value	R value	Cell type (dataset)	*p* value	R value
BZW1	Basophils (xCELL)	3.394E-9	-0.426	CLP (xCELL)	1.718E-6	0.351
	$CD4^+$ Tcm (xCELL)	5.390E-7	-0.366	smooth muscle (xCELL)	3.227E-8	0.401
	MEP (xCELL)	1.155E-13	-0.520	Th2 cells (xCELL)	2.804E-6	0.344
	NKT (xCELL)	1.937E-08	-0.407	neutrophils (MCPcounter)	4.127E-8	0.398
	Th1 cells (xCELL)	7.521E-13	-0.505	neutrophils (quanTIseq)	4.962E-9	0.422
				neutrophils (TIMER)	7.93E-10	0.441
				T cell CD8 (TIMER)	1.450E-16	0.569
				DC (TIMER)	1.655E-15	0.552
				macrophages M1 (quanTIseq)	2.030E-8	0.406
BZW2	hepatocytes (xCELL)	2.062E-10	-0.455	epithelial cells (xCELL)	5.393E-15	0.544
	neurons (xCELL)	2.880E-10	-0.451	keratinocytes (xCELL)	3.884E-10	0.448
	Tgd cells (xCELL)	9.534E-6	-0.326	sebocytes (xCELL)	1.122E-8	0.413
				Th2 cells (xCELL)	2.049E-7	0.379

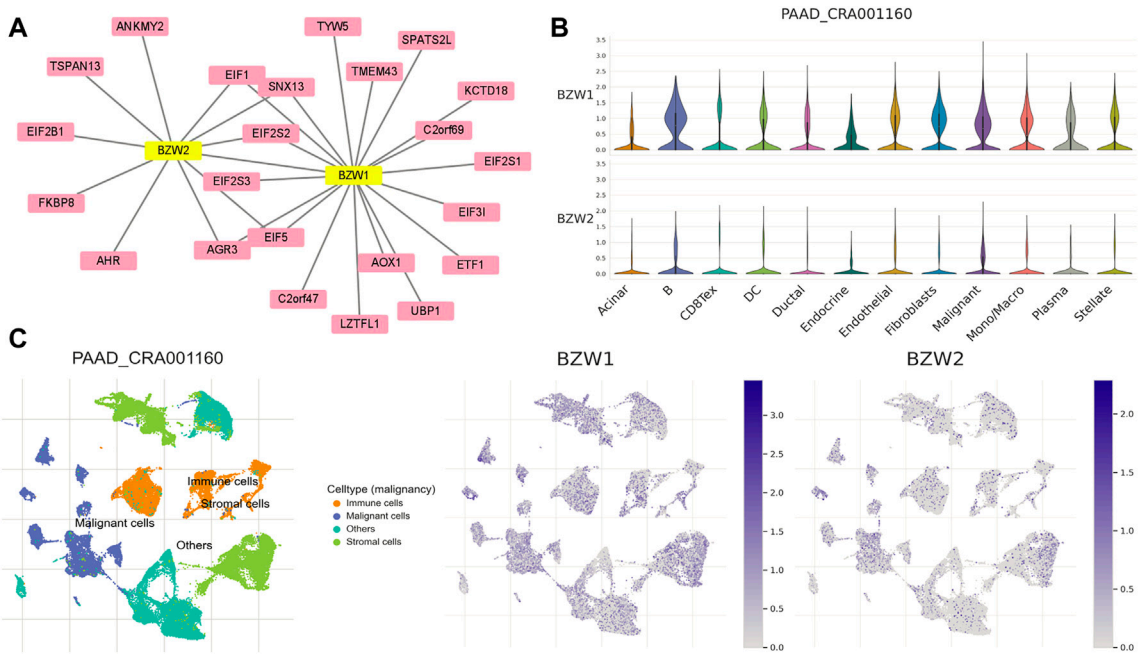


FIGURE 8
(A) PPI network of BZW1- and BZW2-corrected genes based on Cytoscape. **(B)** Violin plots indicated the expression of BZW1 and BZW2 among different cell types in the PAAD_CRA001160 cohort. **(C)** T-distributed stochastic neighbor embedding plots showed the expression of BZW1 and BZW2 in different cell types.

BZW1 high- and low-expression, respectively. 18 patients were identified as BZW2 high-expression, and 31 were low-expression. Correlations between BZW1/2 expression and clinicopathological factors of patients were listed in Supplementary Table S6. The expression of BZW1 and BZW2 showed no significant associations with age, gender, tumor location, differentiation, T stage, N stage, chemotherapy and radiotherapy. Survival analyses indicated

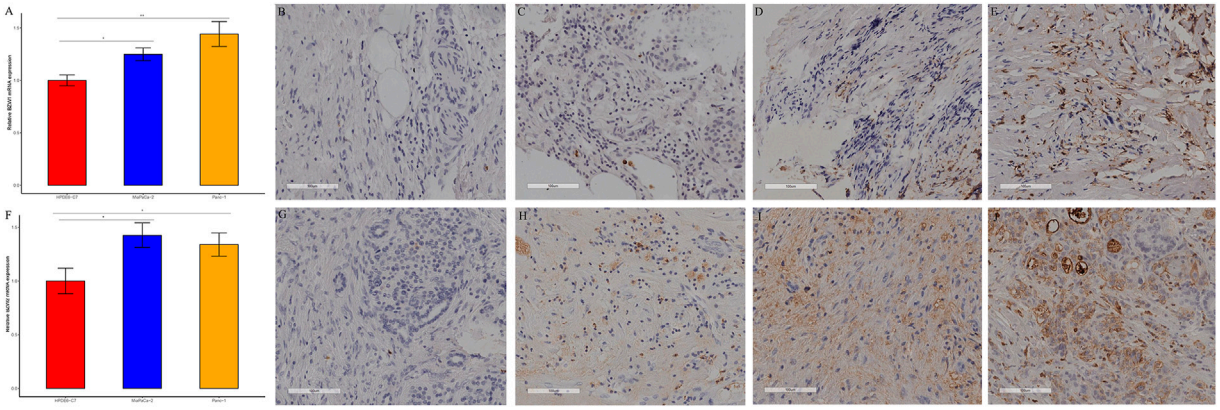


FIGURE 9
RT-qPCR and immunohistochemical staining verified the expression of BZW1 and BZW2 in PAAD samples. The relative mRNA expression levels of **(A)** BZW1 and **(F)** BZW2 in HPDE6-C7, MiaPaCa-2 and Panc-1 cell lines. **(B–E)** BZW1 was mainly expressed in the cell nucleus. Stain intensity: **(B)** BZW1, score 0. **(C)** BZW1, score 1. **(D)** BZW1, score 2. **(E)** BZW1, score 3. **(G–J)** BZW2 mainly expressed in cytoplasm and membrane. Stain intensity: **(G)** BZW2, score 0. **(H)** BZW2, score 1. **(I)** BZW2, score 2. **(J)** BZW2, score 3.

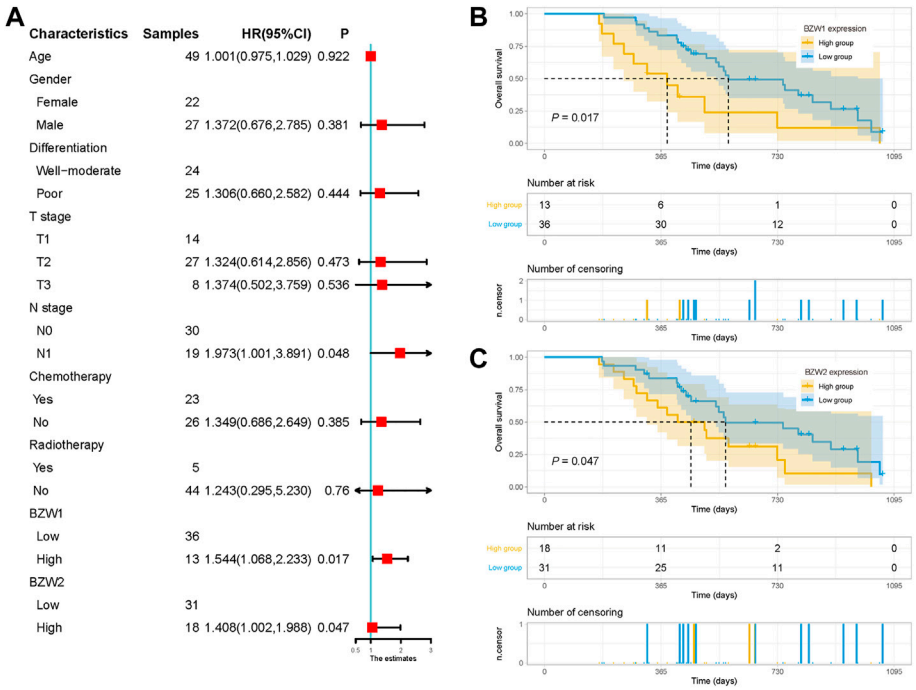


FIGURE 10
(A) Univariate analysis of OS in the external validation cohort. Survival curves for BZW1 **(B)** expression and BZW2 **(C)** expression in the external validation cohort.

that N stage, BZW1 expression and BZW2 expression were independent predictors for OS in PAAD patients (Figure 10A). High expression of BZW1 and BZW2 were significantly corrected with poor prognosis (BZW1, $p = 0.017$; BZW2, $p = 0.047$) (Figures 10B,C).

Discussion and conclusion

In China, PAAD ranks the sixth leading cause of cancer related deaths (Xia et al., 2022). Notwithstanding that significant inroads had been achieved in recent years,

PAAD patients still suffered poor prognoses and relatively short OS time (Xia et al., 2022). Given the insensitivity and drug-resistance, conventional therapy, which included surgery, chemotherapy and radiotherapy, often led to non-satisfactory curative effect nowadays. Targeted and immunological therapies were prized in the treatment of PAAD. Indeed, there was a long way to go for new therapies to be translated clinically. Based on publicly available tumor database, RNA-seq data could be used to search for prognostic-related genes and potential therapeutic targets. BZW1 and BZW2 were known as eukaryotic translation initiation factor 5 (eIF-5) mimic proteins (Singh et al., 2011; Loughran et al., 2018). It was reported that they shared a C-terminal W2 HEAT domain and an N-terminal bZIP domain (Aravind and Koonin, 2000; Loughran et al., 2018). In this study, we analyzed the expression levels of BZW1 and BZW2 in pan-cancer and distinct stages of PAAD based on the mRNA-seq data from TCGA databases. Our results suggested that BZW1 and BZW2 were highly expressed in PAAD tissue compared to normal tissue. Results of RT-qPCR indicated that the expression levels of BZW1 and BZW2 in human pancreatic cancer cell lines were higher than in the human pancreatic epithelium cell line. Immunohistochemical staining showed that BZW1 was mainly located in the nucleus, while BZW2 was mainly expressed in the cytoplasm and membrane. Survival analyses indicated that BZW2 expression was an independent predictor for OS of PAAD patients. A nomogram based on BZW2 expression and clinicopathological factors was established to predict the prognosis of PAAD.

Current evidence suggested that BZW1 and BZW2 could influence the stringency of start codon selection in mammalian cells (Loughran et al., 2018). Besides, BZW1 transcripts were alternatively polyadenylated and expressed in tissue-specific pattern (Yu et al., 2006). The functions of BZW1 and BZW2 had been largely understudied until recently. BZW1 and BZW2 have been established to promote the malignant progression of several cancers. Shi et al. indicated that BZW1 overexpression could promote prostate cancer cell proliferation by regulating TGF-β/Smad pathway (Shi et al., 2021). Jin et al. showed that overexpression of BZW2 in hepatocellular carcinoma cells significantly stimulated the activation of PI3K/AKT/mTOR pathway (Jin et al., 2019). Huang et al. demonstrated that BZW2 promoted malignant progression of colorectal cancer *via* activating ERK/MAPK pathway (Huang et al., 2020). Data presented in our study substantiated that high expression of BZW1 and BZW2 were predictors for poor prognosis of PAAD patients. Li et al. demonstrated that BZW1 could promote cell proliferation and inhibit apoptosis of PAAD cells *via* facilitating glycolysis in mouse xenograft models and organoids (Li et al., 2022). They inferred that BZW1 had the potential to be a new therapeutic target for PAAD (Li et al., 2022).

PPI network analysis in our study showed an interaction between BZW1 and BZW2, and single protein that connected both BZW1 and BZW2 included eIF1, eIF2S2, eIF5, SNX13 and AGR3. A broad range of eIFs had been established to regulate the initiation step of translation (Jackson et al., 2010). Members in the eIFs family were relevant to PAAD biology, and eIF1, eIF2D, eIF3C and eIF6 were identified as new biomarkers of PAAD (Golob-Schwarzl et al., 2020). Singh et al. indicated that BZW1 and BZW2 inhibited the recruitment and recycling of eIF2 by inhibiting its association with eIF5(Singh et al., 2011). Hiraishi et al. reported that BZW2 could serve as a competitor with eIF5, and BZW1 and BZW2 enhanced translation of ATF4, a key protein of endoplasmic reticulum stress and a potential target of PAAD (Hiraishi et al., 2014). Nachmias et al. proposed that BZW1 and BZW2 mediated the stemness and survival of leukemia stem cells (Nachmias et al., 2022). Moreover, BZW1 and BZW2 could regulate TYK2 expression, which acted as an oncogene and a therapeutic target in acute myelocytic leukemia (Sanda et al., 2013; Nachmias et al., 2022). Li et al. indicated that BZW1 facilitated eIF2α phosphorylation and promoted internal ribosome entry site-dependent translation of HIF1α and c-Myc in PAAD samples (Li et al., 2022).

To further clarify the roles of BZW1 and BZW2 in TME of PAAD, we conducted GSVA for each sample in TCGA database. The results revealed that BZW1 and BZW2 expression positively correlated with T cell mediated immune response to tumor cell. Correlations between BZW1/2 expression and TIICs were further analyzed. BZW1 and BZW2 expression were positively correlated with Th2 cells, which might facilitate tumor growth in TME of PAAD (Piro et al., 2017). Interestingly, TISCH analyses showed that BZW1 and BZW2 might regulate TME of PAAD by targeting B cells, which were activated by Th2 cells. Moreover, BZW1 positively correlated with Stromal score and ESTIMATE score. These results may reveal the potential roles of BZW1 and BZW2 in immune TME of PAAD.

Although a significant association with the prognosis of PAAD was found, BZW1 and BZW2 expression were not significantly correlated with gender, grade, AJCC stage, N stage and M stage. These results might indicate that the expression of BZW1/2 was relatively stable during tumor progression. However, several limitations were found in this study. First, subgroup analyses were not carried out due to the relatively small sample size. Therefore, the prediction performance of BZW1/2 remained unclear among different subgroups. Second, the associations between BZW1/2 expression and immune TME were not externally validated. Furthermore, future studies should investigate the role of BZW2 in tumor growth and progression of PAAD. As

potential targets of PAAD, more experiments were needed to clarify the detailed mechanisms underlying the roles of BZW1/2 in PAAD.

In conclusion, we found that BZW1 and BZW2 were highly expressed in malignant cells and B cells in the TME of PAAD. BZW2 was an independent predictor for prognosis of PAAD. BZW1 and BZW2 expression were positively associated with T cell mediated immune response to tumor cell and Th2 cells in PAAD.

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 Ethics Committee of Henan Provincial People's Hospital, People's Hospital of Zhengzhou University (2019 No.58). The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, JG, LT, and DL; investigation, JG and SM; visualization, EX and GT; supervision, LT and DL; validation, JG and DL; funding acquisition, LT and DL; project administration, EX and DL; software and methodology, JG and LT; resources, DL; writing—original draft preparation, JG and LT; writing—review and editing, SM and DL. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by grants from the Joint Project of Medical Science and Technology Research Program of Henan Province (LHGJ20190577), Medical Science and Technology Research Plan of Henan Province, Project Co-built by Provincial Department (SB20190319), and Henan Province Health Science and Technology Innovation Excellent Young Talents Training Project (No. YXKC2020044).

Acknowledgments

We thank Home for Researchers editorial team (www.home-for-researchers.com) for language editing service.

Conflict of interest

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

The reviewer XG declared a shared parent affiliation with the author(s) to the handling editor at the time of review.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1002673/full#supplementary-material

References

Aravind, L., and Koonin, E. V. (2000). Eukaryote-specific domains in translation initiation factors: Implications for translation regulation and evolution of the translation system. *Genome Res.* 10 (8), 1172–1184. doi:10.1101/gr.10.8.1172

Chandrashekar, D. S., Bashel, B., Balasubramanya, S., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19 (8), 649–658. doi:10.1016/j.neo.2017.05.002

Colwill, K., and Gräslund, S. (2011). A roadmap to generate renewable protein binders to the human proteome. *Nat. Methods* 8 (7), 551–558. doi:10.1038/nmeth.1607

Golob-Schwarzl, N., Puchas, P., Gogg-Kamerer, M., Weichert, W., Göppert, B., and Haybaeck, J. (2020). New pancreatic cancer biomarkers eIF1, eIF2D, eIF3C and eIF6 play a major role in translational control in ductal adenocarcinoma. *Anticancer Res.* 40 (6), 3109–3118. doi:10.21873/anticanres.14292

Guo, Z., Neilson, L. J., Zhong, H., Murray, P. S., Zanivan, S., and Zaidel-Bar, R. (2014). E-cadherin interactome complexity and robustness resolved by quantitative proteomics. *Sci. Signal.* 7 (354), rs7. doi:10.1126/scisignal.2005473

Hiraishi, H., Oatman, J., Haller, S. L., Blunk, L., McGivern, B., Morris, J., et al. (2014). Essential role of eIF5-mimic protein in animal development is linked to control of ATF4 expression. *Nucleic Acids Res.* 42 (16), 10321–10330. doi:10.1093/nar/gku670

Huang, L., Chen, S., Fan, H., Ai, F., and Sheng, W. (2020). BZW2 promotes the malignant progression of colorectal cancer via activating the ERK/MAPK pathway. *J. Cell. Physiol.* 235 (5), 4834–4842. doi:10.1002/jcp.29361

Jackson, R. J., Hellen, C. U., and Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11 (2), 113–127. doi:10.1038/nrm2838

Jin, X., Liao, M., Zhang, L., Yang, M., and Zhao, J. (2019). Role of the novel gene BZW2 in the development of hepatocellular carcinoma. *J. Cell. Physiol.* 234, 16592–16600. doi:10.1002/jcp.28331

Li, Z., Ge, Y., Dong, J., Wang, H., Zhao, T., Wang, X., et al. (2022). BZW1 facilitates glycolysis and promotes tumor growth in pancreatic ductal adenocarcinoma through potentiating eIF2α phosphorylation. *Gastroenterology* 162 (4), 1256–1271.e14. doi:10.1053/j.gastro.2021.12.249

Loughran, G., Firth, A. E., Atkins, J. F., and Ivanov, I. P. (2018). Translational autoregulation of BZW1 and BZW2 expression by modulating the stringency of start codon selection. *PLoS ONE* 13 (2), e0192648. doi:10.1371/journal.pone.0192648

Mitra, P., Vaughan, P. S., Stein, J. L., Stein, G. S., and van Wijnen, A. J. (2001). Purification and functional analysis of a novel leucine-zipper/nucleotide-fold protein, BZAP45, stimulating cell cycle regulated histone H4 gene transcription. *Biochemistry* 40 (35), 10693–10699. doi:10.1021/bi010529o

Mizrahi, J. D., Surana, R., Valle, J. W., and Shroff, R. T. (2020). Pancreatic cancer. *Lancet* 395 (10242), 2008–2020. doi:10.1016/S0140-6736(20)30974-0

Nachmias, B., Khan, D. H., Voisin, V., Mer, A. S., Thomas, G. E., Segev, N., et al. (2022). IPO11 regulates the nuclear import of BZW1/2 and is necessary for AML cells and stem cells. *Leukemia* 36 (5), 1283–1295. doi:10.1038/s41375-022-01513-4

Peng, J., Sun, B. F., Chen, C. Y., Zhou, J. Y., Chen, Y. S., Chen, H., et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* 29 (9), 725–738. doi:10.1038/s41422-019-0195-y

Piro, G., Simionato, F., Carbone, C., Frizziero, M., Malleo, G., Zanini, S., et al. (2017). A circulating T(H)2 cytokines profile predicts survival in patients with resectable pancreatic adenocarcinoma. *Oncoimmunology* 6 (9), e1322242. doi:10.1080/2162402X.2017.1322242

Sanda, T., Tyner, J. W., Gutierrez, A., Ngo, V. N., Glover, J., Chang, B. H., et al. (2013). TYK2-STAT1-BCL2 pathway dependence in T-cell acute lymphoblastic leukemia. *Cancer Discov.* 3 (5), 564–577. doi:10.1158/2159-8290.CD-12-0504

Shi, Z., Xiao, C., Lin, T., Wu, J., and Li, K. (2021). BZW1 promotes cell proliferation in prostate cancer by regulating TGF-β1/Smad pathway. *Cell Cycle* 20 (9), 894–902. doi:10.1080/15384101.2021.1909242

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics, 2022. *CA. Cancer J. Clin.* 72 (1), 7–33. doi:10.3322/caac.21708

Singh, C. R., Watanabe, R., Zhou, D., Jennings, M. D., Fukao, A., Lee, B., et al. (2011). Mechanisms of translational regulation by a human eIF5-mimic protein. *Nucleic Acids Res.* 39 (19), 8314–8328. doi:10.1093/nar/gkr339

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247

Truong, L. H., and Pauklin, S. (2021). Pancreatic cancer microenvironment and cellular composition: Current understandings and therapeutic approaches. *Cancers (Basel)* 13 (19), 5028. doi:10.3390/cancers13195028

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33, D433–D437. doi:10.1093/nar/gki005

Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., et al. (2022). Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin. Med. J.* 135 (5), 584–590. doi:10.1097/CM9.0000000000002108

Yin, C., Alqahtani, A., and Noel, M. S. (2022). The next frontier in pancreatic cancer: Targeting the tumor immune milieu and molecular pathways. *Cancers (Basel)* 14 (11), 2619. doi:10.3390/cancers14112619

Yu, M., Sha, H., Gao, Y., Zeng, H., Zhu, M., and Gao, X. (2006). Alternative 3' UTR polyadenylation of Bzw1 transcripts display differential translation efficiency and tissue-specific expression. *Biochem. Biophys. Res. Commun.* 345 (1), 479–485. doi:10.1016/j.bbrc.2006.04.113

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 06 October 2022
DOI 10.3389/fgene.2022.1008644

OPEN ACCESS

EDITED BY
Shun Lu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Luo Yuqi,
Shenzhen Longhua District Central Hospital, China
Ding Lei,
Beijing Shijitan Hospital, Capital Medical University, China

*CORRESPONDENCE
Shijin Yuan,
21818337@zju.edu.cn
Xian Wang,
wangx118@zju.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 09 August 2022
ACCEPTED 20 September 2022
PUBLISHED 06 October 2022

CITATION
Yuan S, Gao Y, Xia Y, Wang Z and Wang X (2022), DNA methylation regulator-mediated modification pattern defines tumor microenvironment immune infiltration landscape in colon cancer. *Front. Genet.* 13:1008644. doi: 10.3389/fgene.2022.1008644

DNA methylation regulator-mediated modification pattern defines tumor microenvironment immune infiltration landscape in colon cancer

Shijin Yuan[1*†], Yuzhen Gao[2†], Yan Xia[2], Zhuo Wang[1] and Xian Wang[1*]

[1]Department of Medical Oncology, Cancer Institute of Zhejiang University, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [2]Department of Clinical Laboratory, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Emerging evidence implies a non-negligible role of DNA methylation in tumor immunity, however, its comprehensive impact on tumor microenvironment (TME) formation and immune activation remains unclear. In this study, we integrated 24 DNA methylation regulators among 754 colon cancer patients to distinguish different modification patterns *via* an unsupervised clustering method, and explore their TME immune characteristics. Three DNA methylation modification patterns with distinct prognosis and biological behaviors were identified, consistent with three known phenotypes of immune-inflamed, immune-excluded, and immune-desert. We then determined a DNA methylation gene signature and constructed a DNA methylation score (DMS) to quantify modification patterns individually through principal component analysis algorithms. DMS-low group had characteristics of specific molecular subtypes, including microsatellite instability, CpG island methylator phenotype positive, and mutant BRAF, presented by increased mutation burden, activation of DNA damage repair and immune-related pathways, highly TME immune cells infiltration, and hence, a preferable prognosis. Further, low DMS was also demonstrated to be correlated to better response and prolonged survival of anti-PD-L1 antibody, indicating that DMS could be considered as an effective predictive tool for immunotherapy. In conclusion, our work presented a landscape of different DNA methylation modification patterns, and their vital role in the formation of TME diversity and complexity, which could help to enhance understanding of TME immune infiltration characteristics and more importantly, guide immunotherapy strategies more effectively and personalized.

KEYWORDS

DNA methylation, immunotherapy, tumor microenvironment, colon cancer, biomarker

Introduction

Colon cancer is common worldwide and remains one of the leading causes of cancer-related mortality (Sung et al., 2021). As a biologically heterogeneous disease, colon cancer derives from the accumulation of a series of genetic and epigenetic changes that transform normal glandular epithelium into malignant invasive adenocarcinoma. Now, it is appreciated that there are multiple molecular pathways involved in these genetic mutations and epigenetic alterations during colon cancer development, including microsatellite instability (MSI), the CpG island methylator phenotype (CIMP), chromosome instability (CIN), and somatic mutations of critical oncogenes like BRAF and KRAS (Nguyen et al., 2020).

Epigenetics, referring to heritable alterations in gene expression that are not dependent on changes in the DNA sequence, plays an important role in the pathogenesis of colon cancer (Lao and Grady. 2011). Therein, aberrant DNA methylation, one of the most widely studied epigenetic modifications, could lead to the dysregulation of gene expression in colon cancer. DNA methylation is a reversible modification process mediated by DNA methyltransferases (DNMTs) that facilitate the catalytic addition of methyl groups to the fifth position of the cytosine of CpG dinucleotides to generate 5-methylcytosine (5mC). On the contrary, 5mC could be reversed to unmodified cytosine through TET dioxygenase-mediated oxidation (Bestor 2000). Over the past two decades, emerging studies have elucidated the epigenetically regulatory mechanism of DNA methylation in colon cancer-specific gene expression patterns. Hypermethylation in the promoter region could silence the expression of tumor-suppressor genes, and contribute to loss of function. For instance, MSI, one of the hallmarks of molecular subtypes of colon cancer (Dienstmann et al., 2017), is the consequence of a deficiency in the DNA mismatch repair (MMR) system, which is not only due to the genetic mutation of MMR-related genes, but also the results of hypermethylation of MLH1 gene (Herman et al., 1998). On the contrary, global hypomethylation of the genome has been demonstrated to influence colon cancer development through inducing CIN and global loss of imprinting (Suter et al., 2004). These insights have improved our understanding of colon cancer pathophysiology and provide clues to discover novel biomarkers and therapeutic targets.

Recently, immunotherapy, especially the inhibitor targeting immune checkpoints like CTLA-4, PD-1, or PD-L1, has achieved durable anti-tumor activity in a range of cancer types. However, there are many patients, particularly in colon cancer with microsatellite stable (MSS), do not benefit from this advanced treatment (Le et al., 2015). The major reason is thought to lack lymphocytes infiltration in the MSS tumor microenvironment (TME), forming a "immune-desert" phenotype and resulting a weak immunoreactivity to the immunotherapy (Chen and Mellman, 2017). Increasing evidence demonstrated DNA methylation regulators mediated regulation is critical in anti-tumor immune response through involving in many processes of the cancer-immunity cycle (Cao and Yan, 2020). To be specific, DNA methylation-associated mutagenesis could generate tumor neoantigens (Alexandrov et al., 2013). In tumor cells, DNMTs could suppress the expression of MHC-I to dysregulate the antigen-presenting machinery, and suppress the expression of pro-inflammatory chemokines, such as CXCL9 and CXCL10, which are required by effector T-cells to permeate the TME and execute an immune attack (Chen et al., 2017; Luo et al., 2018). Hypermethylation of the PD-L1 promotor region inhibits its expression and leads to an inferior prognosis in various cancer types (Goltz et al., 2017; Heiland et al., 2017). Meanwhile, in immune cells, MBD2, the "reader" of methylated DNA, is necessary to induce dendritic cells phenotypic activation and then initiate the T cell response (Cook et al., 2015). By contrast, TET2 and HDAC2, the "eraser" of DNA methylation and histone acetylation respectively, coordinate to suppress IL-6 expression of dendritic cells, inhibiting the inflammatory response (Zhang et al., 2015).

However, to date, the majority of studies focus on the function of one or two DNA methylation regulators, which cannot reflect the whole landscape of DNA methylation in the formation of tumor-permissive immune environment. Therefore, comprehensive recognition of the TME immune characteristics mediated by multiple DNA methylation regulators, including TME infiltrating immune cells and activity of immune/inflammatory-related pathways, could enhance our understanding of TME immune regulation, and further provide novel perspectives for cancer immunotherapy. In this study, we integrated the transcriptomic and clinical information of 754 colon cancer samples to identify DNA methylation modification patterns with distinct TME immune characteristics, which were highly consistent with three known immune phenotypes, including immune-inflamed, immune-excluded, and immune-desert phenotype, respectively. In addition, we determined the DNA methylation gene signature and constructed a scoring system to quantify modification patterns for individual patients, which could be served as an effective biomarker for predicting the efficacy and prognosis of immunotherapy.

Materials and methods

Study design, colon cancer datasets collection, and data processing

Supplementary Figure S1A depicted the workflow of our present study. Generally, we searched publicly available transcriptomic data and clinical annotation of colon cancer samples from the Gene-Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) database. Seven eligible colon cancer datasets with comprehensive survival information [GSE39582 (n = 562), GSE38832 (n = 122), GSE39084 (n = 70), GSE72970 (n = 124), GSE103479 (n = 155),

GSE87211 (n = 196), and TCGA-colon adenocarcinoma cohort (COAD, n = 430)], including 1,659 patients, were collected for our further analysis (Supplementary Table S1). For the RNA sequencing data of the TCGA-COAD cohort, the fragments per kilobase of transcript per million mapped reads (FPKM) value of each sample was downloaded from the Genomic Data Commons (GDC, https://portal.gdc.cancer.gov/) by the "*TCGAbiolinks*" R package directly (Colaprico et al., 2016). And for the microarray data of the GSE cohort, we used the normalized matrix files downloaded from GEO. The somatic mutation data, copy number variation (CNV) data, and TCGA pan-cancer RNA sequencing data were acquired from the University of California Santa Cruz (UCSC) Xena browser (https://xenabrowser.net).

Moreover, we also included an immunotherapeutic cohort (IMvigor210, n = 348), urothelial carcinoma treated with anti-PD-L1 antibody atezolizumab, to evaluate the effect of DNA methylation modification in immunotherapy. The expression data and clinical information were available from the "*IMvigor210*" R package (Mariathasan et al., 2018).

Unsupervised clustering for 24 DNA methylation regulators

After a systematic review of published articles, we identified a total of 24 DNA methylation regulators, including 3 writers (DNMT1, DNMT3A, DNMT3B), 3 erasers (TET1, TET2, TET3), and 18 readers (MBD1, MBD2, MBD3, MBD4, ZBTB33, ZBTB38, ZBTB4, ZBTB24, UHRF1, UHRF2, MECP2, UNG, TDG, NTHL1, SMUG1, NSUN2, MGMT, DMAP1). The protein-protein interactions (PPI) network among 24 regulators were analyzed by the STRING interaction database (https://string-db.org/) (Szklarczyk et al., 2019) and visualized by the Cytoscape software (Shannon et al., 2003).

Three GEO datasets (GSE39582, GSE38832, and GSE39084) with the same microarray platform and no prognostic differences were integrated as meta-cohort (n = 754) to identify different DNA methylation modification patterns mediated by 24 regulators (Supplementary Figure S2A). The non-biological technical biases caused batch effect among each cohort was eliminated by the "*ComBat*" algorithm of the "*sva*" R package. Based on the expression of 24 regulators, the Nonnegative Matrix Factorization (NMF) method was used to determine different DNA methylation modification patterns through the "*NMF*" R package (Gaujoux and Seoighe. 2010).

Identification of differentially expressed genes and generation of DNA methylation gene signature

First, to identify DNA methylation phenotype-related genes, we used the empirical Bayesian approach of the "*limma*" R package to determine differentially expressed genes (DEGs) among three DNA methylation modification patterns (Ritchie et al., 2015). The adjusted p-value was set as < 0.01 to select significant DEGs. Then, we performed a univariate Cox regression analysis to confirm the DEGs that were significantly related to the prognosis (p-value < 0.01). Subsequently, based on the expression of these prognostic DEGs, we conducted the second NMF clustering algorithm to obtain DNA methylation gene clusters as well as validate their stability. Furthermore, through the principal component analysis (PCA) method, we used these prognostic DEGs to construct the DNA methylation gene signature, termed DNA methylation score (DMS), which could quantify the DNA methylation modification pattern for each patient. The procedure of establishing the DMS was similar to a previous study, and we added the principal component 1 and 2 to acquire the signature scores (Zhang et al., 2020; Gao et al., 2021). The formula of DMS was expressed as follow:

$$DMS = \Sigma\ (PC1_i + PC2_i)$$

where i is the expression of DNA methylation signature genes.

Gene set variation analysis, gene set enrichment analysis, and functional annotation

To investigate the difference in the biological processes among different DNA methylation modification patterns, gene clusters, and DMS groups, we performed gene set variation analysis (GSVA) and gene set enrichment analysis (GSEA) analyses through the "*GSVA*" and "*clusterProfiler*" R packages, respectively (Yu et al., 2012; Hanzelmann et al., 2013). The hallmark gene sets (h.all.v7.5.1.symbols) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets (c2.cp.kegg.v7.5.1.symbols) were downloaded from the MSigDB database for running enrichment analysis (http://www.gsea-msigdb.org/gsea/msigdb/). All enrichment p-values were adjusted by the Benjamini-Hochberg methods and less than 0.05 were considered statistically significant (Thissen et al., 2002). Moreover, we performed Gene Ontology (GO) functional annotation for DEGs *via* the "*clusterProfiler*" R package, with the cutoff value of false discovery rate (FDR) < 0.05.

We additionally collected 18 classical biological processes constructed by Mariathasan et al. (2018), including: 1) CD8 T effector; 2) DNA damage repair (DDR); 3) antigen-processing machinery (APM); 4) immune checkpoint; 5) cell cycle regulators; 6) Fanconi anemia; 7) pan-fibroblast TGFβ response signature (Pan-F-TBRS); 8) epithelial-mesenchymal transition (EMT) markers including EMT1, EMT2 and EMT3; 9) WNT targets; 10) fibroblast growth factor receptor 3 (FGFR3) related signature; 11) cell cycle signature; 12) mismatch repair

13) homologous recombination; 14) nucleotide excision repair; 15) DNA replication; 16) base excision repair. The detailed gene set of the corresponding pathway was provided in Supplementary Table S2.

Estimation of the tumor microenvironment infiltrating cell abundance and the immune infiltration score

We used three different algorithms, including the single-sample gene-set enrichment analysis (ssGSEA) algorithm of the "*GSVA*" R package, the CIBERSORT method, and the Tumor Immune Estimation Resource (TIMER) database, to evaluate the infiltrating abundance of various TME immune cells, such as B cell, CD8+ T cell, dendritic cell, and macrophage et al. The gene sets of each type TME infiltrating cell were extracted from the study of Charoentong and listed in Supplementary Table S3 (Charoentong et al., 2017).

Besides, through applying "*xCell*" and "*ESTIMATE*" methods, we calculated the TME stromal score, immune score, estimate score, and microenvironment score based on signature gene expression to infer the fraction of stromal and immune cells in colon cancer samples. The calculation was performed by the "*xCell*" and "*estimate*" R packages, respectively (Yoshihara et al., 2013; Aran et al., 2017).

Cell culture

The colorectal cancer cell line HCT116 and normal colonic epithelial cell line NCM460 were purchased from Type Culture Collection of the Chinese Academy of Science (Shanghai, China). HCT116 cells were cultured in McCoy's 5A Medium (16600-082, Gibco) with 10% fetal bovine serum (10099141C, Gibco) in 5% CO_2 at 37°C. NCM460 cells were cultured in RPMI 1640 Medium (118575-093, Invitrogen) with same condition.

RNA extraction and real-time quantitative reverse transcription-polymerase chain reaction)

Total RNA was isolated using TRIzol™ Reagent (15596026, Invitrogen) and quantified with a NanoDrop 2000™ (Thermo Fisher Scientific, United States). 1 μg RNA was used for the reverse transcription reaction to generate cDNA through the PrimeScript™ RT Reagent Kit with gDNA Eraser (RR047A, TaKaRa) according to the manufacturer's protocols. The mRNA expression was determined by rt-qPCR, which was performed using Ultra SYBR Mixture (CW0957M, CWBIO) and a LightCycler® 480 II system (Roche, Shanghai, China). The mRNA expression of ACTB was used as a reference. The primers used in this study were listed in Supplementary Table S4.

Statistical analysis

The normality of data was tested by the Shapiro-Wilk test. For the comparison of two groups, we used the t-test to detect the significant difference between normally distributed data, and the Wilcoxon test for skewed distributed data. For the comparison of three or more groups, one-way ANOVA and Kruskal-Wallis tests were conducted to detect the significant difference between normal distributed and skewed distributed data, respectively. A chi-squared test was used to compare the frequency differences between the two groups. Correlation coefficients were calculated by the Spearman and distance correlation analyses.

For the survival analysis, we focused on the overall survival (OS) and recurrence-free survival (RFS), and we obtained the best cut-off value through the "*survcutpoint*" function of the "*survminer*" R package. Kaplan-Meier method was used to depict the survival curves, and the log-rank test was utilized to identify significant survival differences between groups. Univariate Cox proportional hazards regression model was applied to calculate the hazard ratio (HR) and 95% confidence interval (95% CI) for each DNA methylation regulator and DNA methylation related gene. And the multivariate Cox model was performed to determine the independent prognostic factors when adjusted by clinical characteristics. The results of Cox regression analyses were visualized by the "*forestplot*" R package. The prediction performance of DMS to evaluate the OS probability at distinct times was assessed by the receiver operating characteristic (ROC) curves and quantified by the area under the curve (AUC), which were conducted *via* the "*timeROC*" R package.

All statistical analyses were accomplished in R 3.6.1 software, and all reported *p*-values were two-sided, with p-value < 0.05 as statistically significant.

Results

Multi-omics landscape analysis of DNA methylation regulators in colon cancer

In this study, we identified 24 DNA methylation regulators, including 3 writers, 3 erasers, and 18 readers. Figure 1A summarized the dynamic reversible DNA methylation modification process of these regulators. Firstly, we analyzed the incidence of somatic mutation and CNV of 24 regulators in colon cancer from the genomic perspective. We visualized the mutation landscape of 24 regulators through the "*waterfall*" function of the "*maftools*" R package. Among 399 available samples in the TCGA-COAD cohort, a total of 113 (28.3%)

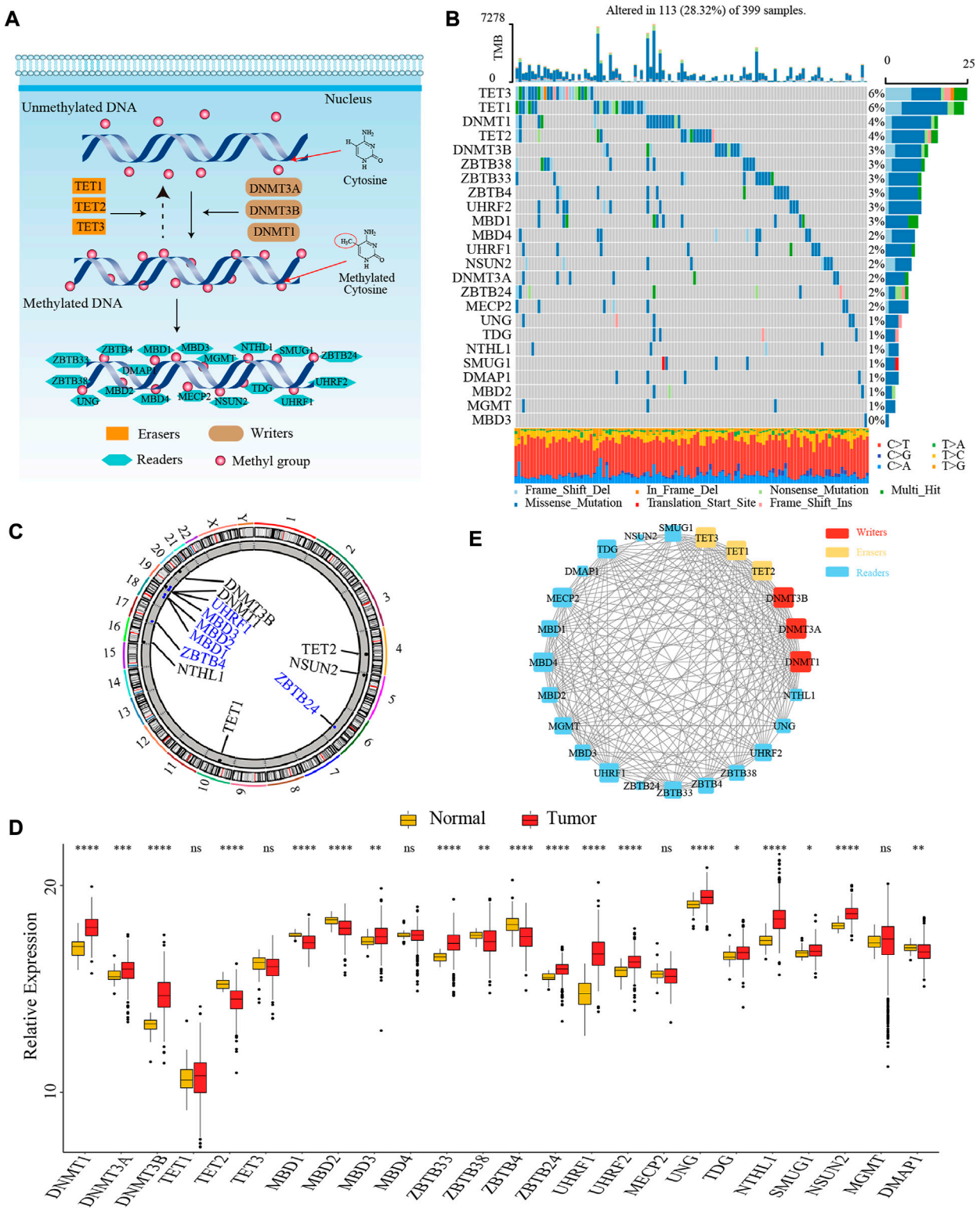


FIGURE 1
The landscape of genetic and transcriptomic variations of 24 DNA methylation regulators in colon cancer. **(A)** The dynamic reversible DNA methylation modification process of 24 regulators. **(B)** The mutation frequency of 24 DNA methylation regulators in the TCGA-COAD cohort. **(C)** The location of CNV alteration of DNA methylation regulators on 23 chromosomes from TCGA-COAD cohort. **(D)** The expression of 24 DNA methylation regulators between colon cancer tissue and normal tissue (Wilcoxon test; ns: no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(E)** The PPI network among 24 DNA methylation regulators.

mutations occurred in 24 regulators (Figure 1B). Therein, 2 erasers, TET3 and TET1, had the highest mutation frequency (6%) with missense mutation as a major mutation type, while MBD3 did not present any mutations. Besides, we found significant co-occurrence mutation relationships between several regulators, such as TET3 and ZBTB38, DNMT1 and MBD1, along with UHRF2 and MECP2 (Supplementary Figure S1B). In addition, the CNV analysis revealed an amplification of copy number in 6 regulators and deletion in other 6 regulators. The location of CNV alterations of 12 regulators on the chromosomes was shown in Figure 1C, which was accomplished by the "*RCircos*" R package.

Secondly, to ascertain whether the above genomic variations affected the transcription of 24 regulators, we further compared the mRNA expression level of these regulators between colon cancer and normal samples from the transcriptomic perspective. We observed most of the regulators were significantly differentially expressed between tumor and normal tissues. Interestingly, compared to normal tissues, we found that many regulators with amplified CNV had a markedly higher expression level in colon cancer tissues (e.g., DNMT1 and DNMT3B), and vice versa (e.g., MBD2 and ZBTB4), suggesting that the alteration of CNV might be the prominent factor resulting in the abnormal expression of DNA methylation regulators (Figures 1C,D). To further confirm the expression level of these regulators in colorectal cancer cell line and normal colonic epithelial cell line, we selected three most significantly differentially expressed regulators between tumor and normal tissues, DNMT3B (writer), TET2 (eraser), and UHRF1 (reader), to compare their mRNA expression in HCT116 and NCM460 cells. The results of quantitative reverse transcription-polymerase chain reaction (qRT-PCR) showed that the mRNA level of DNMT3B and UHRF1 were significantly higher in HCT116 cells while the mRNA level of TET2 was significantly higher in NCM460 cells (Supplementary Figures S1C–E).

Thirdly, we explored the interaction relationship between 24 regulators from the proteomics perspective. The PPI network depicted the extensive protein interactions among these regulators (Figure 1E). Comprehensively, the above analyses presented a highly genetic heterogeneity and expressional anomalism of DNA methylation regulators in colon cancer from a multi-omics landscape perspective, hinting at a critical role of these regulators in tumorigenesis of colon cancer.

Prognosis and immune characteristics of 24 DNA methylation regulators in colon cancer

To clarify the role of 24 DNA methylation regulators in colon cancer clinical prognosis and TME cell infiltration characterization, we gathered three GEO datasets (GSE39582, GSE38832, and GSE39084) without prognostic differences as meta-cohort for further analyses (Supplementary Figure S2A). The Cox regression analysis identified high expression of nearly half of the regulators were associated with a favorable prognosis (Supplementary Figure S2B). Moreover, the interaction network of regulators visualized the correlation of regulators' expression and their prognostic significance (Figure 2A; Supplementary Table S5). We found that the expression of most regulators was positively related to each other, not only in regulators with the same functional category, but also among writers, erasers, and readers.

TME infiltrating immune cells had been widely reported to display an epigenomic reprogramming, especially in aberrant DNA methylation (Loo Yau et al., 2019). Therefore, we investigated the correlation between 24 regulators and TME infiltrating cells. The expression of MBD1, MBD2, and ZBTB4 was positively correlated with infiltrating abundance of majority immune cells, which could explain their favorable prognostic value to some extent (Figure 2B). A previous study demonstrated that colon cancer patients with a MSI status presented an active intra-tumoral immune environment, and hence, had a robust response to immunotherapy and superior prognosis (Andre et al., 2020). Similarly, we found the expression of MBD1 and MBD2 was upregulated in MSI colon cancer patients, implying both regulators might involve in immune activation (Supplementary Figure S2C). In consideration of the MBD2 expression were highly correlated to immune cells quantity, especially for activated CD4+ and CD8+ T-cells, as well as its preferable prognostic value, we next thoroughly analyzed the role of MBD2 in tumor immune environment formation. First, the MBD2 high expression group exhibited a higher infiltrating abundance of most immune cells and TME immune score (Figures 2C,D). Second, GSEA analysis revealed several activated immune and inflammation-related pathways, such as the TNFα, interferon- and interleukin-mediated signaling pathways were significantly enriched in the MBD2 high expression group (Figure 2E). Third, owing to the vital role of MBD2 in tumor immunity, we additionally investigated whether its expression could predict the efficacy and prognosis of immunotherapy. In the IMvigor210 cohort, we observed a marked survival benefit in patients with MBD2 high expression, although the therapeutic response was similar between two groups (Figure 2F; Supplementary Figure S2D).

Collectively, above results exhibited crosstalk among 24 DNA methylation regulators and their significant impact on tumor immunity. The MBD2 expression was positively correlated to TME immune cells infiltration and might be a potential prognostic biomarker in immunotherapy.

DNA methylation modification patterns mediated by 24 regulators in colon cancer

As the markedly different clinical outcomes and TME infiltrating cells characterizations among 24 DNA methylation regulators in colon cancer, we speculated that these regulators

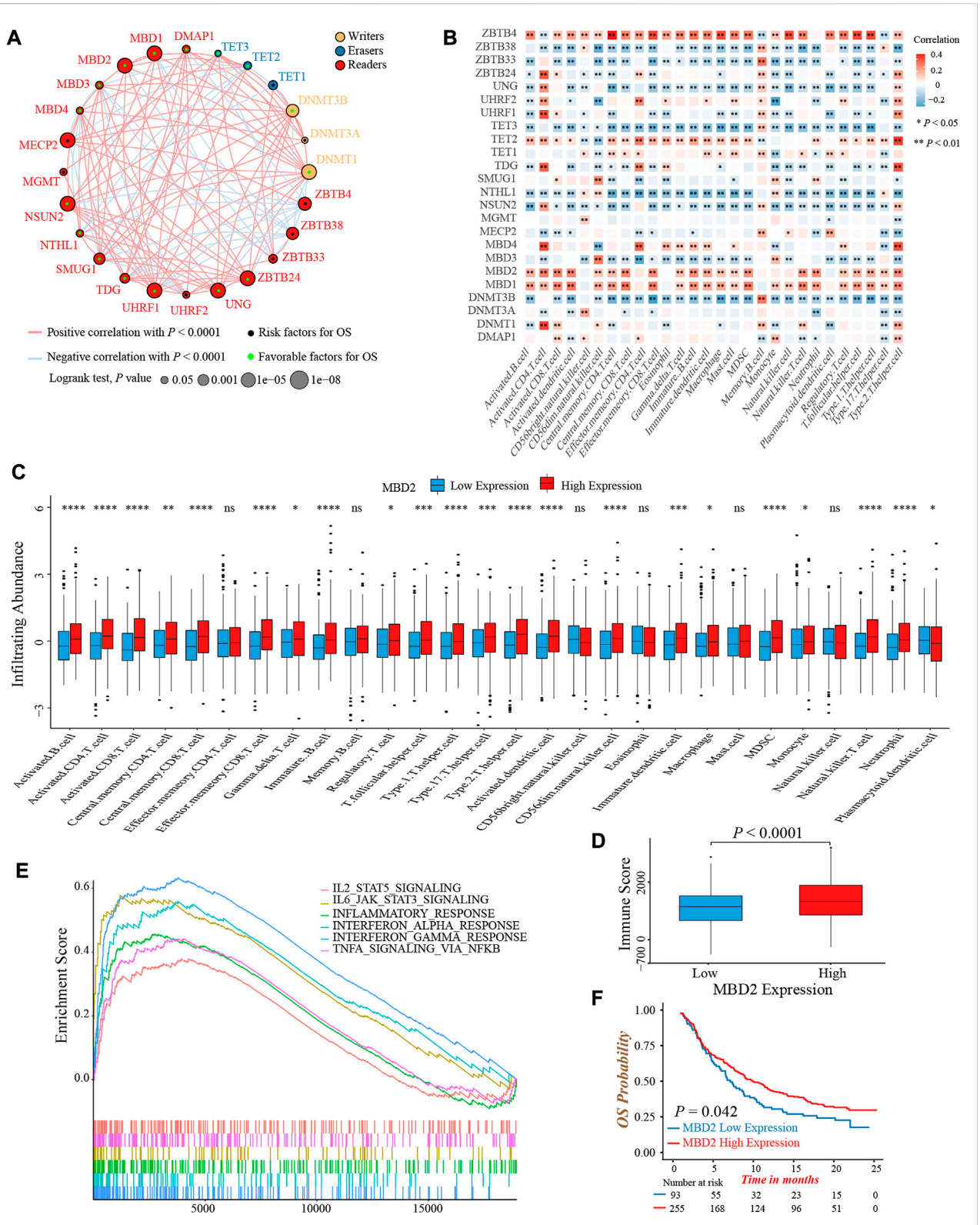


FIGURE 2
Prognosis and immune characteristics of 24 DNA methylation regulators in colon cancer. **(A)** The prognosis and correlation among 24 DNA methylation regulators in colon cancer (Log-rank test and Spearman correlation analysis). **(B)** The correlation between 28 TME infiltrating cell types and 24 DNA methylation regulators (Spearman correlation analysis). **(C)** The infiltrating abundance of 28 TME cell types between the MBD2 high and low expression groups (Wilcoxon test; ns: no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(D)** Comparison of immune score between the MBD2 high and low expression groups. (Wilcoxon test; $p < 0.0001$). **(E)** GSEA analysis indicated six activated immune/inflammation-related pathways that were enriched in the MBD2 high expression group (All $p < 0.05$). **(F)** Overall survival analysis of high and low MBD2 expression groups in the IMvigor210 cohort (Log-rank test, $p = 0.042$).

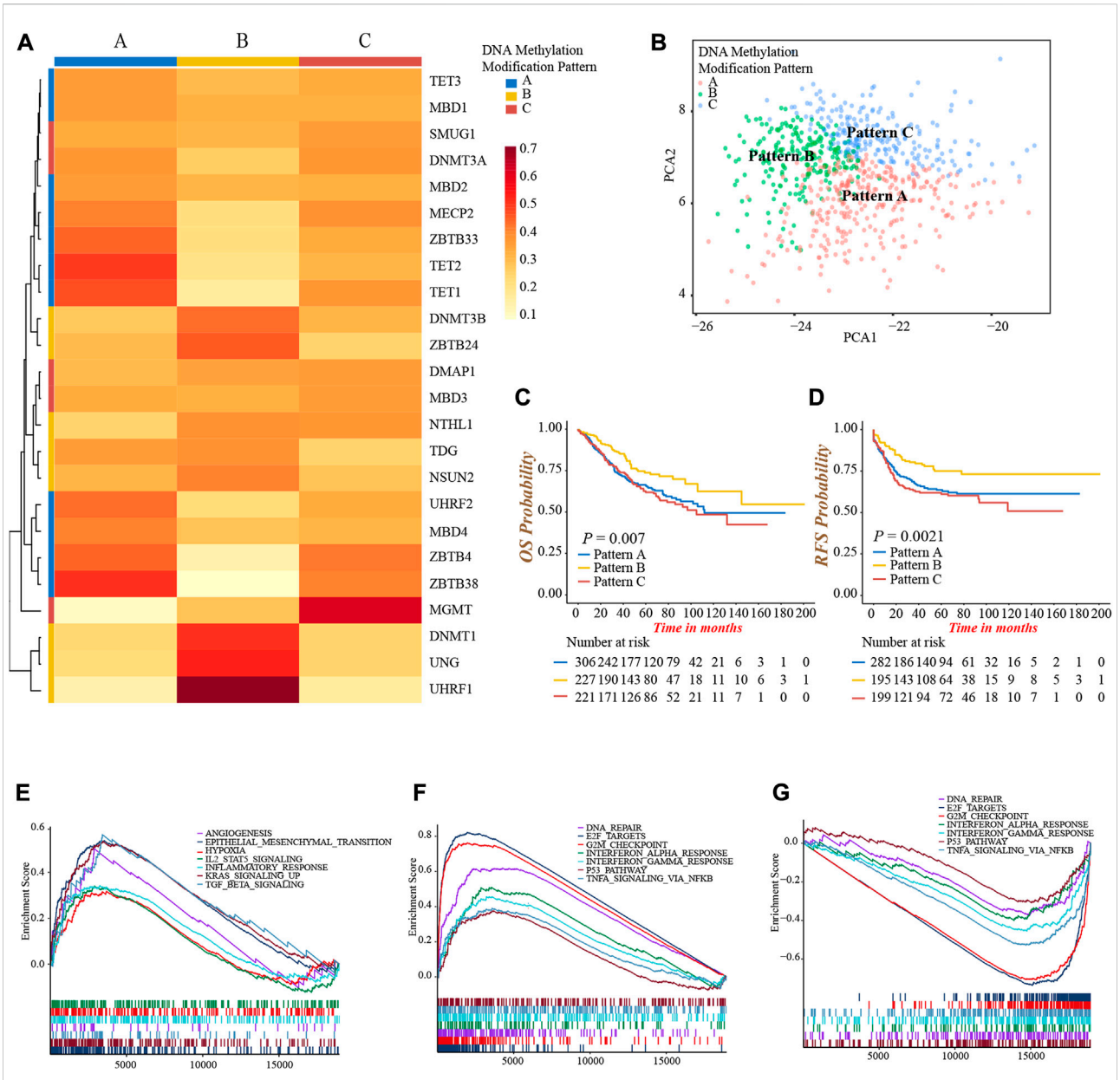


FIGURE 3
Identification of DNA methylation modification patterns and prognostic and pathway analyses. **(A)** Unsupervised clustering of 24 DNA methylation regulators in meta-cohort. **(B)** PCA graph for the transcriptomic profiles of three DNA methylation modification patterns. **(C)** Overall survival analysis for the three DNA methylation modification patterns (Log-rank test, p = 0.007). **(D)** Recurrence-free survival analysis for the three DNA methylation modification patterns (Log-rank test, p = 0.0021). **(E–G)** GSEA enrichment analyses presented the activation status of biological pathways among distinct DNA methylation modification patterns. **(E)** Pattern-A; **(F)** Pattern-B; **(G)** Pattern-C.

may contribute to forming different DNA methylation modification patterns in individual colon cancer. Based on the expression of 24 regulators, we applied the NMF unsupervised clustering to obtain three distinct clusters, including DNA methylation modification pattern-A (n = 306), pattern-B (n = 227), and pattern-C (n = 221), in meta-cohort (Supplementary Figures S3A,B). Accordingly, we found different regulators were differentially expressed in different patterns. Pattern-A was characterized by high expression of TET1, TET2, ZBTB4, ZBTB33, ZBTB38, MBD4, UHRF2, and MECP2; pattern-B was characterized by high expression of DNMT1, DNMT3B, ZBTB24, UHRF1, and UNG; and pattern-C was characterized by high expression of DMAP1, SMUG1 and MGMT (Figure 3A; Supplementary Figure S3C). In addition, the graph of principal component analysis (PCA) showed that three patterns were distinctly segregated, indicating a complete distinguishment of

three patterns (Figure 3B). Moreover, the survival analysis revealed a notably favorable OS and RFS in pattern-B (Figures 3C,D).

Then, we performed a GSEA analysis to explore the different hallmark signaling pathways among three patterns (Supplementary Table S6). We found pattern-A presented the activated enrichment pathways associated with stromal and carcinogenic activation, such as the KRAS, TGFβ, EMT, and angiogenesis pathways, as well as the highest stromal score (Figure 3E; Supplementary Figure S3D). Immune regulation and tumor suppressor-related pathways were significantly enriched in pattern-B, including interferon-α, and -γ, TNFα, P53 signaling, G2M checkpoint, and DNA damage repair related pathways (Figure 3F). In contrast, these pathways were all inhibited in pattern-C (Figure 3G). We subsequently compared the abundance of TME infiltrating cells among three patterns. Not surprisingly, high infiltration of activated CD4+ and CD8+ T-cells were observed in pattern-B (Supplementary Figure S3E; Supplementary Table S7), which was consistent with the results of enrichment into activated immune-related pathways, and may explain the preferable prognosis (Figures 3C,D). Meanwhile, several innate immune cells were remarkably enriched in pattern-A, including eosinophil, mast cell, and plasmacytoid dendritic cell, while the content of the majority of immune cells was relatively lower in pattern-C (Supplementary Figure S3E).

From the above analyses, we were surprised to find three DNA methylation modification patterns exhibited significantly distinct biological behaviors and TME infiltrating cell characterizations, in accordance with the conception of three tumor immune phenotypes (Hegde and Chen, 2020). Pattern-A was classified as immune-excluded phenotype, characterized by innate immune cell infiltration and stromal activation; pattern-B was classified as immune-inflamed phenotype, characterized by a high population of adaptive immune cell infiltration and immune activation; pattern-C was classified as immune-desert phenotype, characterized by few infiltrations of immune cells and immune suppression.

Generation of DNA methylation gene signature

To further investigate the heterogeneity of three DNA methylation modification patterns, we identified 249 DEGs among three patterns (Supplementary Figures S4A,B). The GO functional annotation showed that these DEGs were significantly enriched in pathways related to DNA modification and damage repair events (Figure 4A; Supplementary Table S8), indicating the differences in clinical outcomes and TME characterizations among three patterns might result from these DEGs. Subsequently, we obtained 152 genes associated with prognosis through univariate Cox regression analysis (Figure 4B; Supplementary Table S9). Furthermore, based on the expression of 152 genes, we conducted unsupervised clustering to classify patients into different genomic subtypes. Surprisingly, consistent with three modification patterns, the NMF algorithm still clustered three distinct DNA methylation genomic phenotypes, named DNA methylation gene cluster-A (n = 272), cluster-B (n = 183), and cluster-C (n = 299), respectively (Figure 4C; Supplementary Figures S4C,D). The plot of PCA also presented a distinct separation of three gene clusters (Supplementary Figure S4E). As well, the prominent differences in the expression of 24 DNA methylation regulators were observed among three gene clusters, in line with the results of three modification patterns (Supplementary Figure S4F). In accordance with the feature of modification pattern-B, gene cluster-C had a superior prognosis and the highest CD8+ T-cells infiltrating abundance than other two clusters, demonstrating gene cluster-C was immune-inflamed phenotype (Figures 4D–F). Above all, these analyses verified the perspective that there were indeed three disparate immune phenotypes in colon cancer, representing different clinical and TME immune characterizations.

Characterization of DNA methylation score in different clinical and molecular traits

The above results demonstrated DNA methylation modification played a non-negligible role in TME formation and prognosis of colon cancer patients. Nevertheless, the remarkable heterogeneity between individual tumors limited the accurate application of DNA methylation modification patterns. Therefore, based on 152 prognostic DEGs, we constructed a DNA methylation gene signature, termed as DNA methylation score (DMS) to quantify the DNA methylation modification pattern of each patient.

We first evaluated the prognostic value of DMS in colon cancer, and determined −73.8 as the cut-off value to divide patients into low and high DMS groups. Prognostic analysis showed patients with low DMS had a prominently prolonged survival and recurrence-free time (Figures 5A,B). Besides, the multivariate Cox regression analysis demonstrated DMS was an independent risk prognostic biomarker for colon cancer (HR was 1.70 for OS and 1.66 for RFS; Supplementary Figure S5A). Additionally, to detect the stability of DMS, we enrolled other independent colon cancer cohorts to validate the prognostic value of DMS. Likewise, the DMS-low group had a preferable survival in the TCGA-COAD cohort, GSE72970, GSE103479, and GSE87211 cohorts (all $p < 0.05$, Supplementary Figures S5B–G). Furthermore, we continued to expand the application of DMS in other digestive system tumors, and we found rectum adenocarcinoma and stomach adenocarcinoma patients with low DMS presented a remarkably longer survival (both $p < 0.05$,

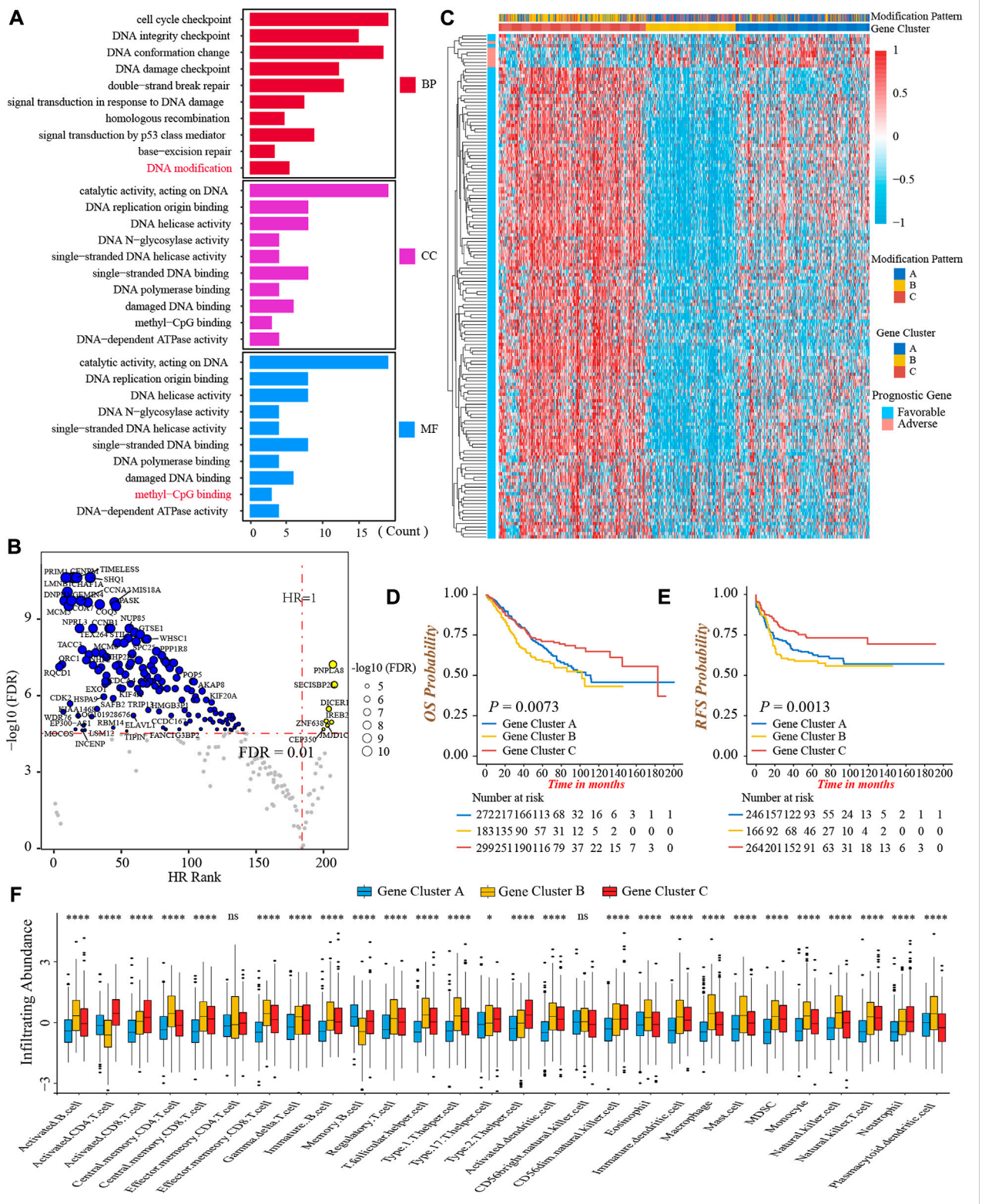


FIGURE 4
Generation of DNA methylation gene signature. **(A)** Functional annotation of 249 DEGs among three DNA methylation modification patterns using GO enrichment analysis. BP, biological process; CC, cellular component; MF, molecular function. **(B)** Volcano Plot presented the DEGs significantly associated with prognosis. **(C)** The heatmap of 152 prognostic DEGs' expressions among three DNA methylation gene clusters-A, -B, and -C. **(D)** Overall survival analysis for the three DNA methylation gene clusters (Log-rank test, $p = 0.0073$). **(E)** Recurrence-free survival analysis for the three DNA methylation gene clusters (Log-rank test, $p = 0.0013$). **(F)** The infiltrating abundance of 28 TME cell types among three DNA methylation gene clusters (Kruskal-Wallis test; ns, no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$).

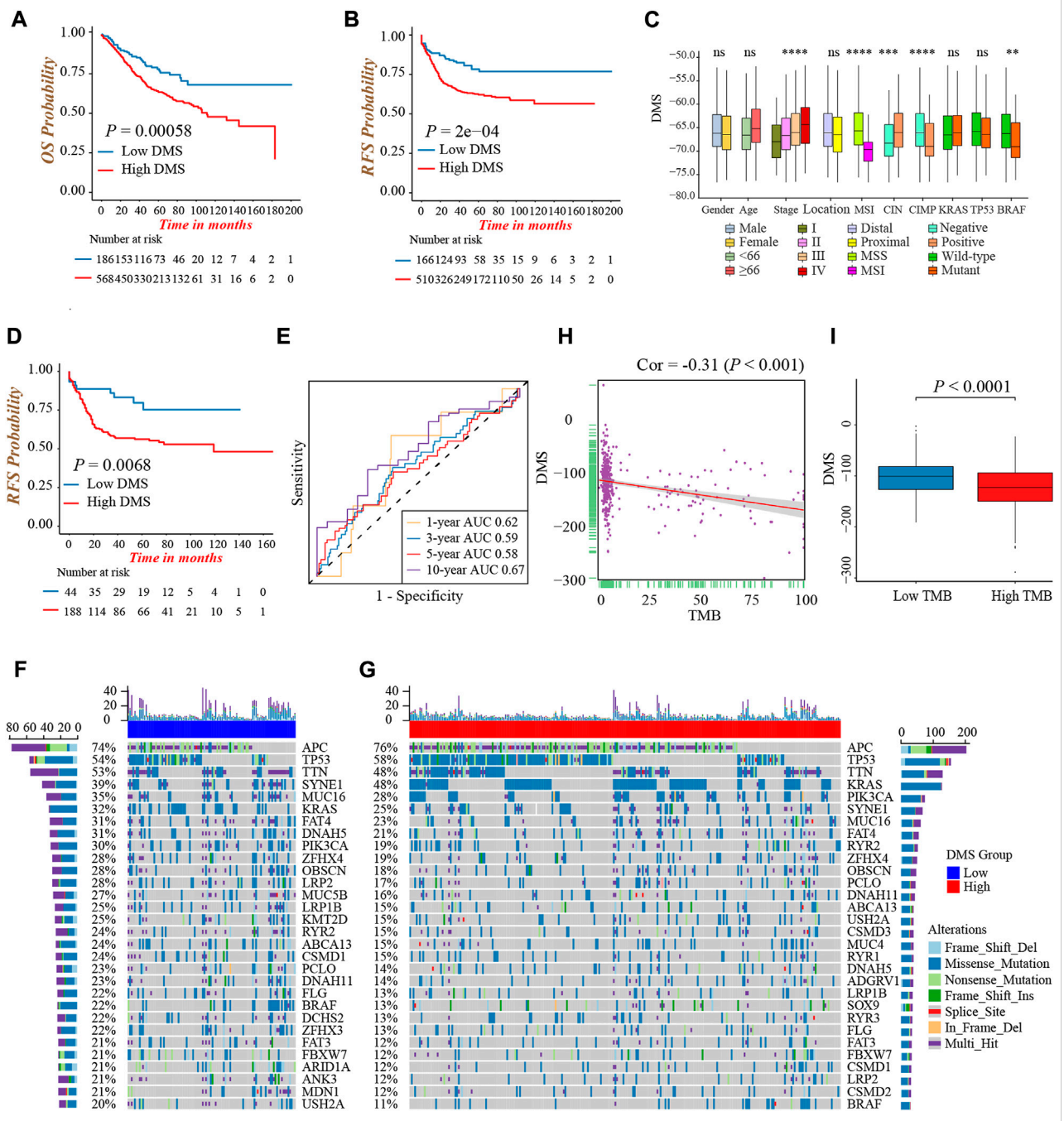


FIGURE 5
Characterization of DNA methylation score in different clinical and molecular traits. **(A)** Overall survival analysis for low and high DMS groups in the meta-cohort (Log-rank test, $p = 0.00058$). **(B)** Recurrence-free survival analysis for low and high DMS groups in the meta-cohort (Log-rank test, $p = 0.0002$). **(C)** Difference in DMS among distinct clinical subgroups in the GSE39582 cohort. MSI, microsatellite instability; CIMP, CpG island methylator phenotype; CIN, chromosome instability. (Wilcoxon test; ns: no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(D)** Recurrence-free survival analysis for low and high DMS groups among patients received chemotherapy in the GSE39582 cohort (Log-rank test, $p = 0.0068$). **(E)** ROC curves to reflect the ability of DMS to predict the 1-, 3-, 5-, and 10-year survival probability for patients received chemotherapy in the GSE39582 cohort. **(F,G)** The waterfall plot presented the distribution differences of tumor somatic mutation between DMS-low **(F)** and -high groups **(G)**. **(H)** The scatter plot depicted the negative correlation between DMS and TMB in the TCGA-COAD cohort (Spearman correlation analysis; $p < 0.001$). **(I)** Comparison of DMS between the TMB high and low groups (Wilcoxon test; $p < 0.0001$).

Supplementary Figures S5H,I). The data above proved that the DMS we constructed could predict the survival of colon cancer effectively and stably.

Next, we explored the characteristics of DMS in different clinical and molecular subtype traits and fixed our attention on the GSE39582 cohort, which had comprehensive clinical information. We found DMS rose gradually by increasing the tumor TNM stage ($p < 0.001$, Figure 5C). Besides, patients with MSI, CIMP positive, CIN negative, and mutant BRAF showed a lower DMS, implying these subtypes may have a better prognosis (all $p < 0.05$, Figure 5C). Chemotherapy was a standard treatment strategy for colon cancer, and hence, we specifically examined the ability of DMS to predict the efficacy of chemotherapy. We revealed significantly longer recurrence-free time in the DMS-low group of colon cancer patients undergoing chemotherapy (Figure 5D), and the ROC curves to reflect the ability of DMS to predict the 1-, 3-, 5-, and 10-year survival probability were shown in Figure 5E.

Then, we investigated the distribution differences of somatic mutation between DMS-low and -high groups. As shown in Figures 5F,G, the DMS-low group exhibited a higher mutation frequency, and APC, TP53, and TTN were three common genes with the highest mutation frequency in both groups. As well, we quantified tumor mutation burden (TMB) value, and discovered a markedly negative correlation between DMS and TMB (Figure 5H). Same as the relation between MSI status and DMS, the TMB-high group presented a lower DMS, hinting low DMS may reflect genomic instability, massive tumor neoantigen generation, and induce immune activation (Figure 5I). Generally, these accumulated results manifested DMS was an effective predictive tool, and distinguished different molecular subtypes of colon cancer.

The role of DNA methylation score in the tumor microenvironment and biological processes

To better illustrate the underlying relevance between DMS and different molecular traits of colon cancer, we first compared DMS among different DNA methylation modification patterns and gene clusters. Pattern-B and gene cluster-C, representing the immune-inflamed phenotype, both had the lowest median DMS (Figures 6A,B). A similar distribution trend of DMS among three patterns and gene clusters was shown in Supplementary Figures S6A,B when we split the meta-cohort into three separate GSE cohorts. These results verified low DMS was related to better survival again, and the Sankey diagram visualized the distribution of individual colon cancer patients in different DNA methylation modification patterns, gene clusters, DMS, and clinical outcomes (Figure 6C).

Subsequently, we analyzed the correlation between DMS and several known biological processes signatures constructed by Mariathasan et al. (2018). The results revealed DMS was mightily negatively correlated with DNA damage repair-related signatures (Supplementary Figure S6C). Further comparative analysis demonstrated a higher activity of immune and DNA damage repair pathways in the DMS-low group, including CD8+ T effector, immune checkpoint, and mismatch repair (Figure 6D). Additional GSVA analysis revealed activated DNA damage repair and tumor suppresser-related pathways were significantly enriched in the DMS-low group, while stromal and carcinogenic activation pathways were markedly enriched in the DMS-high group (Figure 6E). Above analyses confirmed DNA methylation patterns participated in these biological processes and the DMS could reflect the activity of immune and stromal-related pathways.

We then examined the relationship between TME infiltrating cells and DMS using different immunocytes associated algorithms, including ssGSEA, CIBERSORT, and TIMER database, and we found the majority of immune cells, especially CD8+ T cells, presented a high infiltrating abundance in the DMS-low group (Figure 6F; Supplementary Figures S6G,H). Besides, the tumor microenvironment score, estimate score, and stromal score showed a positive correlation with DMS (Figure 6G; Supplementary Figures S6D,E). Furthermore, we compared the expression level of 24 DNA methylation regulators between DMS-low and -high groups and revealed that some regulators positively associated with immune cells were also highly expressed in the DMS-low group, such as TET2, MBD1, and MBD2 (Supplementary Figure S6F). At last, we investigated several molecules, including chemokines and cytokines associated with immune activation or suppression, which were referenced from published literature (Zeng et al., 2019). Of these molecules, CD8A, CXCL10, CXCL9, GZMA, GZMB, IFNG, PRF1, TBX2, and TNF were considered to be correlated to the transcripts of immune activation; and CD80, CD86, CTLA-4, HAVCR2, IDO1, LAG3, PD-1, PD-L1, PD-L2, TIGIT, and TNFRSF9 were related to the transcripts of immune checkpoints. We still found the expression of mRNAs relevant to immune activation was significantly upregulated in the DMS-low group (Supplementary Figure S6I). While two immune checkpoints, PD-1 and PD-L2, were highly expressed in the DMS-high group (Supplementary Figure S6J). Collectively, these results demonstrated DNA methylation involved in TME immunity response. Then combining the above conclusions that low DMS was related to higher TMB, we speculated that DNA methylation modification might influence the genetic mutation and generate a great number of tumoral neoantigens, inducing immune activation. These results provided a novel perspective to explore the mechanism of DNA methylation in tumor somatic mutation, and indirectly hinting DMS may play an important role in tumor immunotherapy.

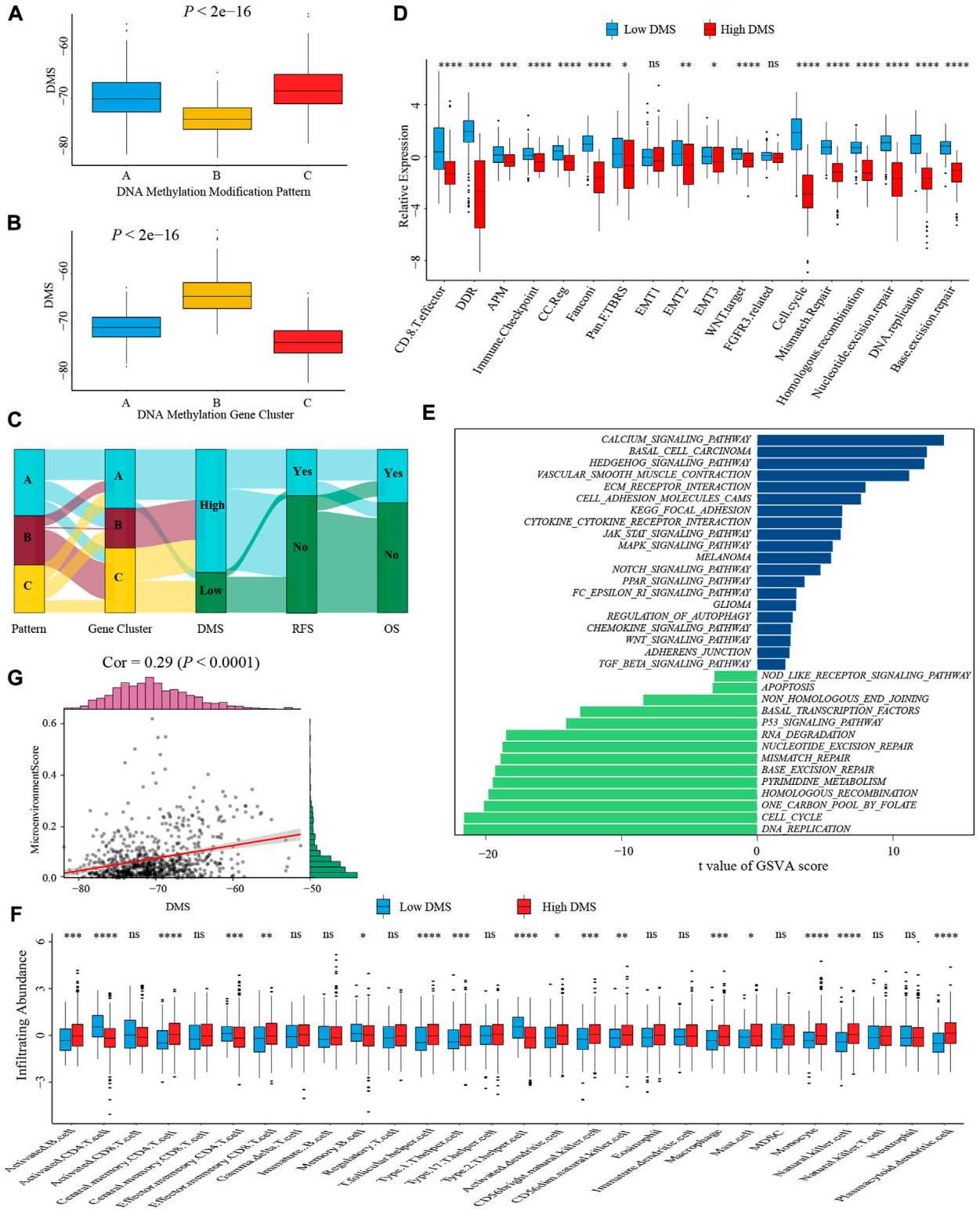


FIGURE 6
Characterization of DMS in TME and biological processes. **(A)** Differences in DMS among three DNA methylation modification patterns in the meta-cohort (Kruskal-Wallis test; $p < 0.001$). **(B)** Differences in DMS among three DNA methylation gene clusters in the meta-cohort (Kruskal-Wallis test; $p < 0.001$). **(C)** Sankey diagram visualized the distribution of individual colon cancer patients in different DNA methylation modification patterns, gene clusters, DMS, and clinical outcomes. **(D)** The differences in biological processes signatures between DMS-high and -low groups

(*Continued*)

FIGURE 6 (Continued)
(Wilcoxon test; ns: no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(E)** GSVA analysis of relatively activated KEGG pathways between DMS-high and -low groups. Blue bars represented activated pathways in the DMS-high group, and green bars represented activated pathways in the DMS-low group. **(F)** The ssGSEA method identified infiltrating abundance of 28 TME cell types between DMS-high and -low groups (Wilcoxon test; ns: no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(G)** The scatter plot depicted the positive correlation between DMS and microenvironment score (Spearman correlation analysis; $p < 0.0001$).

The predictive value of DNA methylation score in anti-PD-1/PD-L1 immunotherapy

Immunotherapy represented by anti-PD-1/PD-L1 or CTLA-4 antibody had broadened the field of cancer treatment and brought huge clinical benefits in recent years. Here, we explored whether DMS could predict the therapeutic response and prognosis of patients treated with immunotherapy. In the IMvigor210 cohort, the DMS-low group presented a remarkably prolonged survival, and the multivariate Cox regression analysis also determined that higher DMS was an independent risk factor for prognosis (Figure 7A; Supplementary Figure S7A). The ROC curves for DMS in predicting 12-, 18-, 24-months survival probability were presented in Supplementary Figure S7B. The objective response rate (percentage of complete response and partial response) of patients with low DMS was significantly higher (Figure 7B). As well, patients in the response group showed a lower DMS compared to the non-response group, indicating a clinical benefit and treatment advantage of immunotherapy in patients with low DMS (Figure 7C; Supplementary Figure S7C).

Although we did not find a significant difference in PD-L1 expression level, a potential biomarker for immunotherapy, between DMS-low and -high groups in IMvigor210 cohort (Supplementary Figure S7D), the pan-cancer analysis demonstrated DMS was negatively related to PD-L1 and PD-L2 expression in the majority of cancer types including colon cancer, indirectly proving the reliability of DMS to predict the efficacy of immunotherapy (Figure 7D; Supplementary Figure S7E). In this immunotherapeutic cohort, the TMB value was also higher in the DMS-low group, which confirmed our above conclusions again (Figure 7E). In addition, we evaluated the prognostic value of TMB and DMS combination in immunotherapy, and observed a marked prognosis benefit in patients with low DMS and high TMB (Figure 7F).

More importantly, the biological processes signatures analyses showed that DNA damage repair-related pathways were significantly activated in the DMS-low group, while EMT and Pan-F-TBRS pathways were highly activated in the DMS-high group, indicating DMS was closely related to the DNA damage repair and stromal signatures in the setting of patients receiving immunotherapy (Figure 7G). Lastly, we investigated the difference of DMS among different immune phenotypes identified by the IMvigor210 study, and found the immune-excluded and -desert phenotypes had higher DMS, implying the response to immunotherapy of these subtypes was limited (Figure 7H). In summary, our work demonstrated a significant correlation between DNA methylation modification patterns and tumor immune phenotypes, and DMS could help to predict the response of immunotherapy.

Discussion

In this study, based on 24 DNA methylation regulators, we first identified three DNA methylation modification patterns with distinct TME infiltrating characteristics and biological behaviors in colon cancer. Moreover, we obtained prognostic DEGs among three modification patterns and established the DNA methylation gene signature, termed DMS, to quantify the DNA methylation modification profile of individual colon cancer, and more importantly, predict the efficacy and clinical outcome of immunotherapy.

Increasing evidence revealed tumors commonly hijacked various epigenetic mechanisms to escape the supervision of the immune system. Particularly, certain regulator mediated DNA methylation and demethylation played an indispensable role in adaptive immune response, including generation of tumoral neoantigen, dysregulation of antigen-presenting machinery, and suppression of anti-tumor cytokine production (Cao and Yan, 2020). However, the analysis of the whole landscape of DNA methylation modification in colon cancer was limited, and its impact on TME immune response remained unclear. Here, we integrated the transcriptomic information of 24 DNA methylation regulators and revealed three distinct DNA methylation modification patterns, and their features were consistent with three classical tumor immunophenotypes: immune-inflamed, -excluded, and -desert (Hegde and Chen, 2020). Modification pattern-B was immune-inflamed phenotype and presented an activated immune status, characterized by activation of immune, tumor suppressor and DNA damage repair related pathways, as well as a high infiltrating abundance of activated CD8+ T cell. Accordingly, patients in pattern-B had a better prognosis than the other two patterns. In detail, abundant immune cells were positioned in proximity to the tumor cells, accompanied by many proinflammatory and effector chemokines and cytokines, suggesting the presence of pre-existing anti-tumor immunity in the tumor parenchyma (Harlin et al., 2009; Gajewski et al., 2013; Herbst et al., 2014). Mechanically, an inflammatory TME was the basis of immune-inflamed phenotype, also known as a hot tumor, containing pro-inflammatory cytokines which provided a more favorable condition for T cell activation and

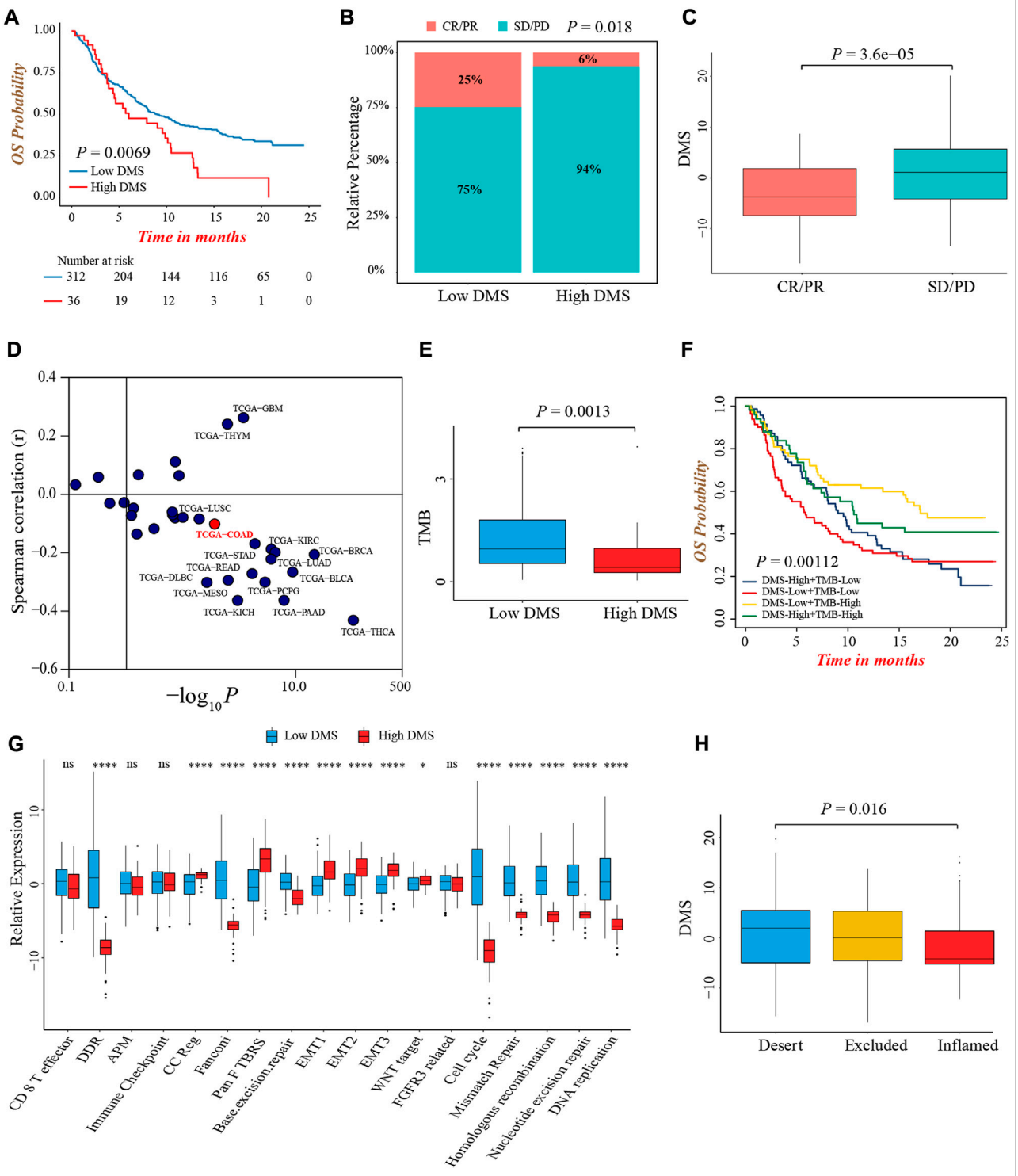


FIGURE 7
The predictive value of DMS in immunotherapy. **(A)** Overall survival analysis for low and high DMS groups in the IMvigor210 cohort (Log-rank test, $p = 0.0069$). **(B)** The proportion of patients with response to immunotherapy between DMS-high and -low groups in the IMvigor210 cohort (Chi-square test, $p = 0.018$). SD, stable disease; PD, progressive disease; CR, complete response; PR, partial response. **(C)** Comparison of DMS between immunotherapy response and non-response groups in the IMvigor210 cohort (Wilcoxon test; $p < 0.0001$). **(D)** Correlations between DMS and PD-L1 expression in pan-cancer cohorts through Spearman analysis (spearman correlation analysis). **(E)** Comparison of TMB between DMS-high and -low groups in the IMvigor210 cohort (Wilcoxon test; $p = 0.0013$). **(F)** Survival analyses for subgroup patients stratified by both DMS and TMB in the IMvigor210 cohort (Log-rank test; $p = 0.00112$). **(G)** The differences in biological processes signatures between DMS-high and -low groups in the IMvigor210 cohort (Wilcoxon test; ns: no significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **(H)** Differences in DMS among distinct tumor immune phenotypes in the IMvigor210 cohort. (Kruskal-Wallis test; $p = 0.016$).

expansion, including type I and type II interferons, TNF-α, IL-2, and IL-12. By contrast, pattern-C was immune-desert phenotype and exhibited an immune suppression status, characterized by inhibition of multiple immune and inflammatory-related pathways and a low proportion of TME infiltrating immune cells. This phenotype might be the result of immunological ignorance, the induction of immune tolerance, and a lack of appropriate environment for T cell activation, and hence, reflected the absence of pre-existing anti-tumor immune response (Gajewski et al., 2013; Herbst et al., 2014; Hegde et al., 2016; Kim and Chen. 2016). Notably, we found several innate immune cells, such as eosinophil, mast cell, natural killer cell, and plasmacytoid dendritic cell, were enriched in the TME of pattern-A, while the patients in this pattern did not present a matching preferable prognosis as pattern B. Conceptually, the immune-excluded phenotype was characterized by stromal activation and the presence of vast immune cells, while they were retained in the stroma surrounding the tumor cells nests rather than penetrated the tumor parenchyma (Salmon et al., 2012; Herbst et al., 2014; Joyce and Fearon, 2015; Hegde et al., 2016). Previous study also provided evidence that T-cells proliferation and activation were observed after anti-PD1/PD-L1 agents but no infiltration (Herbst et al., 2014). Therefore, pattern-A was classified as immune-excluded phenotype, presenting a pre-existing anti-tumor immune response while being rendered ineffective by the retention of immune cells in the surrounding stroma. Generally, the immune-desert and -excluded phenotypes were both considered as the cold tumor, containing numerous immune-suppressive cytokines that contributed to impairing the anti-tumor response. These could explain the alike poorer prognosis of pattern-A and -C.

We next screened DEGs among three patterns, and GO functional annotation revealed they were significantly associated with DNA modification and damage repair related pathways, suggesting the different clinical and biological characteristics among three patterns might be the results of differentially expressed of these genes. We further identified prognostic DEGs, termed as DNA methylation signature genes, to perform unsupervised clustering. Likewise, we found three genomic subtypes, named DNA methylation gene clusters, whose clinical outcomes and immune cell infiltrating traits were similar to three modification patterns. Our comprehensive analyses strongly revealed three immune phenotypes in colon cancer with distinct clinical and TME immune characteristics, which enhanced our understanding of the non-negligible impact of DNA methylation in shaping different TME landscapes.

Whereas, above analyses were performed based on the patient population, which could not accurately predict the specific modification pattern in individual patients. As a consequence, based on the above signature genes, we developed the DNA methylation score, DMS, to quantify the certain modification pattern for each colon cancer patient. We found DMS could precisely discriminate three immune phenotypes, with the lowest median DMS in immune-inflamed type and the highest median DMS in immune-desert type. Additionally, we revealed a markedly positive relation between DMS and tumor stage, with DMS increasing gradually from the stage I to IV. Moreover, DMS was an independent prognostic risk factor, and patients with high DMS presented an inferior survival, which was validated in multiple colon cancer cohorts. Integrally, above results indicated DMS was a reliable tool to reflect the individual DNA methylation modification pattern, and predict clinical outcomes of colon cancer.

Colon cancer was a highly heterogeneous disease, resulting from a series of distinct genetic and epigenetic changes, and a subset of molecular alterations was considered to drive the cellular and clinical behavior of cancer, including MSI, CIMP, CIN, BRAF, and KRAS mutations (Marisa et al., 2013; Phipps et al., 2015). The CIMP subtype was referred to a distinct epigenome with a high frequency of methylated genes, and approximately 20% of colon cancer were CIMP+ tumors (Toyota et al., 1999). Previous study demonstrated that CIMP+ colon tumors had a unique association with $BRAF^{V600E}$ oncogene mutation, and CIMP-associated methylation of MLH1 induced mismatch repair deficiency and resulted in a genomic instability status, also known as MSI, to generate more mutation burden and neoantigen (Weisenberger et al., 2006). In addition, aberrant epigenetic alterations as well contributed to the dysregulation of antigen-presenting machinery in tumor cells, leading to acquiring the adaptive immune response (Alexandrov et al., 2013). Here, we revealed DMS-low subtype was characterized as CIMP positive, mutant BRAF, MSI, and higher TMB, indicating this epigenotype of colon cancer identified by DMS had specific molecular alterations. Further analyses manifested the activation of immune and DNA damage repair related pathways, as well as abundant immune cells infiltration in the DMS-low group. Integrally, our findings substantiated DNA methylation modification was involved in the genomic instability, resulting in the accumulation of tumoral mutation burden and generation of neoantigen, which on the one hand activated the DNA damage repair pathway, and on the other hand enhanced immunogenicity to further activate the immune response.

Emerging evidence demonstrated that MSI and elevated TMB could heighten the anti-tumor activity of immunotherapy (Rizvi et al., 2015; Le et al., 2017; Mouw et al., 2017), and hence, we further investigated the ability of DMS to predict efficacy and prognosis of patients received anti-PD-1/PD-L1 antibody. In an immunotherapy cohort of advanced urothelial carcinoma patients treated with atezolizumab, we confirmed DMS-low group had a higher objective response rate and prolonged survival time than the DMS-high group. Higher PD-L1 expression and TMB value were considered to imply a favorable efficacy of immunotherapy (Gibney et al., 2016), and in this study, we discovered DMS was negatively correlated to TMB and PD-L1 expression in colon cancer, which validated the predictive value of DMS indirectly. Deeply, the pathway and immune phenotype analyses revealed activation of DNA damage

repair and immune-inflamed environment in DMS-low patients, providing clues to illustrate the regulatory mechanism of DNA methylation in shaping TME immune landscape, and also confirmed patients with low DMS could benefit from immunotherapy. Collectively, we considered DMS had the potential to be an excellent predictive biomarker for immune checkpoint inhibitor, and might promote personalized colon cancer immunotherapy in the future.

A major limitation of this work was the public survival and transcriptomic data of colon cancer immunotherapy was not accessible yet. Therefore, the predictive performance of DMS needed to be further certified in the colon cancer immunotherapeutic cohort.

In conclusion, for the first time, we uncovered three distinct DNA methylation modification patterns in colon cancer, and illustrated their extensive regulatory mechanism in tumor immune environment formation, which was a non-negligible factor to cause individual TME heterogeneity and different clinical outcomes. Our integrated analyses of DNA methylation modification would contribute to enhancing the understanding of tumor immune characteristics, and providing novel insights to guide immunotherapy more effectively.

Data availability statement

All data analyzed in this study were available from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/), and Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/).

Ethics statement

Ethical approval was not provided for this study on human participants because all data used in this study were abtained from common database. The ethics committee waived the requirement of written informed consent for participation.

Author contributions

Conception and design: XW and SY; data collection, analyses and interpretation: SY, YG, YX, and ZW; manuscript writing: SY, YG, YX, and ZW; manuscript Supervised: XW.

Acknowledgments

We thanked Lifeng Feng (Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University) and Hongchuan Jin (Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University) for their helpful suggestions and manuscript proof.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1008644/full#supplementary-material

References

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500 (7463), 415–421. doi:10.1038/nature12477

Andre, T., Shiu, K. K., Kim, T. W., Jensen, B. V., Jensen, L. H., Punt, C., et al. (2020). Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *N. Engl. J. Med.* 383 (23), 2207–2218. doi:10.1056/NEJMoa2017699

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Hum. Mol. Genet.* 9 (16), 2395–2402. doi:10.1093/hmg/9.16.2395

Cao, J., and Yan, Q. (2020). Cancer epigenetics, tumor immunity, and immunotherapy. *Trends Cancer* 6 (7), 580–592. doi:10.1016/j.trecan.2020.02.003

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, D. S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. *Nature* 541 (7637), 321–330. doi:10.1038/nature21349

Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., et al. (2017). Methyltransferase SETD2-mediated methylation of STAT1 is critical for interferon antiviral activity. *Cell* 170 (3), 492–506. doi:10.1016/j.cell.2017.06.042

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: An R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507

Cook, P. C., Owen, H., Deaton, A. M., Borger, J. G., Brown, S. L., Clouaire, T., et al. (2015). A dominant role for the methyl-CpG-binding protein Mbd2 in controlling Th2 induction by dendritic cells. *Nat. Commun.* 6, 6920. doi:10.1038/ncomms7920

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* 17 (2), 79–92. doi:10.1038/nrc.2016.126

Gajewski, T. F., Woo, S. R., Zha, Y., Spaapen, R., Zheng, Y., Corrales, L., et al. (2013). Cancer immunotherapy strategies based on overcoming barriers within the tumor microenvironment. *Curr. Opin. Immunol.* 25 (2), 268–276. doi:10.1016/j.coi.2013.02.009

Gao, Y., Wang, H., Li, H., Ye, X., Xia, Y., Yuan, S., et al. (2021). Integrated analyses of m(1)A regulator-mediated modification patterns in tumor microenvironment-infiltrating immune cells in colon cancer. *Oncoimmunology* 10 (1), 1936758. doi:10.1080/2162402X.2021.1936758

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinforma.* 11, 367. doi:10.1186/1471-2105-11-367

Gibney, G. T., Weiner, L. M., and Atkins, M. B. (2016). Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet. Oncol.* 17 (12), e542–e551. doi:10.1016/s1470-2045(16)30406-5

Goltz, D., Gevensleben, H., Grunen, S., Dietrich, J., Kristiansen, G., Landsberg, J., et al. (2017). PD-L1 (CD274) promoter methylation predicts survival in patients with acute myeloid leukemia. *Leukemia* 31 (3), 738–743. doi:10.1038/leu.2016.328

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Harlin, H., Meng, Y., Peterson, A. C., Zha, Y., Tretiakova, M., Slingluff, C., et al. (2009). Chemokine expression in melanoma metastases associated with CD8+ T-cell recruitment. *Cancer Res.* 69 (7), 3077–3085. doi:10.1158/0008-5472.CAN-08-2281

Hegde, P. S., and Chen, D. S. (2020). Top 10 challenges in cancer immunotherapy. *Immunity* 52 (1), 17–35. doi:10.1016/j.immuni.2019.12.011

Hegde, P. S., Karanikas, V., and Evers, S. (2016). The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition. *Clin. Cancer Res.* 22 (8), 1865–1874. doi:10.1158/1078-0432.CCR-15-1507

Heiland, D. H., Haaker, G., Delev, D., Mercas, B., Masalha, W., Heynckes, S., et al. (2017). Comprehensive analysis of PD-L1 expression in glioblastoma multiforme. *Oncotarget* 8 (26), 42214–42225. doi:10.18632/oncotarget.15031

Herbst, R. S., Soria, J. C., Kowanetz, M., Fine, G. D., Hamid, O., Gordon, M. S., et al. (2014). Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 515 (7528), 563–567. doi:10.1038/nature14011

Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P., et al. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 95 (12), 6870–6875. doi:10.1073/pnas.95.12.6870

Joyce, J. A., and Fearon, D. T. (2015). T cell exclusion, immune privilege, and the tumor microenvironment. *Science* 348 (6230), 74–80. doi:10.1126/science.aaa6204

Kim, J. M., and Chen, D. S. (2016). Immune escape to PD-L1/PD-1 blockade: Seven steps to success (or failure). *Ann. Oncol.* 27 (8), 1492–1504. doi:10.1093/annonc/mdw217

Lao, V. V., and Grady, W. M. (2011). Epigenetics and colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* 8 (12), 686–700. doi:10.1038/nrgastro.2011.173

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., et al. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357 (6349), 409–413. doi:10.1126/science.aan6733

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* 372 (26), 2509–2520. doi:10.1056/NEJMoa1500596

Loo Yau, H., Ettayebi, I., and De Carvalho, D. D. (2019). The cancer epigenome: Exploiting its vulnerabilities for immunotherapy. *Trends Cell Biol.* 29 (1), 31–43. doi:10.1016/j.tcb.2018.07.006

Luo, N., Nixon, M. J., Gonzalez-Ericsson, P. I., Sanchez, V., Opalenik, S. R., Li, H., et al. (2018). DNA methyltransferase inhibition upregulates MHC-I to potentiate cytotoxic T lymphocyte responses in breast cancer. *Nat. Commun.* 9 (1), 248. doi:10.1038/s41467-017-02630-w

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

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

Mouw, K. W., Goldberg, M. S., Konstantinopoulos, P. A., and D'Andrea, A. D. (2017). DNA damage and repair biomarkers of immunotherapy response. *Cancer Discov.* 7 (7), 675–693. doi:10.1158/2159-8290.CD-17-0226

Nguyen, L. H., Goel, A., and Chung, D. C. (2020). Pathways of colorectal carcinogenesis. *Gastroenterology* 158 (2), 291–302. doi:10.1053/j.gastro.2019.08.059

Phipps, A. I., Limburg, P. J., Baron, J. A., Burnett-Hartman, A. N., Weisenberger, D. J., Laird, P. W., et al. (2015). Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology* 148 (1), 77–87. doi:10.1053/j.gastro.2014.09.038

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

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., et al. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348 (6230), 124–128. doi:10.1126/science.aaa1348

Salmon, H., Franciszkiewicz, K., Damotte, D., Dieu-Nosjean, M. C., Validire, P., Trautmann, A., et al. (2012). Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J. Clin. Invest.* 122 (3), 899–910. doi:10.1172/JCI45817

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Suter, C. M., Martin, D. I., and Ward, R. L. (2004). Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. *Int. J. Colorectal Dis.* 19 (2), 95–101. doi:10.1007/s00384-003-0539-3

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47 (D1), D607-D613–D613. doi:10.1093/nar/gky1131

Thissen, D., Steinberg, L., and Kuang, D. (2002). Quick and easy implementation of the benjamini-hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Statistics* 27 (1), 77–83. doi:10.3102/10769986027001077

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., and Issa, J. P. (1999). CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* 96 (15), 8681–8686. doi:10.1073/pnas.96.15.8681

Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., et al. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* 38 (7), 787–793. doi:10.1038/ng1834

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

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., et al. (2019). Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol. Res.* 7 (5), 737–750. doi:10.1158/2326-6066.CIR-18-0436

Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020). m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol. Cancer* 19 (1), 53. doi:10.1186/s12943-020-01170-0

Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., et al. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* 525 (7569), 389–393. doi:10.1038/nature15252

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fgene.2022.1036402

OPEN ACCESS

EDITED BY
Shun Lu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Chen Xiaodong,
The First People's Hospital of Foshan, China
Wancong Zhang,
Second Affiliated Hospital of Shantou University Medical College, China

*CORRESPONDENCE
Hong-Wei Liu,
liuhongwei0521@hotmail.com
Xuan Liao,
liaoxuan19860801@163.com

†These authors have contributed equally to this work

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

RECEIVED 04 September 2022
ACCEPTED 07 October 2022
PUBLISHED 24 October 2022

CITATION
Yang L-K, Lin C-X, Li S-H, Liang J-J, Xiao L-L, Xie G-H, Liu H-W and Liao X (2022), Novel IKZF3 transcriptomic signature correlates with positive outcomes of skin cutaneous melanoma: A pan-cancer analysis.
Front. Genet. 13:1036402.
doi: 10.3389/fgene.2022.1036402

Novel IKZF3 transcriptomic signature correlates with positive outcomes of skin cutaneous melanoma: A pan-cancer analysis

Lin-Kai Yang†, Can-Xiang Lin†, Sheng-Hong Li, Jia-Ji Liang, Li-Ling Xiao, Guang-Hui Xie, Hong-Wei Liu* and Xuan Liao*

Department of Plastic Surgery, The First Affiliated Hospital of Jinan University, Innovative Technology Research Institute of Tissue Repair and Registration, Key Laboratory of Regenerative Medicine, Ministry of Education, Guangzhou, Guangdong Province, China

To investigate the potential relationship between Ikaros family genes and skin cutaneous melanoma (SKCM), we undertook a pan-cancer analysis of the transcriptional signature and clinical data of melanoma through multiple databases. First, 10,327 transcriptomic samples from different cancers were included to determine the overall characteristics and clinical prognoses associated with Ikaros gene expression across cancer types. Second, differentially expressed genes analysis, prognostic evaluation, and gene set enrichment analysis were employed to investigate the role of Ikaros (*IKZF*) genes in SKCM. Third, we evaluated the relationship between Ikaros family genes and SKCM immune infiltrates and verified the findings using the GEO single-cell sequencing dataset. The results show that Ikaros genes were widely expressed among different cancer types with independently similar patterns as follows: 1. *IKZF1* and *IKZF3*, and 2. *IKZF2* and *IKZF4–5*. *IKZF2* and *IKZF5* were downregulated in the primary tumor, and *IKZF1–3* expression decreased significantly as the T-stage or metastasis increased in SKCM. Moreover, high *IKZF1–3* expression was associated with better overall survival, disease-specific survival, and progression-free interval. *IKZF3* is an independent prognostic factor of SKCM. Among Ikaros genes, the expression of *IKZF1* and *IKZF3* positively correlated with the infiltration level of $CD4^+$ T cells and $CD8^+$ T cells, B cells, and Tregs in SKCM and negatively correlated with the infiltration level of M0 and M1 macrophages. Moreover, single-cell sequencing data analysis revealed that *IKZF1* and *IKZF3* were mainly expressed by immune cells. Correlation analysis shows the immune factors and drug responses associated with IKZF3 expression. In conclusion, the present study is the first, to our knowledge, to identify a pan-cancer genomic signature of the Ikaros gene family among different cancers. Expression of these family members, particularly high levels of *IKZF3*, indicate positive immunological status and beneficial clinical outcomes of SKCM. *IKZF3* may therefore serve as potential targets for immunotherapy of melanoma.

KEYWORDS

ikaros family genes, SKCM, immune infiltration, immunotherapy biomarker, pan-cancer analysis

Introduction

As one of the most serious cancers, SKCM easily metastasizes to other parts of body through the lymphatic and peripheral circulatory systems. In recent years, SKCM morbidity tends to affect more of the younger population and has become one of the tumors with the fastest increase in incidence across cancer types (Ferlay et al., 2015; Siegel et al., 2016). Early-stage patients with SKCM do not exhibit apparent symptoms or classical characteristics in pathological examinations, which hinders physicians from diagnosing and treating patients in a timely manner (Megahed et al., 2002).

For example, conventional surgical resection of local lesions and radiation therapy hardly achieve significantly positive outcomes of SKCM, particularly for multiorgan metastasis (Balch et al., 2001). In contrast, the antiapoptotic activity of SKCM renders it resistant to most chemotherapies and increases the risk of its recurrence (Soengas and Lowe, 2003; Nazarian et al., 2010). Although SKCM is currently treated with a combination of surgery, radiation, and other systematic modalities, the survival rate of patients with advanced melanoma metastasis remains extremely low.

However, with constant improvements in immunomodulatory technology, cancer immunotherapy appears to serve as an effective method to treat malignant tumors. Mounting studies highlight the role of tumor-infiltrating lymphocytes in eliminating tumors from the tumor microenvironment (TME), thus making it reasonable to treat SKCM with an immunotherapy strategy.

In tumor tissues, fibroblasts, infiltrating lymphocytes, and components such as the surrounding stromal cells and capillaries constitute the TME (Locy et al., 2018). As the "soil" for tumor cell growth, the TME mainly provides nutrients for tumor cell proliferation and invasion. However, once the dysfunctional infiltrated immune cells recover, the tumor microenvironment may become a battlefield to kill tumors. In recent years, mounting studies reveal that quiescent tumor-infiltrating lymphocytes (TILs) are the main forces that kill tumor cells after they respond to effective immunological stimulation *via* the TME (Lee et al., 2016; Fu et al., 2019). In SKCM, effector T cells, a member of the TIL population, continuously eradicate tumor cells once activated by antigens presented by other cells, thus improving the prognostic outcome of patients (Zhou et al., 2005).

Methods for transforming the TME from a quiescent to active immunological state range from intravenous injection of recombinant interleukin (Atkins et al., 2000) and adoptive T-cell therapy (Besser et al., 2010) to monoclonal antibody therapy (Weber, 2010). By blocking the PD-1/PD-L1-induced signaling pathway or activating the costimulatory receptor 4-1BB/4-1BBL, monoclonal antibody therapy of tumors may improve clinical outcomes. However, most patients with SKCM do not positively respond to single-target therapy, particularly those with metastasized melanoma (Tsai et al., 2014). Therefore, it is vitally importance to investigate effective immunological therapies that target SKCM.

The Ikaros family of zinc-finger proteins plays a critical role in lymphatic development, differentiation, and homeostasis. There are five homologous members of the Ikaros gene family including Ikaros (*IKZF1*), Helios (*IKZF2*), Aiolos (*IKZF3*), Eos (*IKZF4*), and Pegasus (*IKZF5*), which are translated as distinct isoforms through alternative splicing (John and Ward, 2011). The Ikaros zinc finger-like N-terminal domain specifically binds DNA, and its zinc finger-like C-terminal domain mediates homologous binding as well as to other family members (Molnár et al., 1994; Sun et al., 1996). These features confer upon Ikaros family members the ability to regulate cell proliferation, differentiation, and apoptosis through the combination of Ikaros family members or other transcription factors (Yoshida and Georgopoulos, 2014).

IKZF4 and *IKZF5* are mainly expressed in skeletal muscle and other solid organs, and *IKZF1–3* expressed by lymphocytes participate in the regulation of the development and differentiation of these cells (Perdomo et al., 2000; Yoshida and Georgopoulos, 2014). It is therefore reasonable to conclude that Ikaros genes influence the immunological status in the TME to affect patients' clinical outcomes. However, the potential relationship between TILs that express Ikaros genes in the TME and SKCM has not been comprehensively investigated.

In the present study, we first applied bioinformatics to analyze RNA-seq data for different cancers to determine the role of Ikaros genes in their pathogenesis. The potential function of Ikaros genes in SKCM was confirmed using gene set enrichment analysis (GSEA), survival evaluation, and other analytical tools. To determine the effects of Ikaros gene expression on the TIL population of SKCM, we performed immunological analyses using CIBERSORT and single-cell RNA-seq. Finally, a novel *IKZF1–3* transcriptomic signature was shown to correlate with positive outcomes of SKCM.

Methods

Data analysis

RNA sequence data and corresponding clinical data of patients with different carcinomas (33 types) were acquired

from the UCSC Xena database (https://xenabrowser.net/datapages/). The Ensemble IDs of the expression profile data were converted to symbol IDs through the human GTF file. Cancer types were included: ACC, BLCA, BRCA, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, OV, PAAD, PRAD, READ, SKCM, STAD, TGCT, THCA, THYM, UCEC, and UCS. At the same time, clinical data were included survival time, progression free survival time, survival status, clinical staging and other clinical follow-up data.

Ikaros transcriptional signature across cancer types

We determined the expression levels of *Ikaros* family members in all cancers and identified the individual expression characteristics of Ikaros family members through Pearson coefficient correlation analysis. In view of the lack of normal samples of SKCM in TCGA data and to evaluate the association between Ikaros and the progression of SKCM, differential expressional levels of normal and tumor samples were retrieved and compared using the GEPIA database (http://gepia.cancer-pku.cn). Furthermore, we selected cancer types with more than three normal data samples and analyzed the differences between the expression of Ikaros family genes in cancer and normal tissues. |logFC>2| and $p < 0.05$ were considered to indicate significant differences in expression.

Prognostic analysis

The prognostic outcomes of Ikaros expression among multiple cancers were determined using Cox univariate analysis with the indexes of hazard ratios (HRs) and 95% confidence intervals (CIs). The log-rank test and the Kaplan–Meier algorithm were used to determine the correlation between expression of individual Ikaros genes and a patient's survival. Several indicators of outcomes, such as overall survival (OS), progression-free interval (PFI), and disease-free survival (DSS) were applied to comprehensively evaluate the prognostic significance of Ikaros gene expression.

GSEA and protein-protein interaction network analysis

A computational method of GSEA (GSEA v.3.0 in the Java environment) was used to explore the potential mechanism of the regulation of Ikaros genes in the occurrence and development of SKCM. We ranked the levels of Ikaros genes expressed by SKCM samples from high to low for GSEA, which was used to identify signaling pathways correlated with Ikaros. Gene-set permutations were performed 1,000 times for each analysis. The expression levels of *IKZF1–IKZF5* were used as phenotypic labels for all SKCM samples. A false discovery rate (FDR) < 0.25 and $p < 0.05$ were used to identify KEGG pathways significantly enriched in each phenotype. Ikaros-interacting proteins were determined using the GeneMANIA protein-protein interaction network (http://genemania.org/).

Analysis of immune infiltration

An algorithm of ESTIMATE (Yoshihara et al., 2013) was adopted to evaluate immune infiltration condition across cancer types using the indexes of stromal score, immune score, and tumor purity. The relationship between Ikaros gene expression and immunological reactivity was determined by comparing the scores and purity indexes among different carcinomas. CIBERSORT was used for characterizing the infiltrating immune cell composition of SKCM tissue from their gene expression profiles. We performed deconvolution calculations with the annotating file LM22. CIBERSORT annotates the abundance of 22 infiltrating immune cells through a 547-gene expression eigenmatrix (Newman et al., 2015). Through the results of CIBERSORT, the correlations between immune-infiltrating cells in the TME of SKCM and the expression of Ikaros genes were demonstrated.

Single-cell RNA-seq analysis

The Tumor Immune-single Cell Hub (TISCH) is an SCRNA-Seq database focused on the TME, which provides detailed cellular annotations at the level of a single cell that enables exploration of the TMEs of different cancer types (Sun et al., 2021). GSE72056, a series of 4,645 single-cell sRNA-seq data from 19 melanoma samples included in the Gene Expression Omnibus by Tirosh et al. was further analyzed using TISCH to determine Ikaros gene expression levels in immunocytes.

Correlation of *IKZF3* expression and immune factors

The TISIDB database (http://cis.hku.hk/TISIDB) is a web portal for tumour and immune system interaction, which integrates multiple heterogeneous data types. Gene signatures of Tumor-infiltrating lymphocytes (TILs) were obtained from TISIDB database. Here, we further analyzed associations for IKZF3 with immunomodulators, chemokines and receptors. Immunomodulators can be divided into three categories, including immunoinhibitors, immunostimulators and major histocompatibility complex (MHC) molecules.

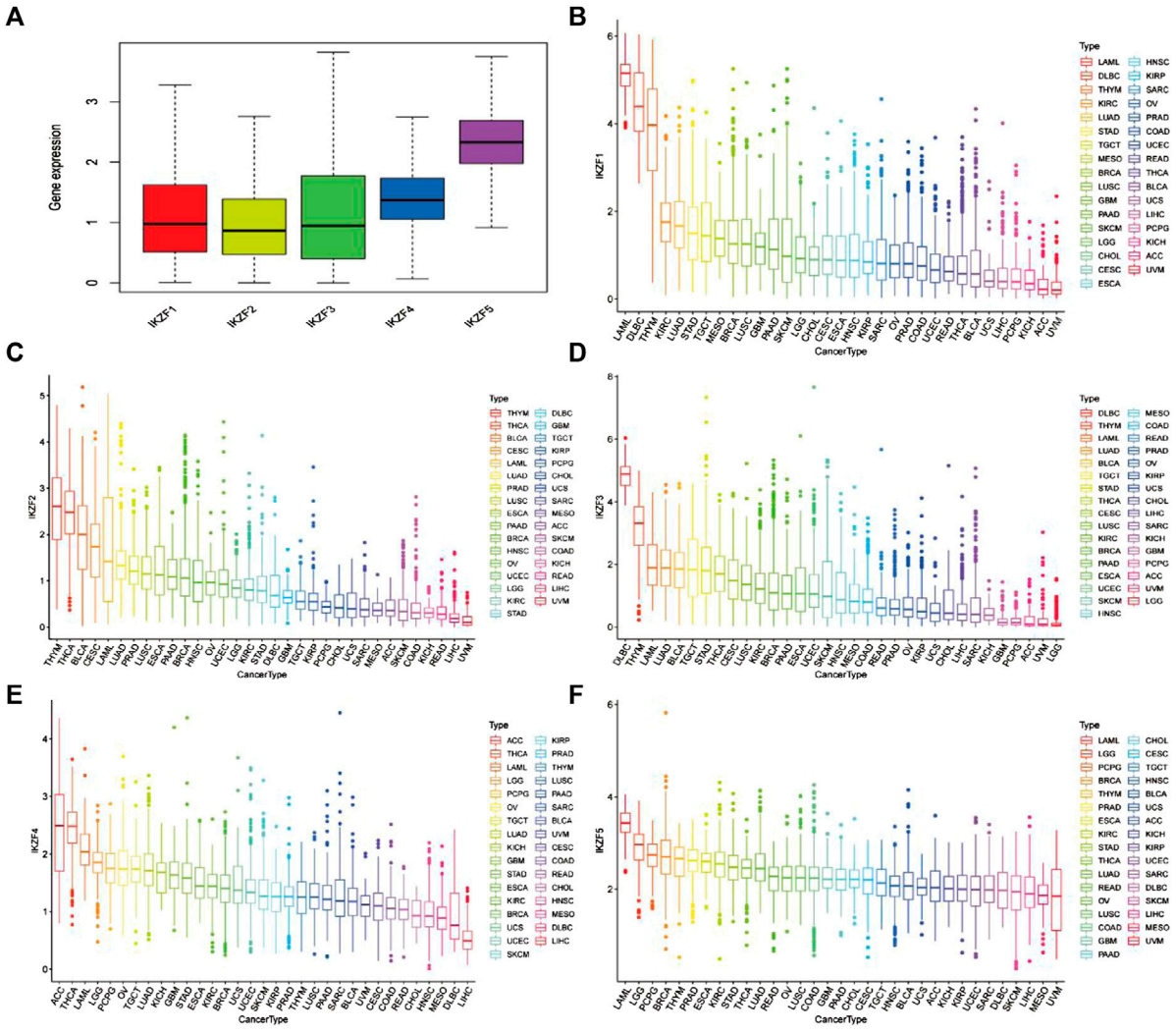


FIGURE 1
(A) Mean mRNA expression values of Ikaros in pan-cancers. **(B–F)** mRNA expression levels of *IKZF1–5* in different human cancers ranked from high to low.

Correlation of *IKZF3* expression and drug responses

The cellminer database contains (NCI)-60 gene expression and drug sensitivity data. We obtained the cellminer database including drug data approved by the US Food and Drug Administration and drug data evaluated in clinical trials. The Pearson correlation coefficient was used to evaluate the correlation between *IKZF3* expression and drug response.

Statistical analysis

The Wilcoxon rank-sum test and the Kruskal–Wallis test were used to compare two and multiple groups, respectively. Pearson analysis was used to perform correlation analysis among members of the Ikaros gene family. $p < 0.05$ indicates a significant difference. All statistical analyses were performed using R (version 3.6.0).

Results

Transcriptomic signatures of ikaros genes across cancer types

Differential transcriptomic levels of Ikaros members (*IKZF1–5*) were determined through the analysis of 11,057 samples of 33 tumors (10,327 tumor and 730 normal samples). To determine the expression

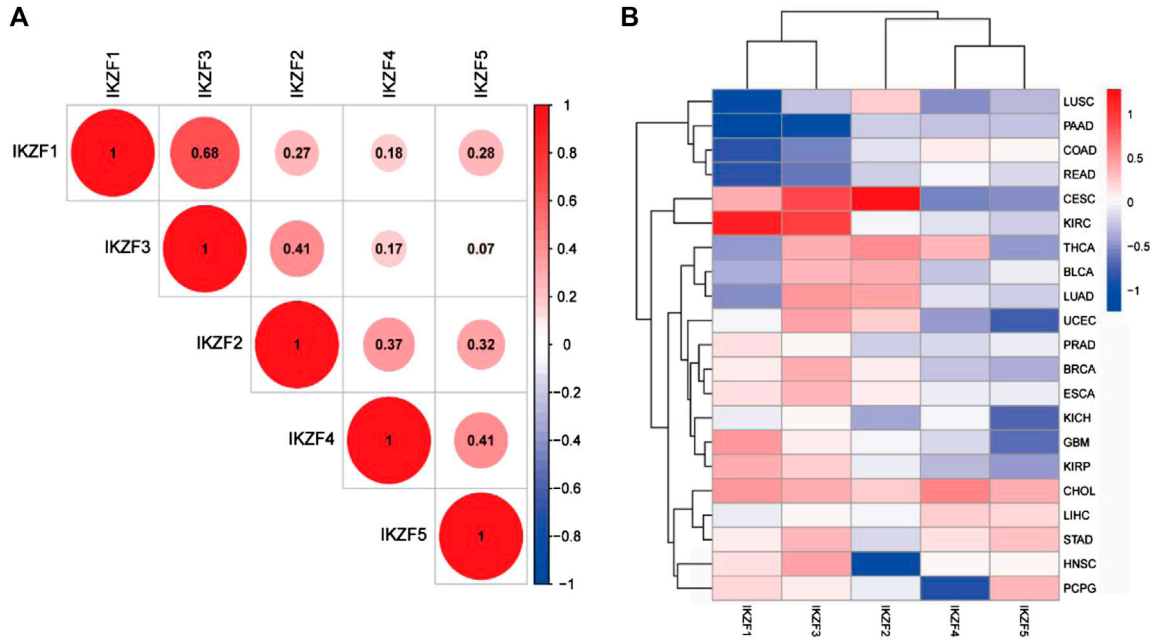


FIGURE 2
(A) Pearson coefficient indexes between Ikaros genes. **(B)** Heatmap of differentially expressed Ikaros genes in 21 TGCA datasets. Red and blue indicate upregulation and downregulation, respectively.

characteristics of Ikaros members, we first calculated the mean expression levels of *IKZF1–5* in pan-cancers. Among them, the mean expression levels of *IKZF1–3* were relatively consistent and that of *IKZF5* was the highest (Figure 1A). Furthermore, the transcriptional signatures of individual Ikaros genes in different cancer types revealed that *IKZF1–5* were mainly overexpressed in hematological tumors such as LAML and DLBC, but were expressed at low levels in digestive tumors, for example, LIHC. The detailed expression levels of Ikaros genes in different cancer types are shown in Figures 1B–F.

The Pearson coefficient index of every pair of Ikaros genes indicated that *IKZF1* and *IKZF3* were closely associated in various cancers, while *IKZF3* and *IKZF5* showed the weakest correlations (Figure 2A). We next screened 21 tumor datasets, each including at least three normal samples, using hierarchical cluster analysis of differentially expressed genes. The results confirmed the transcriptional relationship between *IKZF1* and *IKZF3*, because they ranked in the same cluster with a transcriptional pattern similar to that of *IKZF1* and *IKZF3* in different cancer types (Figure 2B).

To further investigate the role of *IKZF1–5* in predicting clinical outcomes, we conducted univariate Cox regression analysis of gene samples with the corresponding clinical data. We found that most patients with cancer that highly expressed *IKZF1–3* had a significantly improved prognosis. For example, high levels of *IKZF1–3* in patients with CESC, HNSC, LUAD, SARC, and SKCM indicated longer survival but implied poor prognosis of patients with LGG and UVM with the same transcriptional profile (Figure 3). Although there was no significant association between *IKZF4* and *IKZF5*, their high expression levels indicate better outcomes of certain types of cancers. the high expression of *IKZF1, IKZF2, IKZF3,* and *IKZF5* in SKCM will indicate a better prognosis.

IKZF1–3 expression negatively correlates with the progression of SKCM

Compared with normal tissue, *IKZF2* and *IKZF5* mRNAs were expressed at lower levels in SKCM (Figure 4A). Considering that Ikaros genes may participate in the progression of SKCM, we investigated their transcriptional levels associated with different TNM stages. The results show that the expression of *IKZF1–3* mRNAs in SKCM significantly decreased with higher T-stage and significantly decreased when SKCM involved local lymph node metastasis that penetrated the original tissue boundary (Figure 4B). However, the levels of mRNAs of other members of the Ikaros family did not vary with higher T-stage. Moreover, among different tumor locations, Ikaros genes were expressed at the highest levels in regional lymph nodes, *IKZF1* and *IKZF3* were expressed at the lowest levels in distant

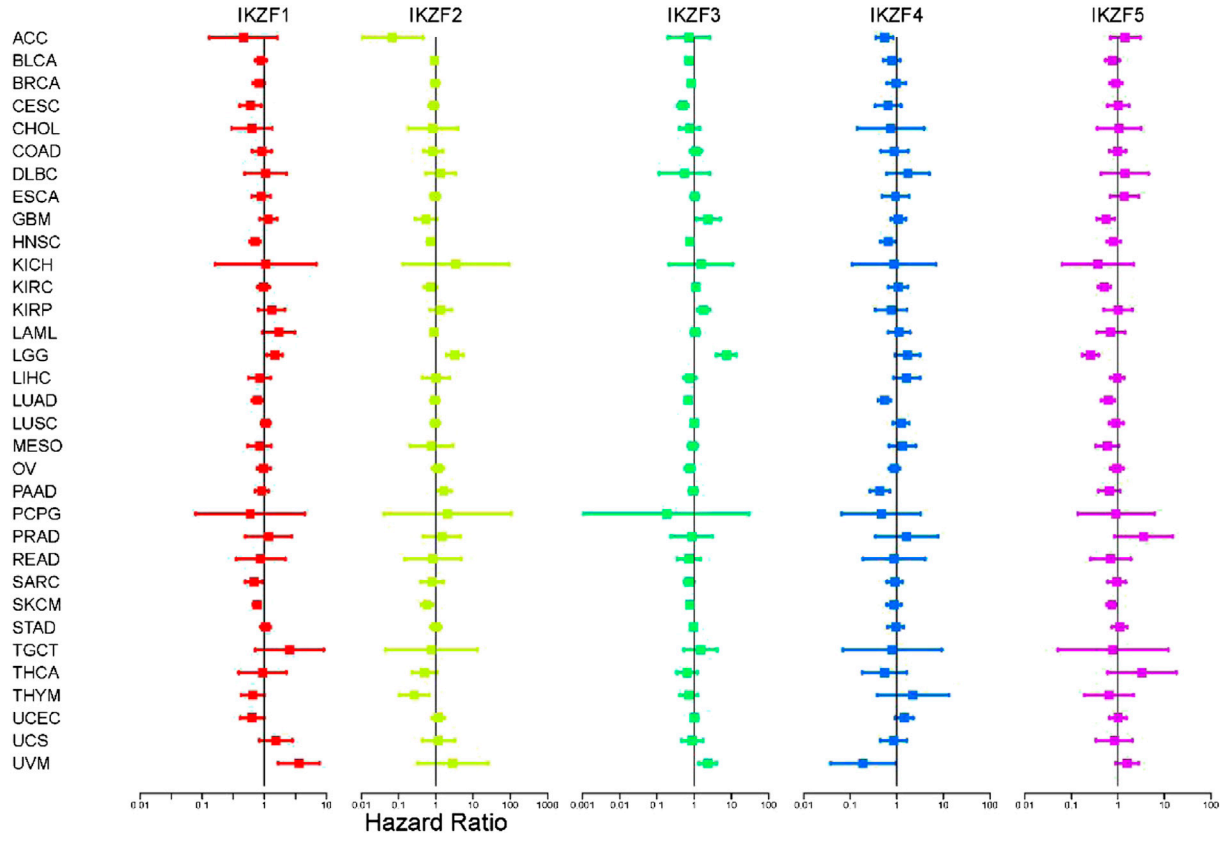


FIGURE 3
Cox univariate regression analysis of the overall survival associated with Ikaros expressed in different cancers. The vertical axis indicates the cancer type. Different colors represent different Ikaros members, and the dots represent the risk ratio (HR) of the gene in cancer. The two lines End represents the 95% confidence interval (CI).

metastases, and *IKZF2*, *IKZF4*, and *IKZF5* were expressed at the lowest levels in the primary tumor.

Highly expressed *IKZF1–3* are associated with favorable clinical outcomes of patients with SKCM

The results of Kaplan-Meier analysis suggested that patients with SKCM with highly expressed *IKZF1–3* experienced significantly longer OS, DFS, and PFI than patients with lower levels ($p < 0.05$) (Figures 5A–C). However, the associations of the OS, DSS, and PFI rates of patients with higher expression levels of *IKZF4–5* were not significantly different (Figures 5D,E), and patients who expressed high levels of *IKZF1–3* experienced significantly longer DSS and longer tumor progression-free survival. These findings indicate that high expression of *IKZF1–3* mRNAs was significantly associated with good prognosis of patients with SKCM. Further multivariate-cox analysis indicates that IKZF3 can be used as an independent prognostic factor (Supplementary Figure S1).

High expression of *IKZF1–3* activates an antitumor immune response

To further investigate Ikaros-associated signaling pathways, transcriptional data of SKCM samples were ranked by the relative levels of *IKZF1–5* expression in the upper 10th percentile (IKZFhi) vs. the lower 10th percentile (IKZFlo) and subjected to GSEA. Figures 6A–C shows that high expression of *IKZF1–3* was associated with several activated KEGG pathways, while *IKZF4* and *IKZF5* were not clustered in known pathways (data not shown). Moreover, the GSEA results demonstrate that high expression of *IKZF1–3* plays a critical role in the regulation of the activation of the immune response through the B-cell receptor signaling pathway, the T-cell receptor signaling pathway, leukocyte transendothelial migration, and natural killer cell-mediated cytotoxicity. Furthermore, IKZF1-3hi was closely

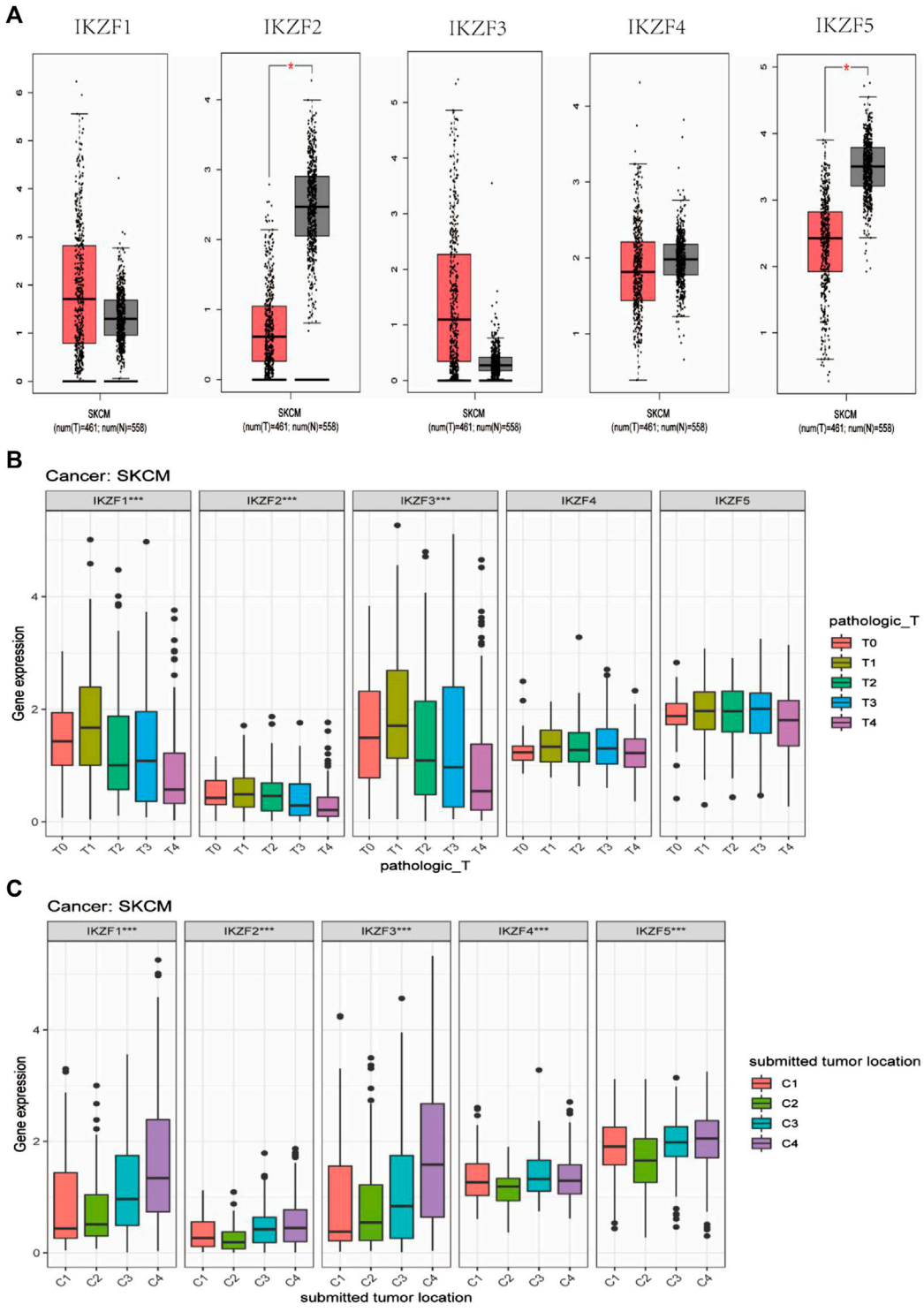


FIGURE 4
(A) GEPIA box diagram shows the expression levels of Ikaros genes in SKCM and healthy skin. T: tumor tissue, N: normal tissue; **(B)** Ikaros gene expression in patients with different pathological T-stages; **(C)** Ikaros gene expression at different tumor locations, C1 (Distant Metastasis), C2 (Primary Tumor), C3 (Regional Cutaneous or Subcutaneous Tissue [includes satellite and in-transit metastasis]), C4 (Regional Lymph Node) (*p < 0.05, **p < 0.01, ***p < 0.001).

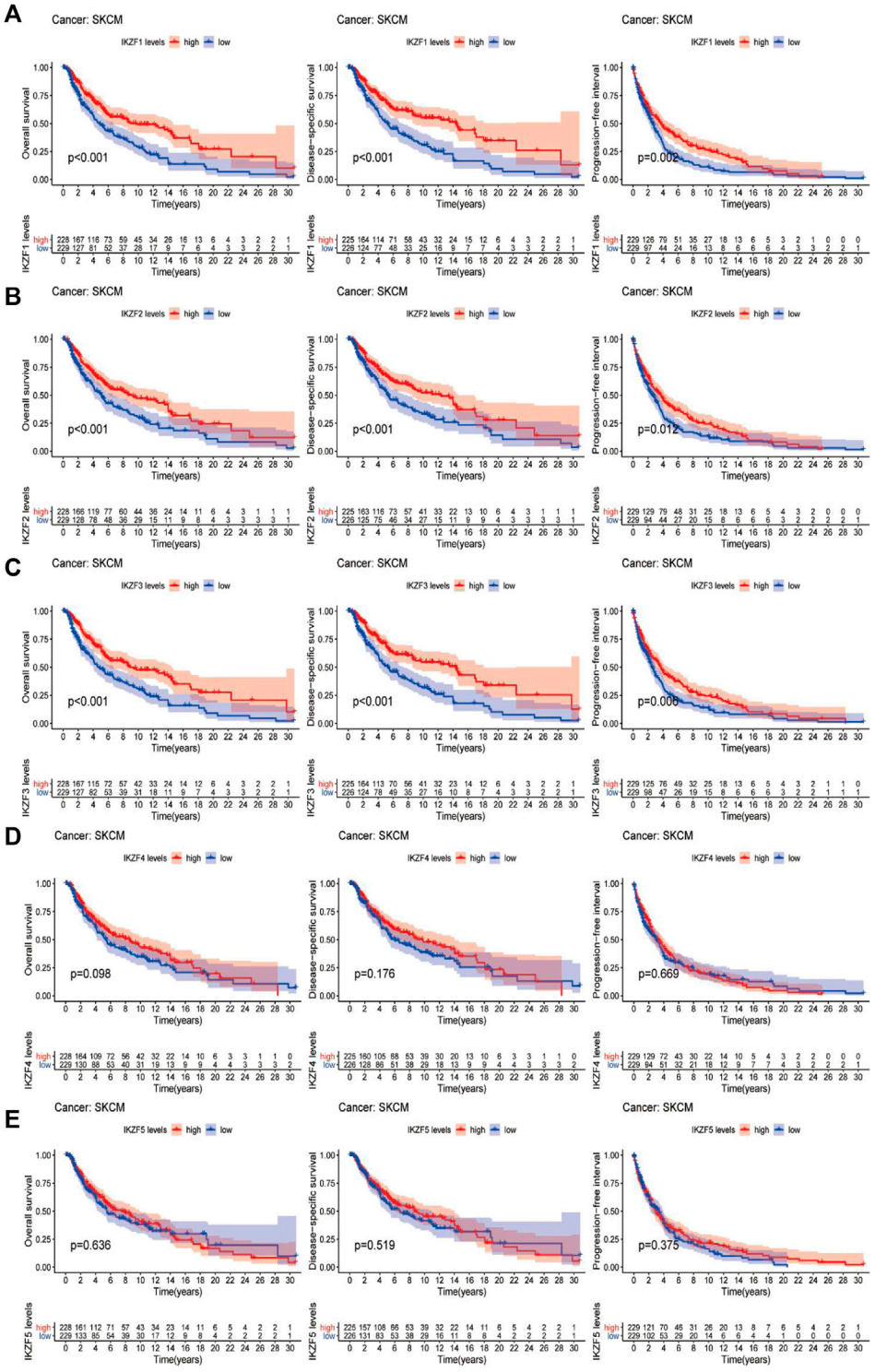


FIGURE 5
Comparison of Kaplan–Meier survival curves of patients with high or low expression of Ikaros. OS, DSS, and PFI rates associated with the expression of Ikaros genes in the SKCM cohort. **(A–E)**, *IKZF1–5*, respectively.

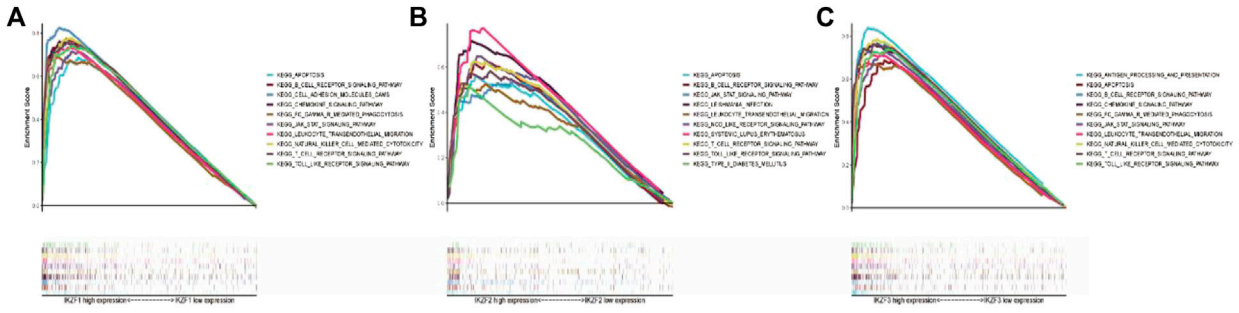


FIGURE 6
GSEA results of single Ikaros genes. **(A–C)**: The first 10 KEGG pathways clustered depending on the relationships of the transcriptional changes of *IKZF1–3*.

associated with apoptosis and chemokine signaling pathways in the SKCM cohort, which indicates the anticancer effect of *IKZF1–3* through its regulation of the immune response.

Increased *IKZF1–3* expression enhances immunological infiltration of SKCM

Tumor infiltration analysis using CIBESORT and Estimate, shows that the stromal and immune scores significantly increased as the expression of *IKZF1–3* was enhanced in SKCM samples. Moreover, there was a positive correlation between increased *IKZF1–3* expression and the abundance of immunocytes such as B cells, CD4-positive T cells and CD8-positive T cells, although there was a negative correlation between increased *IKZF1–3* expression and tumor infiltration with M0/M2-polarized macrophages. Furthermore, increased infiltration of M0/M2 macrophages was a poor prognostic factor for patients. (Supplementary Figure S2). However, there was no significant relationship between tumor infiltration and expression of *IKZF4–5*.

Single-cell sequence analysis of ikaros expression

Single-cell sequence analysis was applied to specifically determine the transcriptomic signature of Ikaros members. For this purpose, we employed melanoma tissue of the GSE72056 dataset. Using the Louvain clustering algorithm and a KNN graph, melanoma tissue was divided into the major cell types as follows: B cells, CD4-positive T cells, CD8-positive exhausted T cells, endothelial cells, fibroblasts, SKCM cells, monocytes, follicular helper T cells, Th1 helper cells, and proliferating T cells (8A, G). Consistent with the results mentioned above, the levels of *IKZF1* and *IKZF3* were highest in immunocytes and lowest in SKCM cells, which further indicates the function of *IKZF1* and *IKZF3* in the immune response to tumor cells and indicates the status of immunocyte infiltration in the TME (Figures 7B,D). However, *IKZF2*, *IKZF4*, and *IKZF5* were basically not detectably expressed by immunocytes, malignant cells, and stromal cells (Figures 7C,E,F).

Correlation of *IKZF3* expression and immune factors

Relationship of IKZF3 expression with immune factors was evaluated using TISIDB and TCGA databases. we calculated the correlation of IKZF3 expression and Gene signatures of TILs, including three immunomodulators, chemokines and receptors (Figures 8A–E). Figure 8A shows correlations between IKZF3 and immunostimulators, IKZF3 strongly correlated with CD27, TNFRSF9 and ICOS, and weakly correlated with TNFSF9 and ULBP1. Figure 8B shows correlations between IKZF3 and Immunoinhibitors, IKZF3 strongly correlated with CD96, PDCD1 and TIGIT, and weakly correlated with KDR and VTCN1. Figure 8C shows correlations between IKZF3 and chemokines, IKZF3 strongly correlated with CXCL13, XCL2 and CCL4, and weakly correlated with CXCL1 and CCL27. Figure 8D shows correlations between IKZF3 and receptors, IKZF3 strongly correlated with CXCR3, CCR5 and CCR4, and weakly correlated with CXCR1 and CXCR2. Figure 8E shows IKZF3 expression moderately to strongly correlated with MHC.

Correlation of *IKZF3* expression and drug responses

In order to understand the potential relationship between IKZF3 expression and various types of drug response, we performed a correlation analysis to use CellMiner to identify

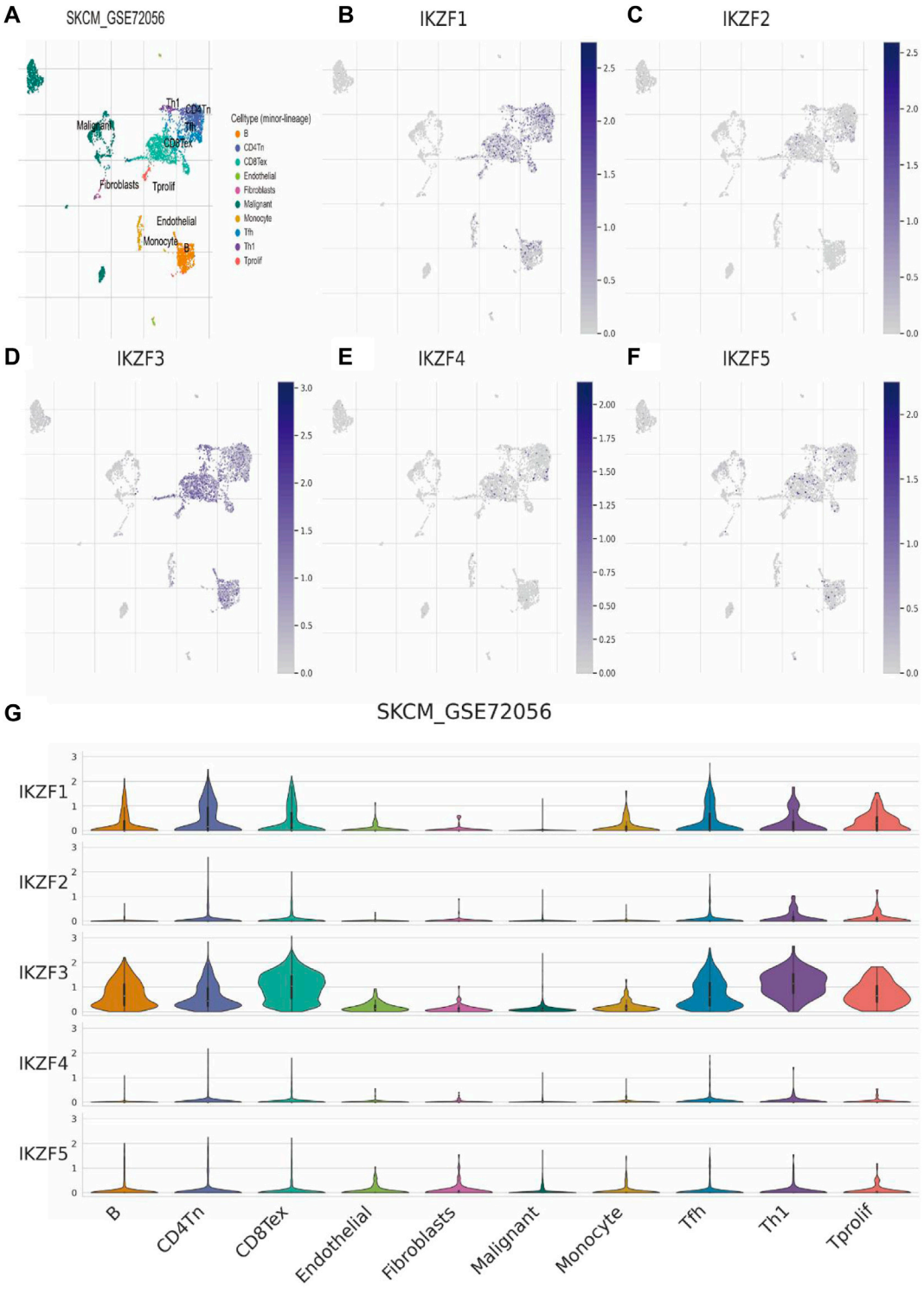


FIGURE 7
Transcriptomic signatures of different cell types in the melanoma microenvironment. **(A)**: There were 10 major cell types in the SKCM microenvironment as follows: B cells, CD4-positive T cells, CD8-positive T cells, endothelial cells, fibroblasts, malignant melanoma cells, monocytes, follicular helper T cells, Th1 helper cells, and proliferating T cells. **(B–E)**: Expression levels of *IKZF1–5* among different cell types in the melanoma.

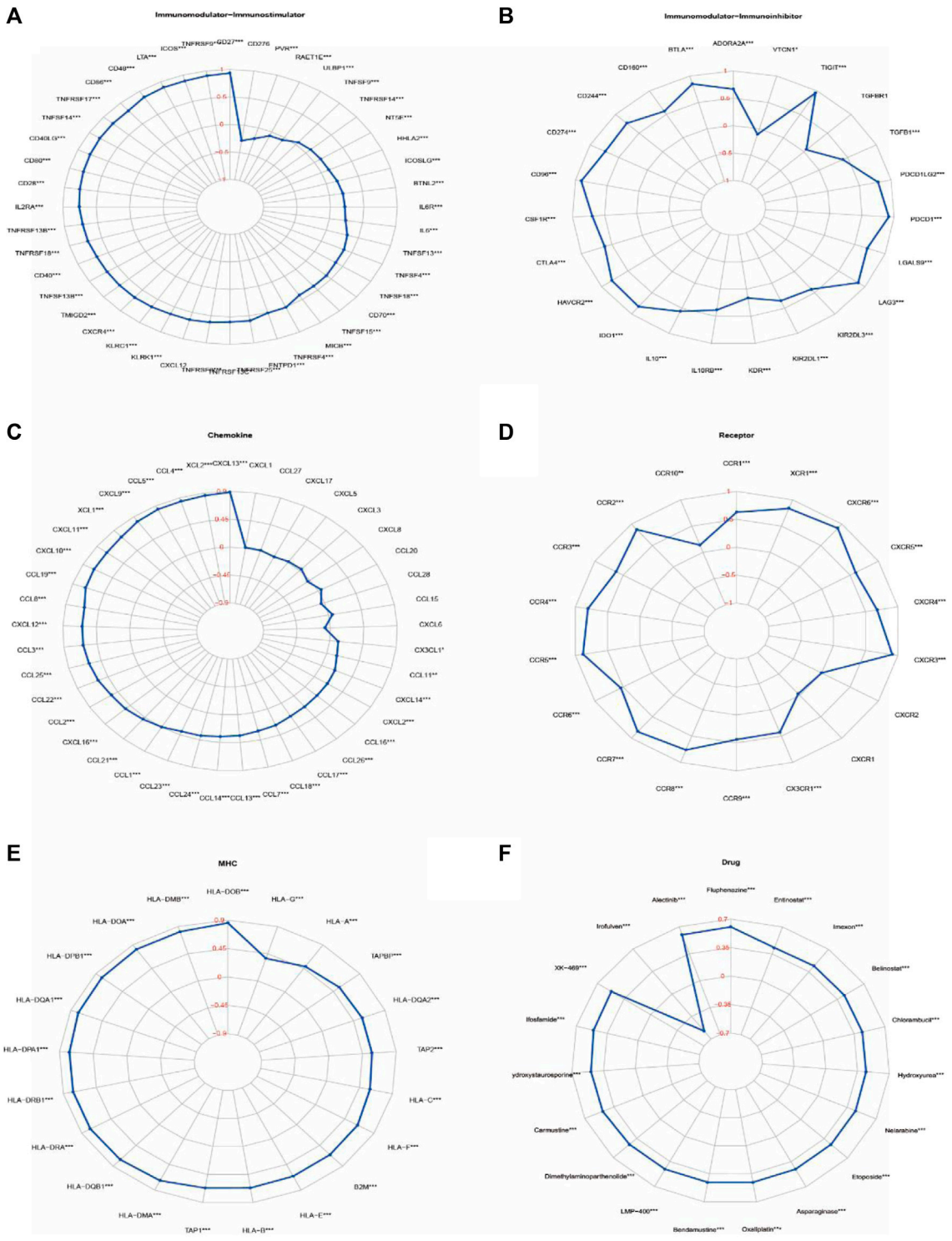


FIGURE 8
Correlations of IKZF3 expression with immunomodulatory genes, drugs. **(A–E)** Correlation between immunomodulators and chemokines (or receptors) with *IKZF3* expression. **(F)** Correlation between *IKZF3* expression and drug responses. The red numbers of radar chart represent Spearman's correlation coefficient (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

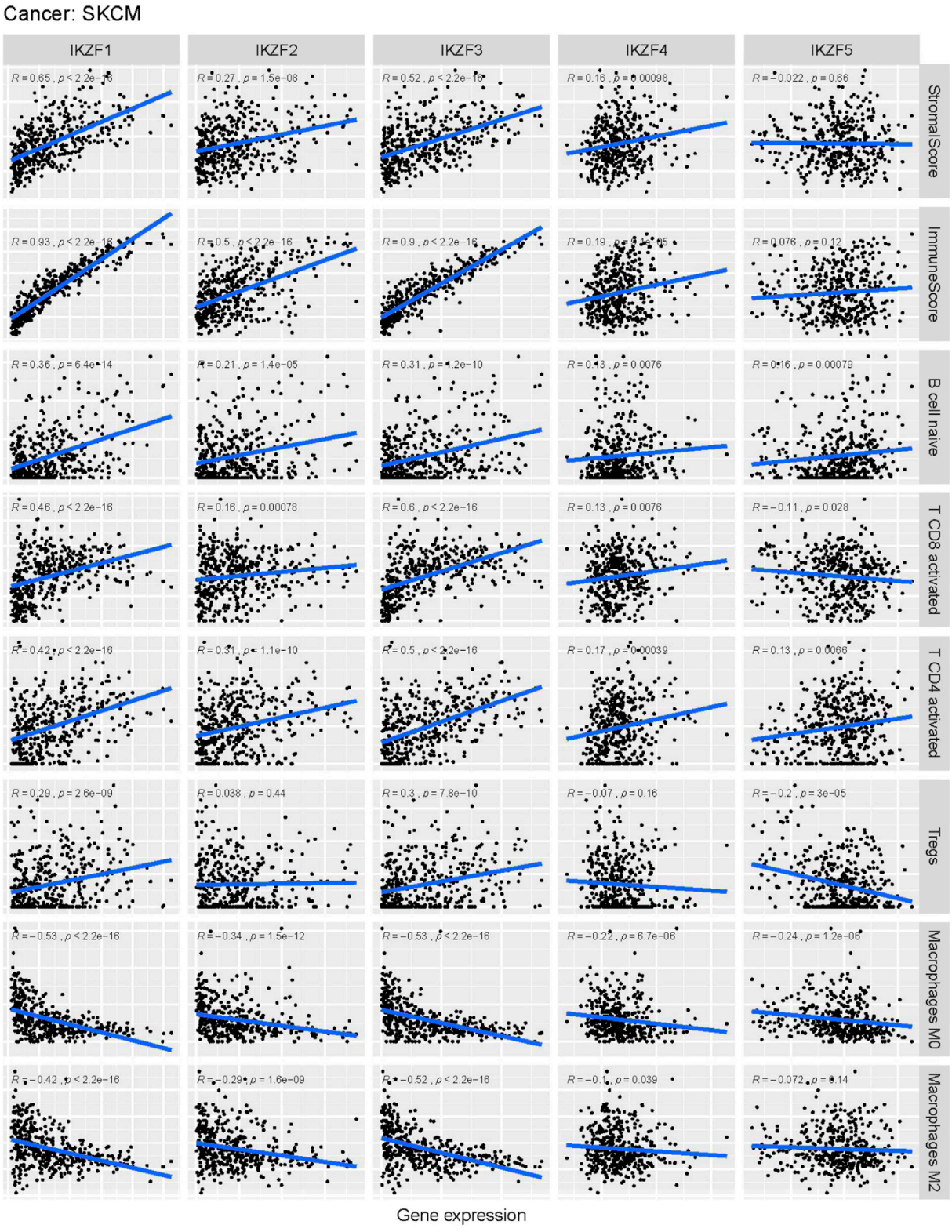


FIGURE 9
Correlation analysis of Ikaros family genes and the abundance of immune-infiltrating cells in the tumor microenvironment. Horizontal labels represent different levels of Ikaros gene expression, vertical labels represent different infiltrating immune cells, and the line represents the correlation between them. The correlation value is represented by the index R.

potential drug candidates. IKZF3 expression in tumor cells was positively associated with increased drug sensitivity to Fluphenazine、Alectinib, and negatively associated with Irofulven. We have screened a total of 20 drugs that are most related to the expression of IKZF3, which are also shown in the radar chart (Figure 8F).

Discussion

With an extremely low survival rate and high potential for metastatic growth, SKCM causes heavy economic losses and medical burdens annually worldwide. Except for conventional surgical removal of the carcinoma and regular radiation therapy, immunotherapy is one of the most promising treatments for SKCM (Besser et al., 2010; Hodi et al., 2010).

The recent research have determined the effect of immunotherapy in the treatment of tumors. As the development of cancer is highly associated with immune micro-environment, determination of the pivotal role of immune response in the tumor is of importance in developing new immunotherapeutic strategies (Roma-Rodrigues et al., 2019). IMiD (Immunomodulatory drugs) covers thalidomide, Len and pomalidomide, and is clinically approved for the treatment of MM (multiple myeloma) and other malignant tumors. Among them, MM and MDS (myelodysplastic syndrome) are widely used and studied. IMiD can achieve the therapeutic effect by inducing selective ubiquitination and proteasome degradation of Ikaros (Gao et al., 2020). In addition, IMiD has the ability to directly inhibit the growth of tumor cells and strong immune stimulation characteristics, so it has multiple effects on the existence of different cell components in the tumor microenvironment. The efficacy and safety of IMiD have been verified in a wide range of clinical trials. It is increasingly clear that IMiDs are promising in the treatment of MM (Gao et al., 2020), MDS and CLL (chronic lymphocytic leukemia) (Vitale et al., 2016). At the same time, Ikaros can be used as prognostic factors for MM patients treated by Len (Tachita et al., 2020). Previous Research has shown that patients with lower IKZF3 expression level have a poorer therapeutic effect with Len, which will also lead to shorter progression-free survival and overall survival (Pourabdollah et al., 2016). Moreover, inhibiting RUNX to protect Ikaros from degradation can significantly improve the drug resistance of IMiDs in MM(Zhou et al., 2019). In addition, the combined use of imatinib in children with B-ALL can enhance the therapeutic effect of IKZF1-deletion patients. Imatinib, as an intensive treatment for B-ALL in IKZF1- deletion children, significantly reduces the risk of recurrence and improves the 5-year overall survival of patients (Yeoh et al., 2018). CX4945, an inhibitor of CK2, can restore Ikaros function and play an anti-leukemia role *in vitro* or preclinical leukemia model (Borgo et al., 2021). In breast cancer, the application of ginseng polysaccharides can inhibit the proliferation of MDA-MB-231 cells by activating IKZF1 (Zhou et al., 2020). Through the inhibition of the CTLA4 *via* ipilimumab and the application of adoptive cellular immunotherapy with tumor-infiltrated lymphocytes, it enhanced the anti-tumor effects of T cells (Besser et al., 2010; Hodi et al., 2010).

The previous studies have revealed that SKCM patients with a high abundance of T and B cells or lowly infiltrated M0 and M2 macrophages have a significant prognosis, while the metastatic SKCM patients with low infiltration of B and $CD8^+$ T cells have a worse outcome than the previous condition (Iglesia et al., 2016; Wang et al., 2020). The present study suggests that expressed *IKZF1-3* could activate the B cell receptor signaling pathway, T cell receptor signaling pathway, leukocyte trans-endothelial migration, antigen processing and presentation and several other immune regulation processes. Obviously, *IKZF1-3* is positively correlated with T and B cells infiltration, but negatively correlated with the M0 and M2 infiltration. Moreover, Ikaros is specifically expressed in the immunocytes. It seems that *IKZF1-3* participate the regulation of immunocytes' differentiation, thus making it could be considered as an indicator for the status of TME in SKCM patients. Besides the regulation of hematopoietic process, Ikaros also regulates the balance of autoimmunity and suppresses the growth of tumor (Fan and Lu, 2016). Ikaros genes mediate lymphocyte proliferation and differentiation (Fan and Lu, 2016). *IKZF1-3* play an important role in the regulation of lymphatic system differentiation. The disorder of its expression has the directive relationship with the hematological malignancies and primary immunodeficiency. Lack of Ikaros family may lead to a variety of immune related diseases, including immune thrombocytopenia (Sriaroon et al., 2019), systemic lupus erythematosus (Chen et al., 2020), rheumatoid arthritis (Yang et al., 2019). For example, mutations of *IKZF1–3* in leukemias are associated with poor prognosis, mainly caused by the disruption of lymphocyte fates (Rebollo and Schmitt, 2003; Payne and Dovat, 2011). In addition, Some solid tumors are related to the abnormal expression of Ikaros family proteins. Higher levels of Ikaros have been proved to be related to poor differentiation and late stage of ovarian cancer (He et al., 2012). Moreover, hypermethylation of Ikaros levels can be regarded as a marker of colorectal cancer progression and poor prognosis (Javierre et al., 2011). As a nuclear transcription factor, IKZF1 could inhibit the proliferation of HCC through suppressing the promotor of ANXA4D (Liu et al., 2017). In lung cancer, overexpressed IKZF3 upregulates the expression of TWIST and matrix metalloproteinase-16, which promotes the epithelial-mesenchymal-transition and the transformation of cancer stem cells and leads to poor prognosis (Hung et al., 2020). Moreover, the highly expressed Ikaros are closely associated with recurrence and metastasis lung adenocarcinoma (Zhao et al., 2020). But, in several tumors that normally lack IKZF1 expression, overexpression of Ikaros leads to enhanced immune recruitment infiltration and tumor sensitivity to CTLA4 and PD1 inhibitors (Chen et al., 2018). However, no research, to our knowledge, has investigated the correlation

between Ikaros expression and SKCM with the aim of developing markers to predict prognosis and tumor progression.

In the present study, we first determined the associations between patients' overall characteristics and clinical prognosis with the expression of Ikaros across cancer types. We employed different analytical methods to identify the relationship between the transcriptional properties of *IKZF1–5* and different cancers (Figures 1–3). For patients with SKCM, the expression of *IKZF1–3* significantly decreased with increased T-stage or metastasis (Figure 4). Moreover, highly expressed *IKZF1–3*, but not *IKZF4–5*, were positively associated with a favorable prognostic outcome (Figure 5) of patients with SKCM through regulating immunocyte infiltration (Figure 9) and the immune response (Figure 6, 7). Together, these results indicate that highly expressed *IKZF1–3* gain clinical significance for predicting the prognosis of patients with SKCM, *IKZF3* is an independent prognostic factor of SKCM, thus suggesting that *IKZF3* are prognostic biomarkers and immunotherapeutic targets of melanoma. To the best of our knowledge, there is no prior study that comprehensively investigated the relationship between Ikaros gene expression and SKCM.

A limitation of the present study is the lack of experimental evidence that activation of *IKZF1–3*-related pathways predict the prognosis of SKCM. Therefore, detailed epigenetic regulation of Ikaros during the pathogenesis of melanoma should be further investigated.

In conclusion, our present study systematically describes the distributions and functions of Ikaros family genes across cancer types. Although the roles of other Ikaros family members such as *IKZF4–5* in melanoma remain unclear, *IKZF3* may serve as biomarkers for the outcomes of treatments for SKCM that achieve a positive clinical prognosis.

Data availability statement

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

Ethics statement

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

Author contributions

L-KY: Conception and design, Provision of study material, Collection and/or assembly of data, Data analysis and interpretation, Manuscript writing. C-XL: Manuscript writing. S-HL: Provision of study material, Collection and/or assembly of data. J-JL:Provision of study material, Collection and/or assembly of data. L-LX: Provision of study material, Collection and/or assembly of data. G-HX: Provision of study material, Collection and/or assembly of data, Data analysis and interpretation. H-WL: Conception and design, Administrative support, Data analysis and interpretation, Final approval of manuscript. XL: Conception and design, Financial support, Administrative support, Data analysis and interpretation, Final approval of manuscript.

Funding

This work was supported by the National Nature and Science Foundation, China (82102345, 81871563), the Medical Research Foundation of Guangdong Province(A2021165), the Fundamental Research Funds for the Central Universities (21619350), Guangdong Foundation for Basic and Applied Basic Research (2019A1515110161). Funding by Science and Technology Projects in Guangzhou Ensure to add all grant numbers and funding information, as after publication this will no longer be possible. All funders should be credited and all grant numbers should be correctly included in this section.(202201020002,202201020004,202201020468), the Clinical Frontier Technology Program of the First Affiliated Hospital of Jinan University, China (No. JNU1AF-CFTP-2022-a01231,No. JNU1AF-CFTP-2022-a01208).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1036402/full#supplementary-material

References

Atkins, M. B., Kunkel, L., Sznol, M., and Rosenberg, S. A. (2000). High-dose recombinant interleukin-2 therapy in patients with metastatic melanoma: Long-term survival update. *Cancer J. Sci. Am.* 6 (1), S11–S14.

Balch, C. M., Buzaid, A. C., Soong, S. J., Atkins, M. B., CasciNelliN.Coit, D. G., et al. (2001). Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J. Clin. Oncol.* 19 (16), 3635–3648. doi:10.1200/JCO.2001.19.16.3635

Besser, M. J., Shapira-Frommer, R., Treves, A. J., Zippel, D., Itzhaki, O., Hershkovitz, L., et al. (2010). Clinical responses in a phase II study using adoptive transfer of short-term cultured tumor infiltration lymphocytes in metastatic melanoma patients. *Clin. Cancer Res.* 16 (9), 2646–2655. doi:10.1158/1078-0432.CCR-10-0041

Borgo, C., D'Amore, C., Sarno, S., Salvi, M., and Ruzzene, M. (2021). Protein kinase CK2: A potential therapeutic target for diverse human diseases. *Signal Transduct. Target. Ther.* 6 (1), 183. doi:10.1038/s41392-021-00567-7

Chen, J. C., Perez-Lorenzo, R., Saenger, Y. M., Drake, C. G., and Christiano, A. M. (2018). IKZF1 enhances immune infiltrate recruitment in solid tumors and susceptibility to immunotherapy. *Cell Syst.* 7 (1), 92–103. doi:10.1016/j.cels.2018.05.020

Chen, L., Niu, Q., Huang, Z., Yang, B., Wu, Y., and Zhang, J. (2020). IKZF1 polymorphisms are associated with susceptibility, cytokine levels, and clinical features in systemic lupus erythematosus. *Med. Baltim.* 99 (41), e22607. doi:10.1097/MD.0000000000022607

Fan, Y., and Lu, D. (2016). The Ikaros family of zinc-finger proteins. *Acta Pharm. Sin. B* 6 (6), 513–521. doi:10.1016/j.apsb.2016.06.002

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136 (5), E359–E386. doi:10.1002/ijc.29210

Fu, Q., Chen, N., Ge, C., Li, R., Li, Z., Zeng, B., et al. (2019). Prognostic value of tumor-infiltrating lymphocytes in melanoma: A systematic review and meta-analysis. *Oncoimmunology* 8 (7), 1593806. doi:10.1080/2162402X.2019.1593806

Gao, S., Wang, S., and Song, Y. (2020). Novel immunomodulatory drugs and neo-substrates. *Biomark. Res.* 8, 2. doi:10.1186/s40364-020-0182-y

He, L. C., Gao, F. H., Xu, H. Z., Zhao, S., Ma, C. M., Li, J., et al. (2012). Ikaros inhibits proliferation and, through upregulation of Slug, increases metastatic ability of ovarian serous adenocarcinoma cells. *Oncol. Rep.* 28 (4), 1399–1405. doi:10.3892/or.2012.1946

Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., et al. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* 363 (8), 711–723. doi:10.1056/NEJMoa1003466

Hung, J. J., Kao, Y. S., Huang, C. H., and Hsu, W. H. (2020). Author Correction: Overexpression of Aiolos promotes epithelial-mesenchymal transition and cancer stem cell-like properties in lung cancer cells. *Sci. Rep.* 10 (1), 1309. doi:10.1038/s41598-020-57957-0

Iglesia, M. D., Parker, J. S., Hoadley, K. A., Serody, J. S., Perou, C. M., and Vincent, B. G. (2016). Genomic analysis of immune cell infiltrates across 11 tumor types. *J. Natl. Cancer Inst.* 108 (11), djw144. doi:10.1093/jnci/djw144

Javierre, B. M., Rodriguez-Ubreva, J., Al-Shahrour, F., Corominas, M., Grana, O., Ciudad, L., et al. (2011). Long-range epigenetic silencing associates with deregulation of Ikaros targets in colorectal cancer cells. *Mol. Cancer Res.* 9 (8), 1139–1151. doi:10.1158/1541-7786.MCR-10-0515

John, L. B., and Ward, A. C. (2011). The Ikaros gene family: Transcriptional regulators of hematopoiesis and immunity. *Mol. Immunol.* 48 (9-10), 1272–1278. doi:10.1016/j.molimm.2011.03.006

Lee, N., Zakka, L. R., Mihm, M. C., Jr., and Schatton, T. (2016). Tumour-infiltrating lymphocytes in melanoma prognosis and cancer immunotherapy. *Pathology* 48 (2), 177–187. doi:10.1016/j.pathol.2015.12.006

Liu, Y. Y., Ge, C., Tian, H., Jiang, J. Y., Zhao, F. Y., Li, H., et al. (2017). The transcription factor Ikaros inhibits cell proliferation by downregulating ANXA4 expression in hepatocellular carcinoma. *Am. J. Cancer Res.* 7 (6), 1285–1297.

Locy, H., de Mey, S., de Mey, W., De Ridder, M., Thielemans, K., and Maenhout, S. K. (2018). Immunomodulation of the tumor microenvironment: Turn foe into friend. *Front. Immunol.* 9, 2909. doi:10.3389/fimmu.2018.02909

Megahed, M., Schön, M., Selimovic, D., and Schön, M. P. (2002). Reliability of diagnosis of melanoma *in situ*. *Lancet (London, Engl.* 359 (9321), 1921–1922. doi:10.1016/S0140-6736(02)08741-X

Molnár, A., Georgopoulos, K., and Molnar, A. (1994). The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Mol. Cell. Biol.* 14 (12), 8292–8303. doi:10.1128/mcb.14.12.8292

Nazarian, R., Shi, H., Wang, Q., Kong, X., Koya, R. C., Lee, H., et al. (2010). Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* 468 (7326), 973–977. doi:10.1038/nature09626

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Payne, K. J., and Dovat, S. (2011). Ikaros and tumor suppression in acute lymphoblastic leukemia. *Crit. Rev. Oncog.* 16 (1-2), 3–12. doi:10.1615/critrevoncog.v16.i1-2.20

Perdomo, J., Holmes, M., Chong, B., and Crossley, M. (2000). Eos and pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. *J. Biol. Chem.* 275 (49), 38347–38354. doi:10.1074/jbc.M005457200

Pourabdollah, M., Bahmanyar, M., Atenafu, E. G., Reece, D., Hou, J., and Chang, H. (2016). High IKZF1/3 protein expression is a favorable prognostic factor for survival of relapsed/refractory multiple myeloma patients treated with lenalidomide. *J. Hematol. Oncol.* 9 (1), 123. doi:10.1186/s13045-016-0354-2

Rebollo, A., and Schmitt, C. (2003). Ikaros, Aiolos and Helios: Transcription regulators and lymphoid malignancies. *Immunol. Cell Biol.* 81 (3), 171–175. doi:10.1046/j.1440-1711.2003.01159.x

Roma-Rodrigues, C., Mendes, R., Baptista, P. V., and Fernandes, A. R. (2019). Targeting tumor microenvironment for cancer therapy. *Int. J. Mol. Sci.* 20 (4), E840. doi:10.3390/ijms20040840

Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics. *Ca. Cancer J. Clin.* 66 (1), 7–30. doi:10.3322/caac.21332

Soengas, M. S., and Lowe, S. W. (2003). Apoptosis and melanoma chemoresistance. *Oncogene* 22 (20), 3138–3151. doi:10.1038/sj.onc.1206454

Sriaroon, P., Chang, Y., Ujhazi, B., Csomos, K., Joshi, H. R., Zhou, Q., et al. (2019). Familial immune thrombocytopenia associated with a novel variant in IKZF1. *Front. Pediatr.* 7, 139. doi:10.3389/fped.2019.00139

Sun, D., Wang, J., Han, Y., Dong, X., Ge, J., Zheng, R., et al. (2021). Tisch: A comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* 49 (D1), D1420–d1430. doi:10.1093/nar/gkaa1020

Sun, L., Liu, A., and Georgopoulos, K. (1996). Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. *EMBO J.* 15 (19), 5358–5369. doi:10.1002/j.1460-2075.1996.tb00920.x

Tachita, T., Kinoshita, S., Ri, M., Aoki, S., Asano, A., Kanamori, T., et al. (2020). Expression, mutation, and methylation of cereblon-pathway genes at pre- and post-lenalidomide treatment in multiple myeloma. *Cancer Sci.* 111 (4), 1333–1343. doi:10.1111/cas.14352

Tsai, K. K., Zarzoso, I., and Daud, A. I. (2014). PD-1 and PD-L1 antibodies for melanoma. *Hum. Vaccin. Immunother.* 10 (11), 3111–3116. doi:10.4161/21645515.2014.983409

Vitale, C., Falchi, L., Ten Hacken, E., Gao, H., Shaim, H., Van Roosbroeck, K., et al. (2016). Ofatumumab and lenalidomide for patients with relapsed or refractory chronic lymphocytic leukemia: Correlation between responses and immune characteristics. *Clin. Cancer Res.* 22 (10), 2359–2367. doi:10.1158/1078-0432.CCR-15-2476

Wang, X., Xiong, H., Liang, D., Chen, Z., Li, X., and Zhang, K. (2020). The role of SRGN in the survival and immune infiltrates of skin cutaneous melanoma (SKCM) and SKCM-metastasis patients. *BMC cancer* 20 (1), 378. doi:10.1186/s12885-020-06849-7

Weber, J. (2010). Immune checkpoint proteins: A new therapeutic paradigm for cancer--preclinical background: CTLA-4 and PD-1 blockade. *Semin. Oncol.* 37 (5), 430–439. doi:10.1053/j.seminoncol.2010.09.005

Yang, M., Liu, Y., Mo, B., Xue, Y., Ye, C., Jiang, Y., et al. (2019). Helios but not CD226, TIGIT and Foxp3 is a potential marker for CD4+ treg cells in patients with rheumatoid arthritis. *Cell. Physiol. biochem.* 52 (5), 1178–1192. doi:10.33594/000000080

Yeoh, A., Lu, Y., Chin, W., Chiew, E. K. H., Lim, E. H., Li, Z., et al. (2018). Intensifying treatment of childhood B-lymphoblastic leukemia with IKZF1 deletion reduces relapse and improves overall survival: Results of Malaysia-Singapore ALL 2010 study. *J. Clin. Oncol.* 36 (26), 2726–2735. doi:10.1200/JCO.2018.78.3050

Yoshida, T., and Georgopoulos, K. (2014). Ikaros fingers on lymphocyte differentiation. *Int. J. Hematol.* 100 (3), 220–229. doi:10.1007/s12185-014-1644-5

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

Zhao, W., Chen, T. B., and Wang, H. (2020). Ikaros is heterogeneously expressed in lung adenocarcinoma and is involved in its progression. *J. Int. Med. Res.* 48 (8), 300060520945860. doi:10.1177/0300060520945860

Zhou, H., Yan, Y., Zhang, X., Zhao, T., Xu, J., and Han, R. (2020). Ginseng polysaccharide inhibits MDA-MB-231 cell proliferation by activating the inflammatory response. *Exp. Ther. Med.* 20 (6), 229. doi:10.3892/etm.2020.9359

Zhou, J., Dudley, M. E., Rosenberg, S. A., and Robbins, P. F. (2005). Persistence of multiple tumor-specific T-cell clones is associated with complete tumor regression in a melanoma patient receiving adoptive cell transfer therapy. *J. Immunother.* 28 (1), 53–62. doi:10.1097/00002371-200501000-00007

Zhou, N., Gutierrez-Uzquiza, A., Zheng, X. Y., Chang, R., Vogl, D. T., Garfall, A. L., et al. (2019). RUNX proteins desensitize multiple myeloma to lenalidomide via protecting IKZFs from degradation. *Leukemia* 33 (8), 2006–2021. doi:10.1038/s41375-019-0403-2

Glossary

TME Tumor microenvironment

TIL Tumor-infiltrating lymphocyte

GSEA Gene Set Enrichment Analysis

HR Hazard ratio

OS Overall survival

PFI Progression-free interval

DSS Disease-free survival

ACC Adrenocortical carcinoma

BLCA Bladder Urothelial Carcinoma

BRCA Breast invasive carcinoma

CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma

CHOL Cholangiocarcinoma

COAD Colon adenocarcinoma

DLBC Lymphoid Neoplasm Diffuse Large B-cell Lymphoma

ESCA Esophageal carcinoma

GBM Glioblastoma multiforme

HNSC Head and Neck squamous cell carcinoma

KICH Kidney Chromophobe

KIRC Kidney renal clear cell carcinoma

KIRP Kidney renal papillary cell carcinoma

LAML Acute Myeloid Leukemia

LGG Brain Lower Grade Glioma

LIHC Liver hepatocellular carcinoma

LUAD Lung adenocarcinoma

LUSC Lung squamous cell carcinoma

MESO Mesothelioma

OV Ovarian serous cystadenocarcinoma

PAAD Pancreatic adenocarcinoma

PCPG Pheochromocytoma and Paraganglioma

PRAD Prostate adenocarcinoma

READ Rectum adenocarcinoma

SARC Sarcoma

SKCM Skin Cutaneous Melanoma

STAD Stomach adenocarcinoma

TGCT Testicular Germ Cell Tumors

THCA Thyroid carcinoma

THYM Thymoma

UCEC Uterine Corpus Endometrial Carcinoma

UCS Uterine Carcinosarcoma

UVM Uveal Melanoma

MDS Myelodysplastic syndrome

MM Multiple myeloma

IMiD Immunomodulatory drugs

CLL Chronic lymphocytic leukemia

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 November 2022
DOI 10.3389/fgene.2022.1003754

OPEN ACCESS

EDITED BY
Yi Yao,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Wan Qin,
Tongji Hospital, China
Zhengyang Hu,
Fudan University, China

*CORRESPONDENCE
Yueran Li,
408635799@qq.com

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

RECEIVED 26 July 2022
ACCEPTED 14 November 2022
PUBLISHED 25 November 2022

CITATION
Li Y, Peng G, Qin C, Wang X, Li Y and Li Y (2022), Positive regulators of T cell proliferation as biomarkers for predicting prognosis and characterizing the immune landscape in lung adenocarcinoma.
Front. Genet. 13:1003754.
doi: 10.3389/fgene.2022.1003754

Positive regulators of T cell proliferation as biomarkers for predicting prognosis and characterizing the immune landscape in lung adenocarcinoma

Yang Li[1,2], Gang Peng[2], Chaoying Qin[2], Xiangyu Wang[2], Yue Li[2] and Yueran Li [1,3]*

[1]Department of Laboratory Medicine, Third Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, Hunan, China, [3]Department of Gynecology, Third Xiangya Hospital, Central South University, Changsha, Hunan, China

Lung adenocarcinoma (LUAD) is the one of the most prevalent and fatal form of malignant tumors worldwide. Recently, immunotherapy is widely used in the treatment of patients with LUAD and has proved to be clinically effective in improve the prognosis of patients. But there still has been a tremendous thrust to further improve the efficacy of immunotherapy in individual patients with LUAD. The suppression of T cells and their effector functions in the tumor microenvironment (TME) of LUAD is one of the primary reasons for the low efficacy of immunotherapy in some patients with LUAD. Therefore, identifying positive regulators of T cell proliferation (TPRs) may offer novel avenues for LUAD immunotherapy. In this study, we comprehensively evaluated the infiltration patterns of TPRs in 1,066 patients with LUAD using unsupervised consensus clustering and identified correlations with genomic and clinicopathological characteristics. Three infiltrating TPR clusters were defined, and a TPR-related risk signature composed of nine TPRs was constructed using least absolute shrinkage and selection operator-Cox regression algorithms to classify the individual TPR infiltration patterns. Cluster 1 exhibited high levels of T cell infiltration and activation of immune-related signaling pathways, whereas cluster 2 was characterized by robust T cell immune infiltration and enrichment of pathways associated with carcinogenic gene sets and tumor immunity. Cluster 3 was characterized as an immune-desert phenotype. Moreover, the TPR signature was confirmed as an independent prognostic biomarker for drug sensitivity in patients with LUAD. In conclusion, the TPR signature may serve as a novel tool for effectively characterizing immune characteristics and evaluating the prognosis of patients with LUAD.

KEYWORDS

lung adenocarcinoma, driver of T cell proliferation, tumor microenvironment, prognosis, immune prognostic model

Introduction

Lung cancer is the most frequently diagnosed and most lethal cancer worldwide, with a 5-year relative survival rate of 21% (Siegel et al., 2021). Lung adenocarcinoma (LUAD) is the most common pathological type of non-small cell lung cancer (NSCLC), accounting for 50% of lung cancers (Travis, 2011; Brustugun et al., 2018). Surgery remains the primary treatment for patients with stage I LUAD, but the prognosis remains poor, owing to the prevalence of metastasis before diagnosis (Herbst et al., 2018). The risk of recurrence 5 years after surgery is as high as 27% (Yan et al., 2009). Recent advances in targeted therapies for driver genes of LUAD may reduce metastasis, delay postoperative recurrence, and improve patient survival rates (Mayekar and Bivona, 2017). For example, targeted therapies that employ epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) against tumors with EGFR mutations or anaplastic lymphoma kinase (ALK) TKIs against tumors with ALK fusions have improved the outcomes in a subset of patients (Saito et al., 2018; Harrison et al., 2020). However, these target gene mutations are only present in 15%–20% of patients, and targeting agents are ineffective in a small portion of patients with advanced LUAD (Mayekar and Bivona, 2017). Therefore, novel biomarkers and therapeutic targets are needed to predict prognosis and improve the survival of patients with LUAD.

The tumor microenvironment (TME) refers to the ecosystem surrounding the tumor, which includes immune cells, blood vessels, extracellular matrix, stromal cells, and signaling molecules (Anderson and Simon, 2020). Recent studies have shown that interactions between the tumor and TME play an important role in LUAD initiation, development, and progression (Hanahan and Coussens, 2012; Altorki et al., 2019). Studies elucidating the molecular and cellular biology of the TME have led to the development of novel immunotherapy strategies, including checkpoint blockade, adoptive cellular therapy, and cancer vaccinology (Waldman et al., 2020). Drugs targeting various components of the TME have been used in clinical trials and have demonstrated durable responses in patients with NSCLC (Gettinger et al., 2016; Herbst et al., 2018). Immune checkpoint blockades of programmed death-1 (PD-1) and its ligand, PD-L1, are the most effective treatments for LUAD, as they positively regulate T cell activation. As one of the most effective anti-PD-1 drugs, nivolumab has been shown to significantly improve the 5-year overall survival (OS) of patients with advanced NSCLC, compared with chemotherapy (Saka et al., 2021). However, the clinical efficacy of anti-PD-1 drugs has been reported in only 10% of patients with PD-L1-expressing tumors (Borghaei et al., 2015), and most patients with PD-L1$^+$ tumors respond shortly.

Adoptive T cell (ATC) therapy, which involves the infusion of autologous or allogeneic T cells, is an efficient and promising cancer treatment approach. Allogenic hematopoietic stem cell transplantation was the first effective adoptive transfer approach used for the clinical treatment of leukemia, and the T cell graft-versus-tumor effect produced an improved prognosis (Weiden et al., 1979). Recently, a novel ATC therapy using autologous patient T cells redirected against specific antigens was shown to be an efficient treatment for blood cancers and has been approved for clinical applications (Munshi et al., 2021). However, the response and cure rates still require improvement, especially in the treatment of solid tumors. Owing to the suppression of T cell effector functions in the TME of solid tumors, the efficiency of chimeric antigen receptor T therapy in solid tumors is much lower than that in blood cancers. Moreover, the generation of adaptive immune responses in patients with cancer depends on the antigen-specific activation of naive T cells and the coordination of T cell signaling. Thus, regulators of T cell proliferation (TPRs) in solid tumors may be ideal targets for improving ATC immunotherapy.

In this study, we comprehensively evaluated the characteristics of TPRs in 526 patients with LUAD and identified three subgroups of TPRs associated with distinct immune infiltration patterns, prognoses, genomic features, and clinicopathological characteristics. We then established a TPR-related risk model to quantify T cell activation patterns in individuals. The model was shown to be a robust prognostic factor and predictive biomarker for the response to drugs in patients with LUAD.

Materials and methods

Data collection

A total of 1,883 patients with LUAD from six independent datasets were included in this study. TRP-related genes were extracted from the Gene Set Enrichment Analysis (GSEA) database (https://www.gsea-msigdb.org/gsea/msigdb/cards/GOBP_ACTIVATED_T_CELL_PROLIFERATION.html) and Legut et al. (2022) report (Supplementary Table S1). Gene expression data, gene mutation data, and LUAD clinical profiles from The Cancer Genome Atlas (TCGA) were acquired from the XENA database (https://xena.ucsc.edu/). The gene expression matrix and corresponding survival files from GSE68465, GSE50081, and GSE72094, based on the Affymetrix Human Genome platform, were downloaded from the Gene Expression Omnibus database. Gene expression data and survival profiles for the validated cohort were extracted from the GEO database (GSE42127 and GSE36471 datasets). The Affy package was used to perform a background adjustment among these datasets (Gautier et al., 2004). According to the empirical Bayes framework using the

TABLE 1 Clinical baseline features of the LUAD patients in three databases.

	GSE36471		GSE42127		TCGA-LUAD		*p*-value
	High	Low	High	Low	High	Low	
	(*N* = 58)	(*N* = 58)	(*N* = 88)	(*N* = 88)	(*N* = 250)	(*N* = 250)	
TPS							
Subtype 1	37 (63.8%)	8 (13.8%)	15 (17.0%)	23 (26.1%)	46 (18.4%)	161 (64.4%)	<0.001
Subtype 2	11 (19.0%)	31 (53.4%)	23 (26.1%)	50 (56.8%)	99 (39.6%)	38 (15.2%)	
Subtype 3	10 (17.2%)	19 (32.8%)	50 (56.8%)	15 (17.0%)	105 (42.0%)	51 (20.4%)	
Event							
Death	36 (62.1%)	30 (51.7%)	41 (46.6%)	23 (26.1%)	116 (46.4%)	66 (26.4%)	<0.001
Alive	22 (37.9%)	27 (46.6%)	47 (53.4%)	65 (73.9%)	134 (53.6%)	184 (73.6%)	
Missing	0 (0%)	1 (1.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	

sva package, we adjusted and removed batch effects between the different expression profiles, which were subsequently merged to form a mixed cohort for further analyses. To prevent influencing the accuracy of patient survival predictions, this study did not include patients without prognostic data. The baseline information of the patients with LUAD is shown in Table 1.

Unsupervised consensus clustering of T cell proliferations

The ConsensusClusterPlus package was used to perform consensus analysis. LUAD samples were divided into three clusters based on significantly differential TPR-associated gene expression levels (false discovery rate <0.05 and |fold change| > 0.5) (Wilkerson and Hayes, 2010). Among the different k-means clustering results (k = 2–7), three groups (k = 3) demonstrated the most stable discrimination.

Survival and clinical analysis

OS was evaluated in each group using the Kaplan-Meier (KM) method and compared among groups using the log-rank test. The chi-square test was used to compare differences between groups. The threshold for statistical significance was defined as a p-value less than 0.05.

Pathway enrichment analysis

To exploit the potential processes between TPR group, we utilized the limma package [PMID: 25605792] to perform the differential expression analysis between different TPR and risk group. Firstly, the differential expression genes (DEGs) were obtained between TPRs group and risk group by differential expression analysis using limma package. Then, we screened the DEGs at certain condition (log2FC > 1 and adjust p-value < 0.05). Final, those DEGs were enrolled to performed next step enrichment analysis. The GSEA analysis of different groups from two independent cohorts was performed (Powers et al., 2018). The profiles extracted from the GSEA database (http://www.gsea-msigdb.org/gsea/downloads.jsp; project: h.all.v7.5.1. symbols.gmt) were analyzed using a reference gene set. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were conducted for the different groups using the clusterProfiler package.

Immune infiltration analysis

The MCP-counter method can be used to infer the immune and stromal cell composition of heterogeneous tissue (Aran et al., 2017; Petitprez et al., 2020). The IOBR package was used to assess T cell infiltration *via* the MCP-counter method (Zeng et al., 2021). The ESTIMATE algorithm was used to evaluate the immune score and stromal score in different groups *via* the IOBR package (Yoshihara et al., 2013). Tumor immune dysfunction and exclusion (TIDE) scores were calculated using the TIDE online database (http://tide.dfci.harvard.edu/) (Fu et al., 2020). The T cell exhaust score (gene set come from IOBR package) and T cell activation score (gene set come from: http://cis.hku.hk/TISIDB/index.php) were calculated by ssGSEA algorithm.

Construction and validation of the least absolute shrinkage and selection operator-Cox regression model

We first identified TPR-related genes that were significantly differentially expressed between LUAD tissues and normal lung

tissues. Univariate Cox regression analysis was used to determine the OS associated with TPR-related genes ($p < 0.05$). Finally, LASSO-Cox regression analysis was performed. Nine key TPRs in LUAD were identified and used to construct a TPR-related risk model for LUAD. The risk score for each patient was calculated using the following formula: risk score = −0.004491532* AGER (gene expression level) + (−0.077486959* CYP27A1) + 0.113176339* CDK1 + (−0.0453135648* CADM1) + 0.342268624* FADD+ 0.133393571* ADA +0.024214523* LTBR + (−0.022959715* FYN) + (−0.199900841* CRTAM).

Predictive efficacy of the model

Time-dependent receiver operating characteristic (ROC) curves were used to assess 1-, 3-, and 5-year OS. The predictive efficacy of the risk model was determined by assessing the area under the curve (AUC).

Correlations between clinical characteristics and T cell proliferation signature

Correlations between clinical features (age, sex, stage, and TNM stage) and the TPR signature were evaluated using the chi-square test. The TPR signature was then differentiated into subgroups based on these clinical characteristics. Univariate and multivariate Cox regression analyses were used to identify independent indicators of patient survival.

Nomogram construction and assessment

Univariate Cox regression analysis was performed to screen for significant factors ($p < 0.05$), which were subjected to further multivariate analysis and used for nomogram construction. The concordance index (C-index) was used to compare the predictive ability of the nomogram and the clinical features. Calibration plots were constructed to determine the fitting efficiency between the nomogram-predicted OS and actual OS. Decision curve analysis was used to assess the threshold expectation range of the nomogram in association with clinical characteristics.

Relationship between chemoresistance and T cell proliferation signature

The half-maximal inhibitory concentration (IC50) of FDA-approved drugs (rapamycin, cisplatin, paclitaxel, bortezomib, elesclomol, tipiifarnib, nilotinib, and doxorubicin) was determined for each TCGA-LUAD patient using the pRRophetic package. The IC50 was used to differentiate between high and low risk scores.

RNA extraction and quantitative PCR

Total RNA was extracted from U87 cells using TRIzol reagent (Invitrogen, Carlsbad, CA, United States), and reverse transcription was performed using the PrimeScript™ RT Reagent Kit (Takara, Dalian, China). cDNA was subjected to RT-qPCR using the SYBR Green Real-Time PCR Kit (Takara, Dalian, China). Relative mRNA expression were normalized to that of β-actin. The relative expression were calculated using the $2^{-\Delta\Delta CT}$ method.

Statistical analysis

R (version 4.0.2) was used for statistical analysis. A p-value < 0.05 was regarded as indicative of a statistically significant difference. Comparisons between two groups were conducted using Student's t-test or the Kruskal-Wallis H test, and comparisons among three or more groups were conducted using the Wilcoxon signed-rank test. Clinicopathological data for TCGA-LUAD patients grouped by the TPR model were analyzed using the chi-square test, and the log-rank test was used for survival analysis.

Results

Characterization of T cell proliferation patterns

The TPR infiltration patterns were systematically evaluated, and a TPR signature was constructed (Figure 1A). We integrated 1,066 LUAD samples from the same GEO platform and constructed T cell proliferation clusters (TPCs) in the mixed cohort (GSE68465, GSE50081, and GSE72094). Principal component analysis revealed changes in the sample distribution before and after integration (Figures 1B,C).

To determine the optimal cluster number, we identified differentially expressed TPRs between LUAD tumors and normal lung tissues in TCGA-LUAD cohort (Figures 2A,B). Next, we evaluated the clustering stability using the ConsensusClusterPlus package, which indicated the existence of three powerful TPCs in both the mixed cohort and TCGA-LUAD cohort (Figure 2C). In addition, the KM survival curves revealed that the three main TPCs in the mixed and TCGA-LUAD cohort exhibited significant differences (log-rank test, $p < 0.05$; Figure 2D–E). In particular, cluster 2 was associated with worse survival outcomes than the other clusters.

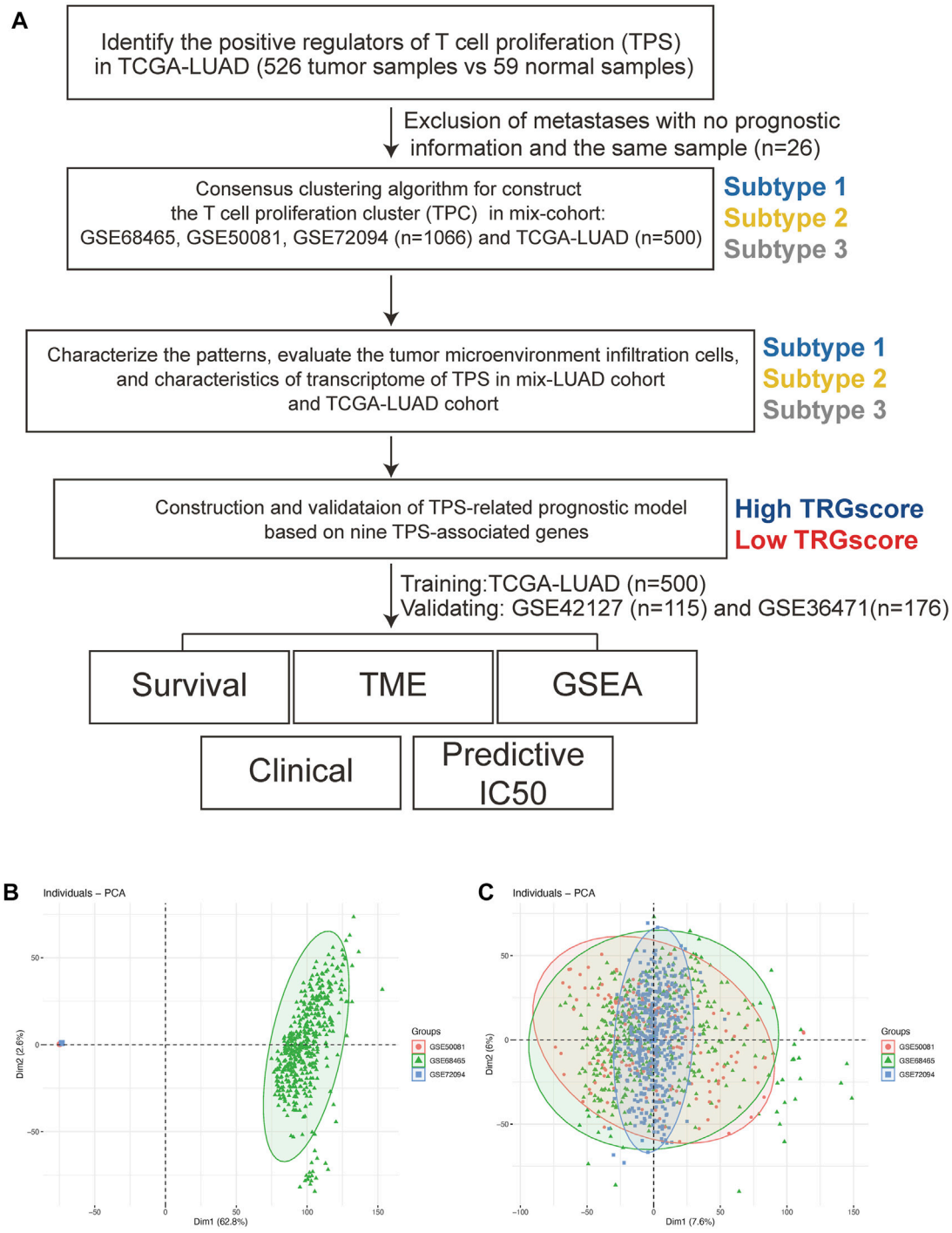


FIGURE 1
The workflow of this study. **(A)** The workflow chart of this study. **(B,C)** Principal component analyses before **(B)** and after **(C)** batch removal.

To exploit the potential processes between TPR group, we performed the differential analysis by Limma package. We explored the biological processes associated with the three TPR clusters by using the clusterProfiler package to perform KEGG pathway enrichment analysis and GSEA in the mixed cohort. Cluster 1 was markedly enriched in carcinogenic pathways, such as the cAMP signaling pathway, WNT signaling pathway, KRAS signaling pathway, and P53 pathway (Figures 3A,B; Supplementary Table S2). Cluster 2 exhibited enrichment in carcinogenic pathways (PI3K-AKT signaling pathway, MAPK

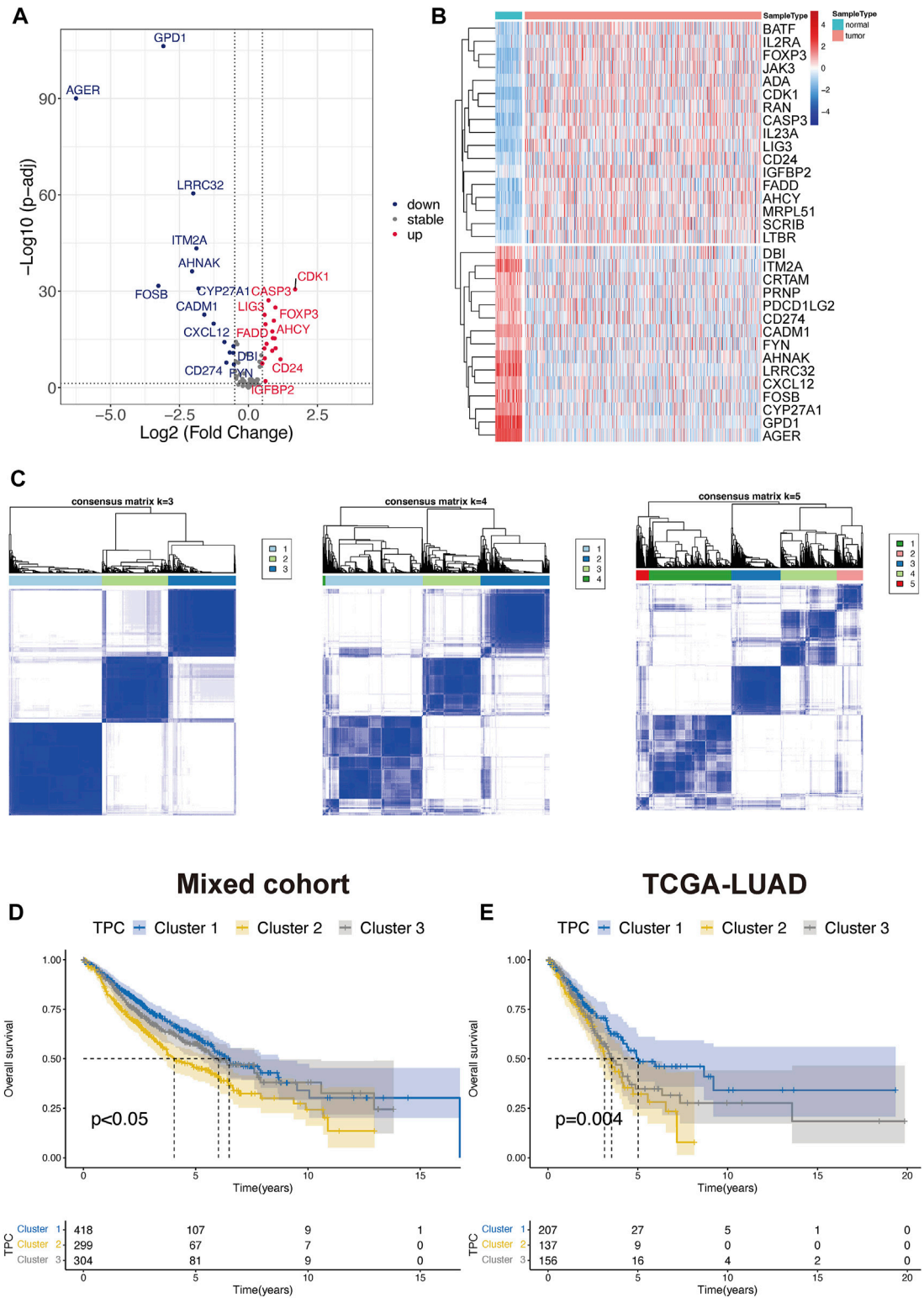


FIGURE 2
Identify the T cell proliferation cluster (TPC). **(A,B)** Volcano diagram and heatmap of the positive drive of T cell proliferation that depicts the abnormal differentially express pattern in lung adenocarcinoma and normal tissue samples. Blue dots: down-regulation, grey dots: none significance differential genes, and red dots: up-regulation. **(C)** Unsupervised hierarchical analyses of the differential expression patterns of these T cell-associated genes in mix-cohort (k-means = 3–5). **(D,E)** Comparison of overall survival between TPC by using Kaplan-Meier survival curves in mix-cohort **(D)** and TCGA-LUAD **(E)**.

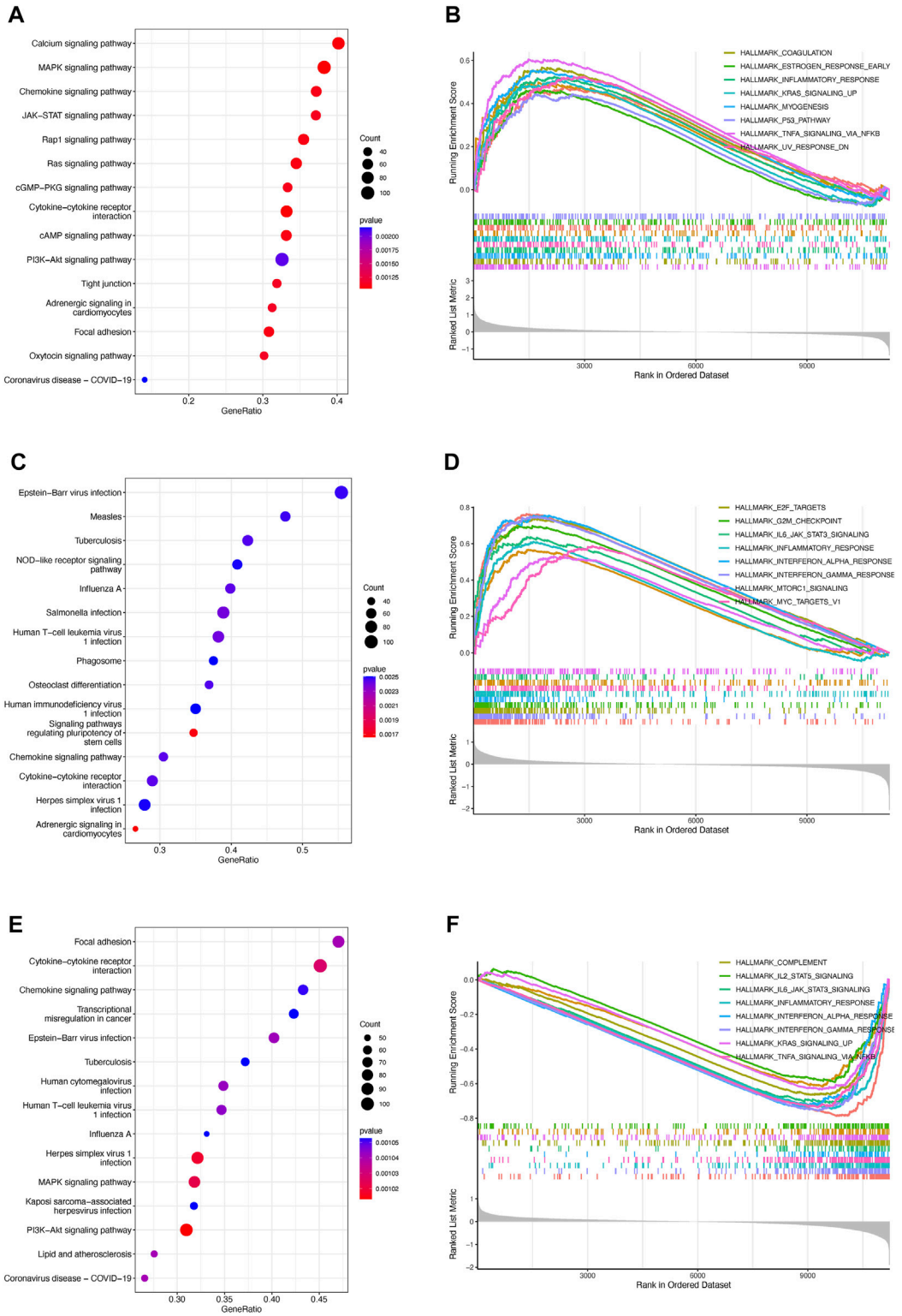


FIGURE 3
Enrichment analysis of TPRs **(A–F)**. Top 5 KEGG enriched gene pathway-related catalogs and top 10 GSEA of tumor-associated items in cluster 1 **(A,B)**, cluster 2 **(C,D)**, and cluster 3 **(E,F)**.

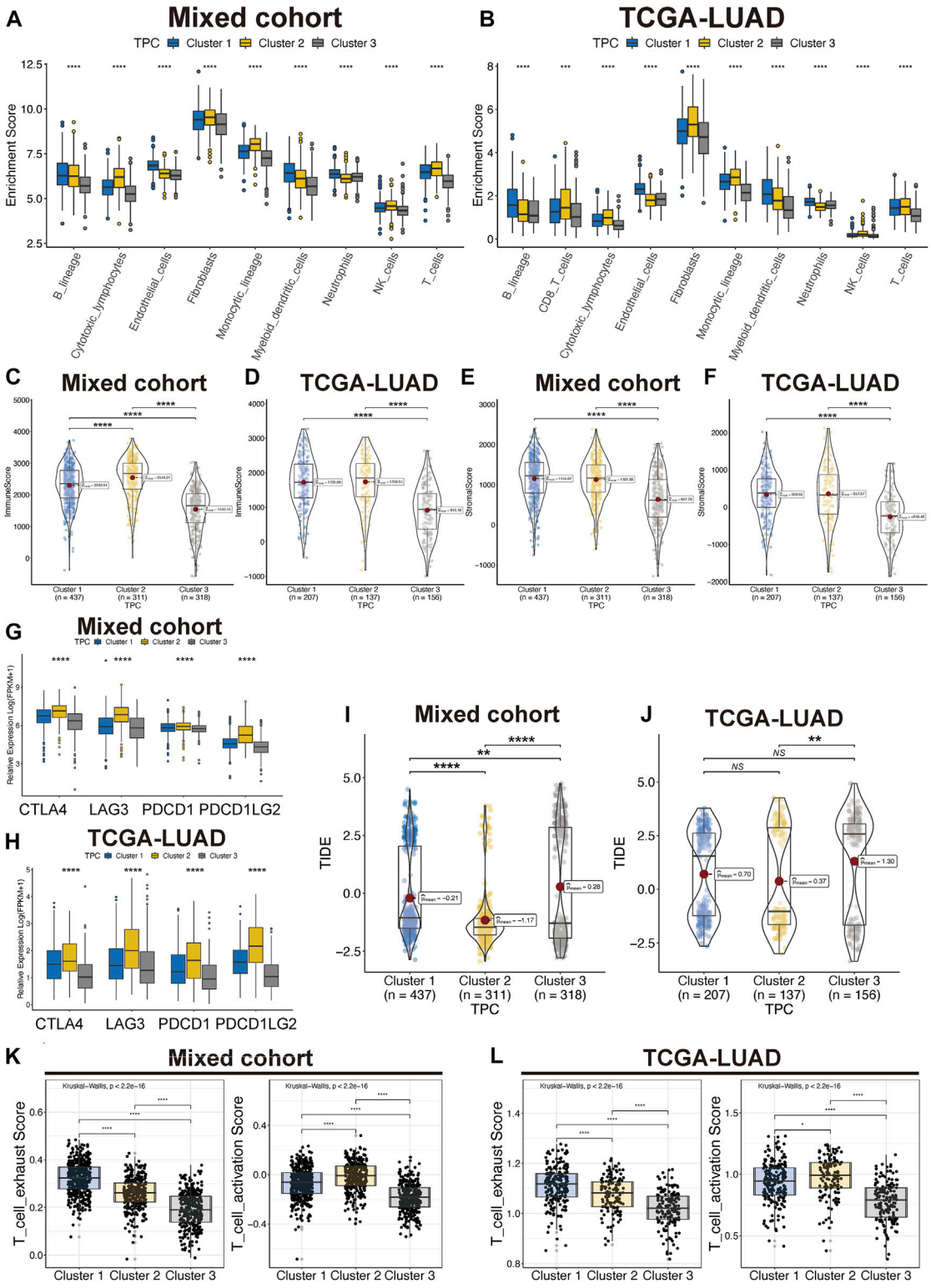


FIGURE 4
Tumor immune microenvironment of TPC. **(A,B)** Distribution of T cell in lung cancer and TME in mix-cohort **(A)** and TCGA-LUAD **(B)**. **(C–F)** ESTIMATE tumor purity algorithm was used to calculate the immune score and the stromal score of three TPC patients in the mix-cohort **(C,E)** and TCGA-LUAD **(D,F)** cohort. **(G,H)** Abnormal expression of immune checkpoint markers between TPC in mix-cohort **(G)** and TCGA-LUAD cohort **(H)**. **(I,J)** TIDE score of mix-cohort **(I)** and TCGA-LUAD **(J)** between TPC group. **(K,L)** T cell activation/exhaust score of mix-cohort **(K)** and TCGA-LUAD **(L)** between TPC group.

signaling pathway, and IL6-JAK-STAT3 signaling pathway) but also exhibited robust positive correlations with biological processes associated with immune activation, including cytokine-cytokine receptor interaction, interferon-gamma response, and inflammatory response (Figures 3C,D). Interestingly, cluster 3 annotations included pathways that were negatively associated with cluster 2 (Figures 3E,F).

Characteristics of T cell proliferation patterns in the tumor immune microenvironment

To identify correlations between TPR patterns and the TIME, we calculated the degree of infiltration of different types of immune and stromal cells by using single-sample GSEA to analyze the TPCs. The findings were consistent with the results shown in Figure 3, indicating that the degree of infiltration of the TPCs decreased in the following order in the mixed cohort and TCGA-LUAD cohort (Figures 4A,B): cluster 2 > cluster 1 > cluster 3. However, cluster 2, which had the highest degree of T cell infiltration and the strongest association with immune-related response pathways, was not associated with a corresponding survival advantage. Therefore, we evaluated the immune and stromal scores of the TPCs. In the mixed cohort and TCGA-LUAD cohort, cluster 2 had the highest scores, and cluster 1 had the lowest scores (Figures 4C–F; cluster 2 > cluster 1 > cluster 3). Although cluster 2 exhibited CD4/8 + T cell activation, this cluster also exhibited stromal cell activation, which exerts an immunosuppressive effect. Therefore, for these patients, immunosuppressive therapy may be suitable as a first-line treatment.

In addition, we examined the expression of four immune checkpoint genes (*PDCD1*, *PDCD1LG2*, *CTLA4*, and *LAG3*), which are related to immune blockage. The expression of these genes in the mixed cohort (Figure 4G) and TCGA-LUAD cohort (Figure 4H; Supplementary Table S3) decreased in the following order: cluster 2 > cluster 1 > cluster 3. Patients in cluster 2 had the lowest TIDE scores in both the TCGA-LUAD and mixed cohorts, suggesting that these patients are most likely to benefit from immunotherapy (Figures 4I,J). In addition, Cluster 2 have moderate exhaust and activation score of T cell (Figures 4K,L). Cluster 3 was characterized as an immune-desert phenotype. Cluster 2, which featured robust T cell immune filtration and a high stromal score, was characterized as an immune-inflamed phenotype. Cluster 1, which featured a moderate immune score and moderate immune infiltration with T cells, was characterized as an intermediate phenotype.

Construction and validation of T cell proliferation-related risk model

TPRs play a critical role in the regulation of different T cell functions. As TPR pattern prediction in individuals is not a suitable analysis method, TPCs were identified in the population. To account for the individual heterogeneity and complexity of TPR patterns, we aimed to construct a TPR-associated risk model to quantify the TPR patterns of individuals with LUAD. To illustrate TPR patterns in transcriptomic data, 11 TPRs were selected using univariate Cox regression analysis (Supplementary Table S4). LASSO-Cox regression was used to identify nine candidate prognostic genes, which were then used to establish the risk score (Figure 5A). The coefficients of each TPR are shown in Supplementary Table S3. The heatmap depicts the transcriptome characteristics associated with the risk score and the distribution of risk scores among the TPCs (Figure 5B). KM survival curves showed that the OS of LUAD patients with low-risk scores was better than that of those with high-risk scores (Figure 5C). The AUC for the time-dependent ROC curve, which was used to evaluate the predictive efficacy of the prognostic model, was 0.69 for 1-year survival, 0.70 for 3-year survival, and 0.73 for 5-year survival (Figure 5D).

To confirm the reproducibility and stability of the TPR signature, two independent LUAD cohorts acquired from the GEO database were used for external validation. The transcriptome features of the two validation sets were consistent with those of the training set (Figures 6A,B). KM survival analysis also indicated that patients in the validation cohorts with high-risk scores were associated with a poor OS, compared with those with low-risk scores (Figures 6C,D). Similarly, the AUCs for GSE42127 (Figure 6E; 1-year AUC = 0.80, 3-year AUC = 0.82, 5-year AUC = 0.80) and GSE36471 (Figure 6F; 1-year AUC = 0.68, 3-year AUC = 0.69, 5-year AUC = 0.68) indicated that the TPR signature exhibited excellent performance when used to predict the OS of LUAD patients.

Tumor mutation characteristics of T cell proliferation clusters

To investigate whether the distinct T cell prognostic phenotypes were determined by genetic events, we conducted an integrative analysis of the mutation data. We first explored the quantity and quality of somatic mutations in the high- and low-risk groups of the TCGA-LUAD cohort. As depicted in Figures 7A,B, the frequency of TP53 mutations was significantly higher in the high-risk group than in the low-risk group.

Mutually exclusive or co-occurring gene mutations are frequently observed in cancer patients (Kang et al., 2008; Kim et al., 2017). Detecting such mutation patterns is critical for identifying novel cancer signaling pathways and developing potential therapeutic strategies. As shown by the differences among the top 10 genes in the heatmap, TP53 and KRAS mutations exhibited mutual exclusivity ($p < 0.05$) in the high-risk group, but not in the low-risk group (Figure 7C). In addition, co-occurring mutations in TTN and KRAS were identified in the

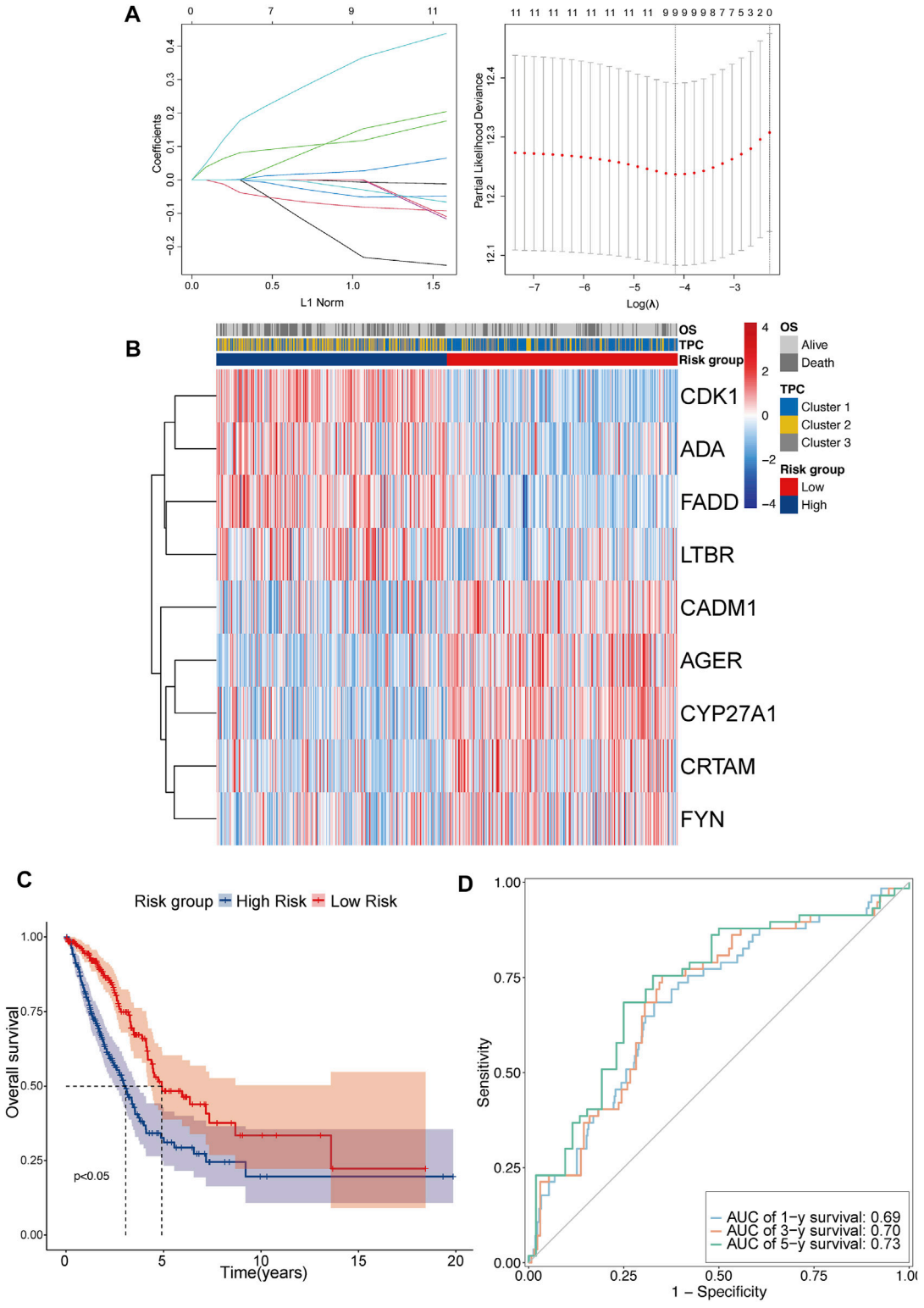


FIGURE 5
Construct the TPR-related risk model in TCGA-LUAD. **(A,B)** Lasso-Cox regression model analysis results and heatmap of T cell proliferation prognostic model signature: **(A)** lasso regression analysis (left panel); partial likelihood deviance for the lasso regression (right panel); dotted line: lambda.min (left) and lambda.se (right); **(B)** heatmap between high and low-risk scores and clinical parameters. **(C)** KM curves showing OS in patients with risk group; blue line: high-risk score (*n* = 250) and red line: low-risk score (*n* = 250). **(D)** Receiver operating characteristic (ROC) curve of 1 (light blue), 3 (orange), and 5 years (green).

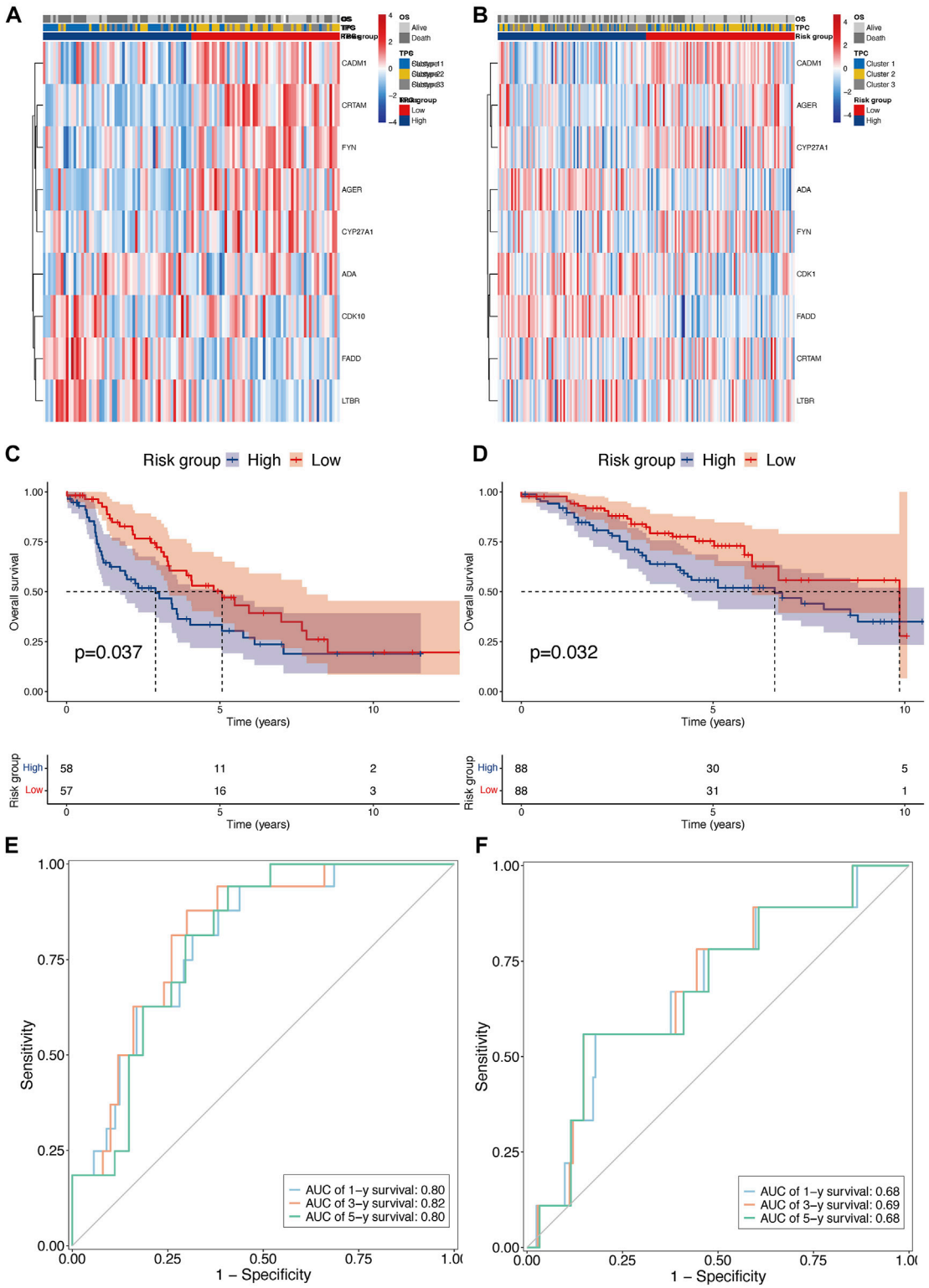


FIGURE 6
External validation of TPR signature in GSE42127 and GSE36471 cohorts. **(A,B)** Heatmap of nine TPR signature in external validate cohort (GSE36471 *n* = 115: left, and GSE42127 *n* = 176: right). **(C,D)** KM curves showing OS in patients with TPR signature (Left: GSE36471, and right: GSE42127). **(E,F)** AUC curves of 1 (light blue), 3 (orange), and 5 years (green) in both external cohorts (Left: GSE36471, and right: GSE42127).

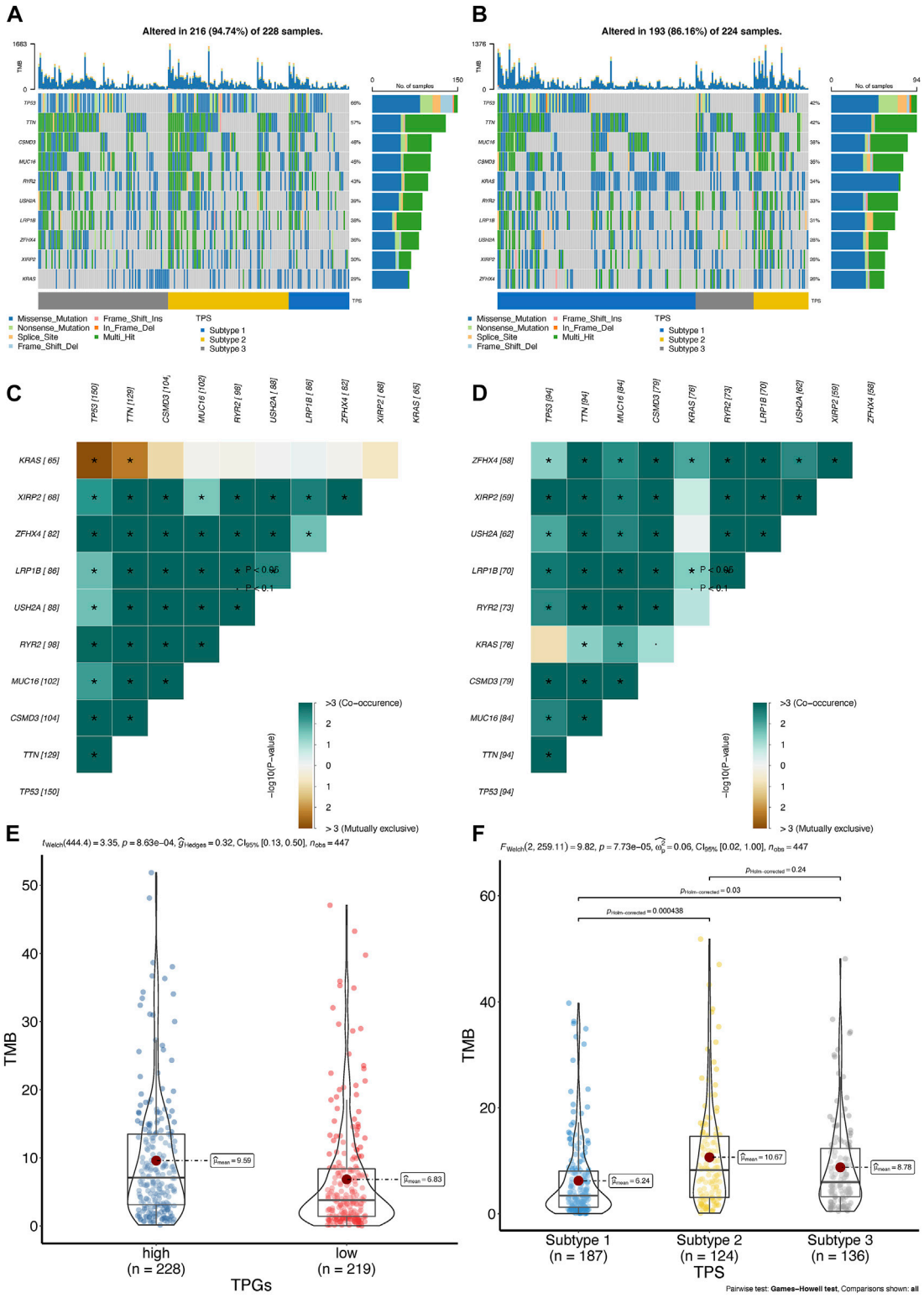


FIGURE 7
The landscape of somatic mutation between high- and low-risk group in TCGA-LUAD cohort. **(A,B)** Oncoplot of genes with highest counts of variants between high- **(A)**, *n* = 250 and low-risk score **(B)**, *n* = 250 in TCGA-LUAD cohort. **(C,D)** Significant exclusive or co-occurrence top 10 mutation gene sets are indicated in the high-risk score **(C)** and low-risk score **(D)** in TCGA-LUAD. **(E,F)** Distribution of tumor mutation burden in TPR signature **(E)** and TPC **(F)** group patients; TMB, tumor mutation burden.

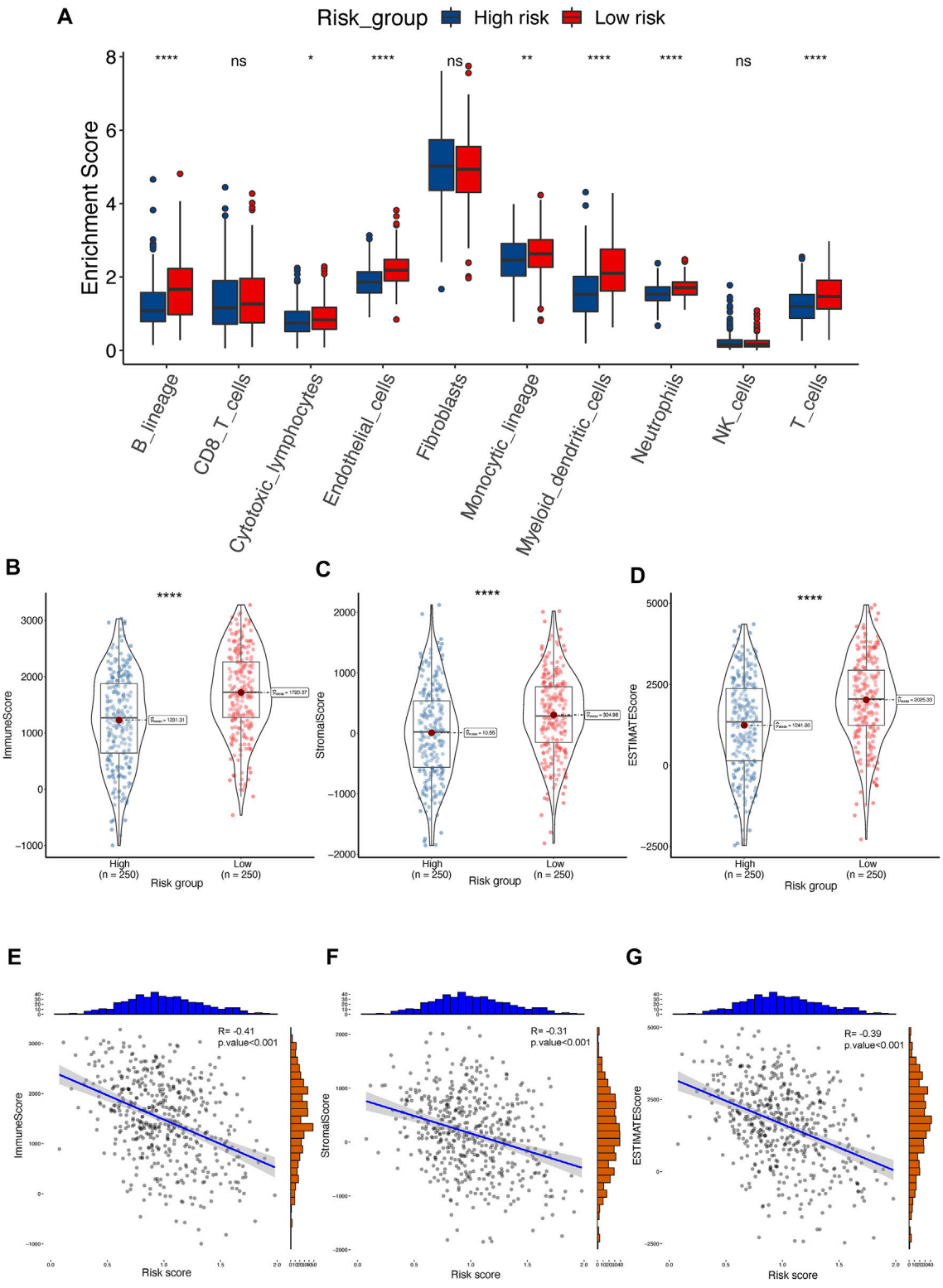


FIGURE 8
The landscape of TME between high- and low-risk groups in TCGA-LUAD cohort. **(A)** Distribution of T cell between TPR signature in TCGA-LUAD. **(B–D)** ESTIMATE tumor purity algorithm was utilized to calculate an immune score **(B)**, stromal score **(C)**, and tumor purity **(D)** of risk-related patients in the TCGA-LUAD cohort. **(E–G)** Pearson correlation analyses between risk score and immune score **(E)**, stromal score **(F)**, and estimate score **(G)** in the TCGA-LUAD cohort.

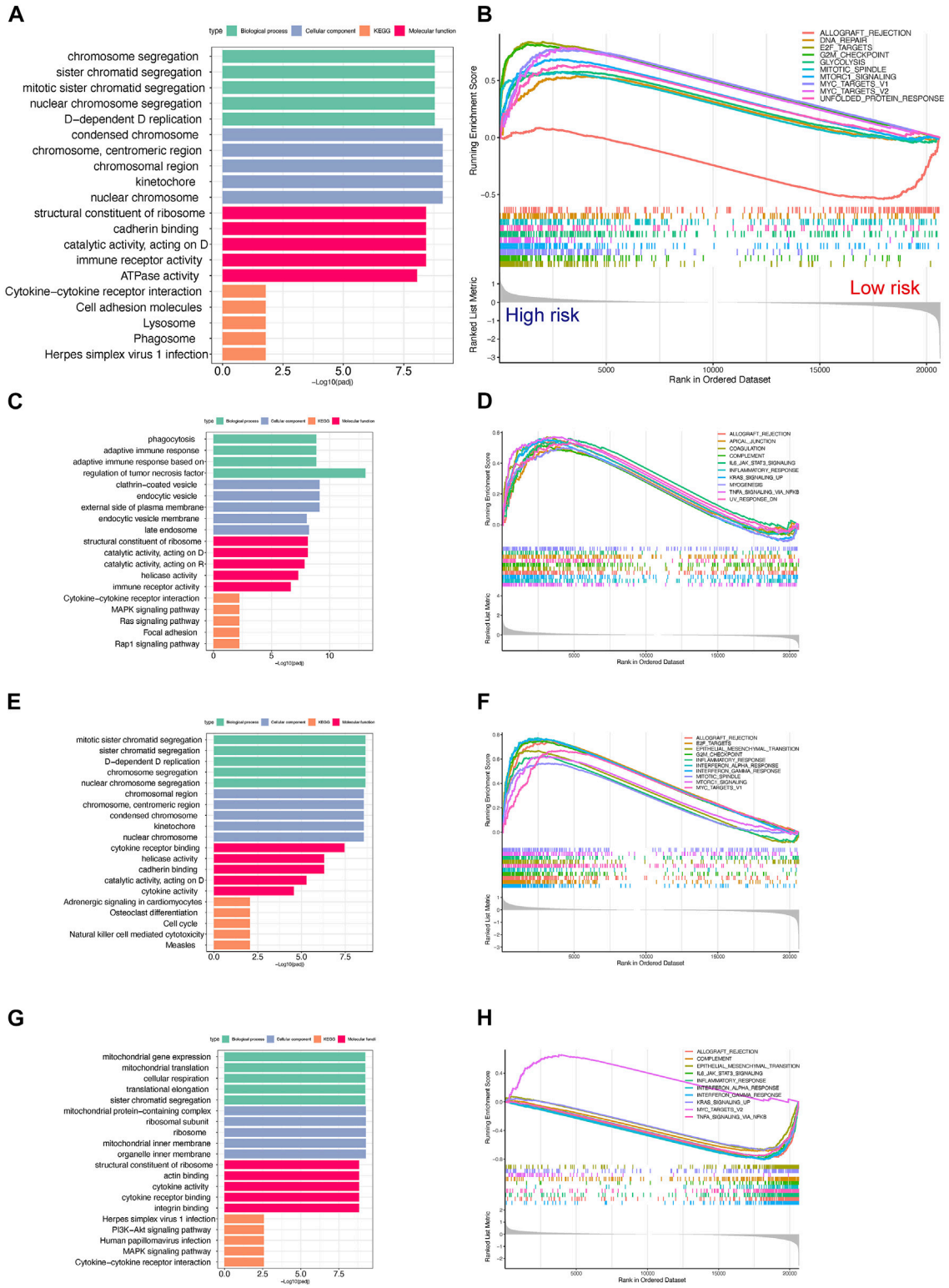


FIGURE 9
Gene set enrichment analysis for high- and low-risk scores in the TCGA-LUAD cohort. **(A,B)** Top 5 GO and KEGG enriched gene pathway-related catalogs **(A)** and top 10 GSEA **(B)** of tumor-related items between high- and low-risk scores in TCGA-LUAD **(C,D)**. Cluster 1 in GO/KEGG analysis **(C)** and GSEA **(D)** results in the TCGA-LUAD cohort (n = 207) **(E,F)**. Cluster 2 in GO/KEGG analysis **(E)** and GSEA **(F)** results in the TCGA-LUAD cohort (n = 137). **(G,H)** Cluster 3 in GO/KEGG analysis **(G)** and GSEA **(H)** results in the TCGA-LUAD cohort (n = 156).

low-risk group, whereas mutual exclusivity was observed in the high-risk group (Figure 7D).

In patients with LUAD, tumor mutation burden (TMB) has been regarded as an independent predictor of immunotherapy success (Goodman et al., 2017; Hellmann et al., 2018). As shown in the violin plot, patients belonging to the low-risk group or cluster 2 had a higher TMB, indicating that they may respond to PD-1/PD-L1 blockade therapy (Figures 7E,F). Therefore, we further assessed the differences in the TME among the TPR signature groups.

Characteristics of T cell proliferation-related risk model

We quantified immune cell infiltration using the MCP-counter algorithm to further investigate the association between TPRs and the TME. Consistent with the TMB analysis results, the low-risk group exhibited a higher degree of T cell infiltration than the high-risk group (Figure 8A). Moreover, the ESTIMATE algorithm results revealed a greater elevation in the immune score, stromal score, and estimate score in the low-risk group (Figures 8B–D). Similarly, the Pearson correlation coefficients also indicated that risk was negatively associated with the immune-associated scores (Figures 8E–G; immune score R = −0.41, stromal score R = −0.31, and estimate score R = −0.39; $p < 0.001$).

To further explore the potential differences in biological functions between the TPR-associated groups, GO and KEGG enrichment analyses of hallmark pathways in the high-risk and low-risk groups were performed. Chromatid segregation-related pathways and cytokine-cytokine receptor interaction pathways were significantly enriched (Figure 9A; Supplementary Table S5). Similar to cluster 3, the low-risk group displayed more enrichment in immune-related functions than the high-risk group, including the interferon-associated response, inflammatory response, and IL2/STAT5 signaling pathway (Figure 9B). Like the previous TPC results, the enrichment results for TCGA-LUAD cohort were consistent with those for the mixed cohort (Figures 9C–H).

Subgroup overall survival analysis

Clinical subgroup OS analysis demonstrated that the TPR signature was suitable for predicting survival in older (≥65 years), N2-N3 stage, M0 stage, or stage I-II LUAD patients. Among these patients, high risk was correlated with a notably poor OS. Sex and T stage did not affect the TPR model. Furthermore, a statistical difference (log-rank test) in OS between the high-risk group and the low-risk group was not observed in younger (<65 years), N0-N1 stage, M1 stage, or stage III-IV patients (Figures 10A–L).

Prediction of drug sensitivity

The predictive IC50 was calculated using the pRRophetic package. The high-risk group exhibited more sensitivity to doxorubicin (Figure 11A, High_median = −1.96 vs. Low_median = −1.90, $p < 0.001$), rapamycin (Figure 11B, High_median = −0.28 vs. Low_median = −0.03, $p < 0.001$), cisplatin (Figure 11C, High_median = −3.06 vs. Low_median = −3.30, $p < 0.001$), paclitaxel (Figure 11D, High_median = −3.18 vs. Low_median = −2.64, $p < 0.001$), bortezomib (Figure 11E, High_median = −5.31 vs. Low_median = −5.10, $p < 0.001$), and elesclomol (Figure 11F, High_median = −2.98 vs. Low_median = −2.77, $p < 0.001$). The low-risk group exhibited more sensitivity to tipifarnib (Figure 11G, High_median = 2.18 vs. Low_median = −2.14, $p < 0.001$) and nilotinib (Figure 11H, High_median = 4.42 vs. Low_median = −4.28, $p < 0.001$). In addition, we predicted the response rate to immunotherapy in the TCGA-LUAD cohort using the TIDE algorithm. These results indicated that immunotherapy may be more suitable for patients with a lower risk score (Figure 11I).

Nomogram construction and assessment

The results of the univariate analysis indicated that TPR signature, T stage, N stage, and M stage and stage were associated with OS (Figure 12A, $p < 0.001$). To exclude the interference of other phenotypes to the prognosis, we perform a multifactorial cox analysis of TPR signature in the TCGA-LUAD cohort. As shown in Figure 12B, the TPR signature was identified as independent prognostic variables associated with OS. Stage and TPR signature were included in a nomogram model established for predicting OS in clinical settings (Figure 12C). Calibration and C-index curves were used to assess the agreement between the actual prognosis value and the value predicted by the nomogram. The calibration curves for the 1-, 3-, and 5-year survival rates exhibited a close fit with the nomogram values (Figure 12D). According to the C-index curves, in terms of predictive ability, the nomogram and clinical data performed in the following order: nomogram > TPR signature > T/N stage > M stage (Figure 12E).

Validation of the expression of T cell proliferation signature in lung adenocarcinoma cell lines

To explore the clinical significance of the TPR signature, mRNA expression were validated in LUAD cell lines (PC9 and HCC827) and a normal lung cell line (HBE). As shown in Figure 13, the qPCR results indicated that the mRNA expression of *CDK1*, *FADD*, and *LBTR* were

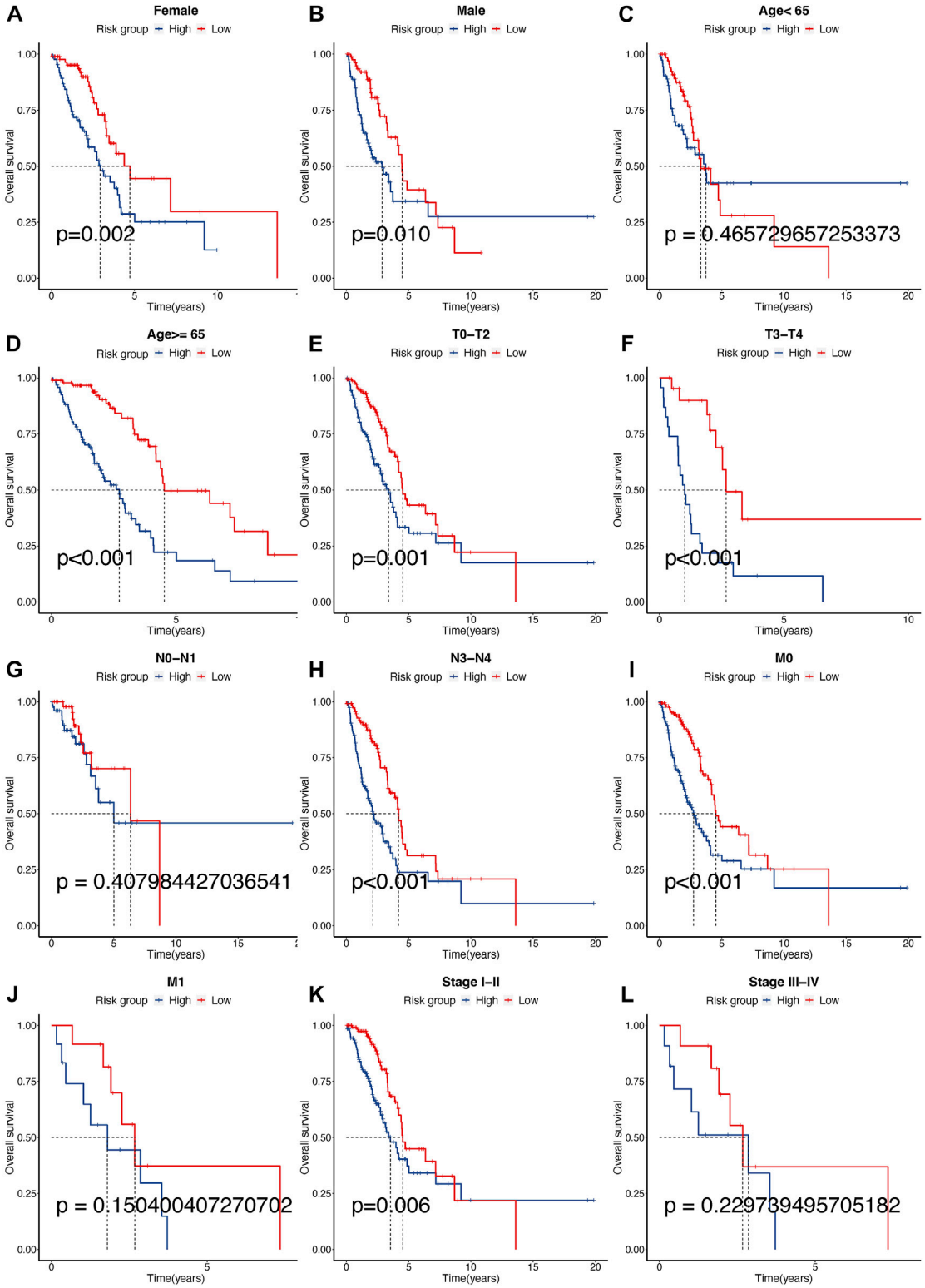


FIGURE 10
Different clinical sub-group survival analyses between high- and low-risk groups in the TCGA-LUAD cohort. **(A–L)** KM analysis between gender (Female: a, Male: b), age (<65: c, ≥65: d), T (TX-T2: e, T3-T4: f), N (N0-N1: g, N2-N3: h), M (M0: i, M1: j), and stage (I-II: k, III-IV: l) for high- and low-risk scores in TCGA-LUAD cohort.

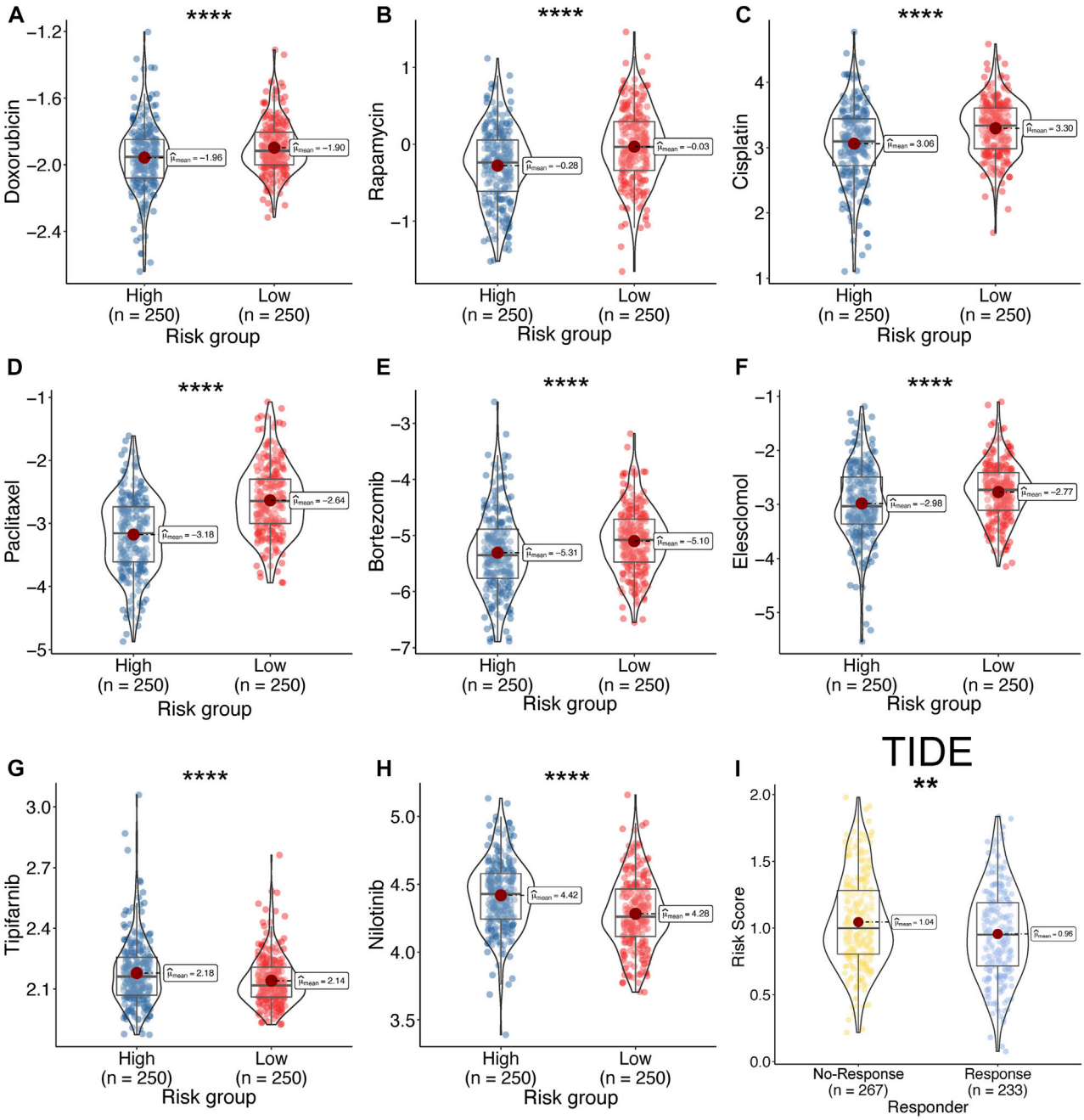


FIGURE 11
Drug sensitivity between high- and low-risk. **(A–H)** Calculate the half-maximal inhibitory concentration of FDA-approved drugs in the risk model, including, Doxorubicin **(A)**, Rapamycin **(B)**, Cisplatin **(C)**, Paclitaxel **(D)**, Bortezomib **(E)**, Elesclomol **(F)**, Tipifarnib **(G)**, and Nilotinib **(H)**. **(I)** Response rate between high- and low-risk group by TIDE algorithm in TCGA-LUAD cohort.

significantly increased in LUAD cell lines compared with those in the normal lung cell line, whereas the expression of *CADM*, *CRTAM*, *FYN*, *AGER*, and *CYP27A1* exhibited the opposite trend. No statistical difference was observed in the expression of *ADA* between LUAD cells and normal lung cells.

Discussion

LUAD is characterized by multiple mutations and copy number alterations (Chen et al., 2020), posing challenges to establishing individualized immunotherapy. However, advances in immunotherapies that target the components of the TME have

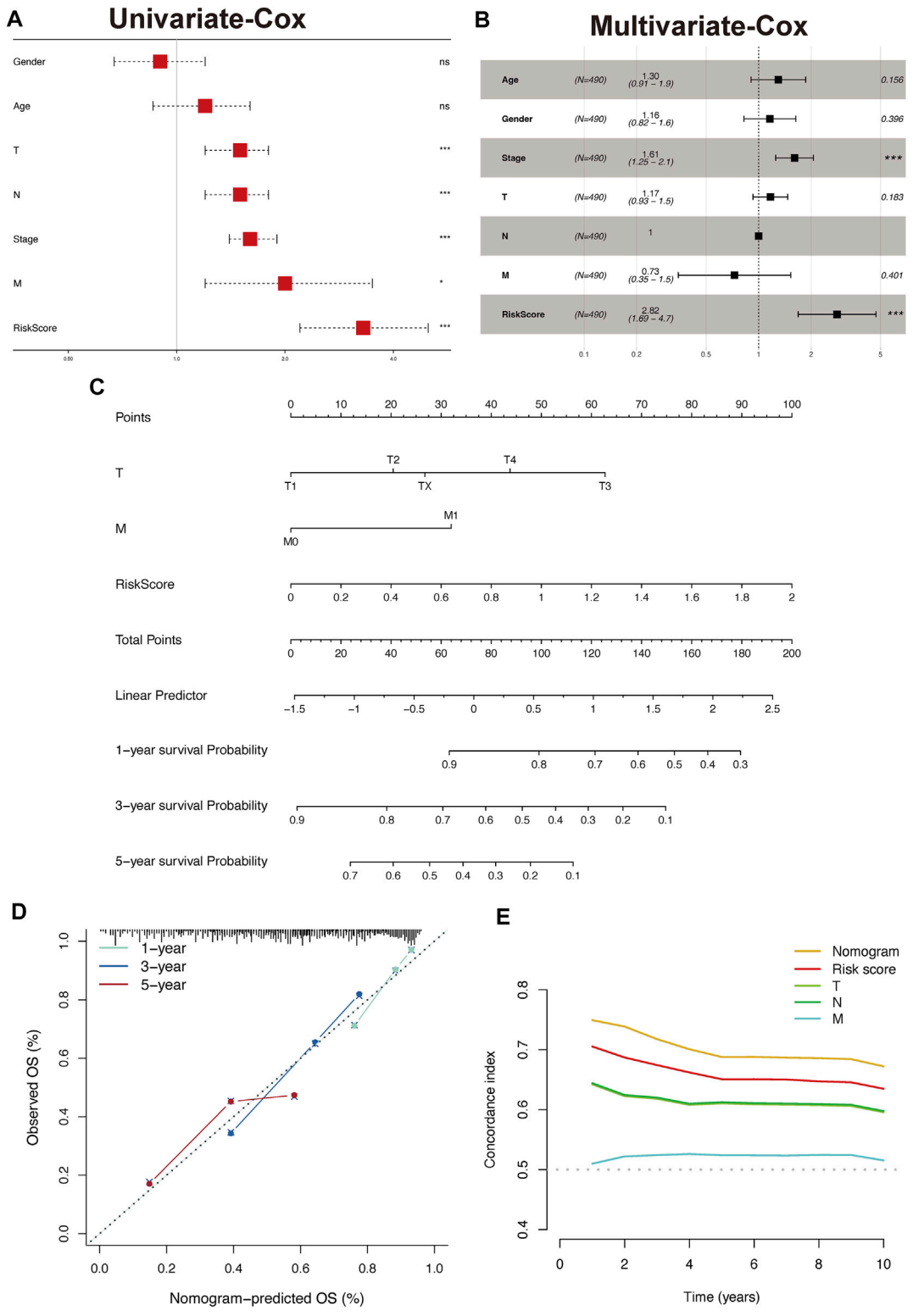


FIGURE 12
Construct and validate the nomogram model. **(A,B)** Univariate **(A)** and multivariate **(B)** analyses were performed using Cox regression of the TCGA-LUAD cohort. **(C)** Nomogram based on TPR signature and Stage in TCGA-LUAD. **(D)** Calibration of a nomogram predicting 1-, 3-, and 5-year OS. **(E)** Distributions of concordance index values in a nomogram and relevant clinical data.

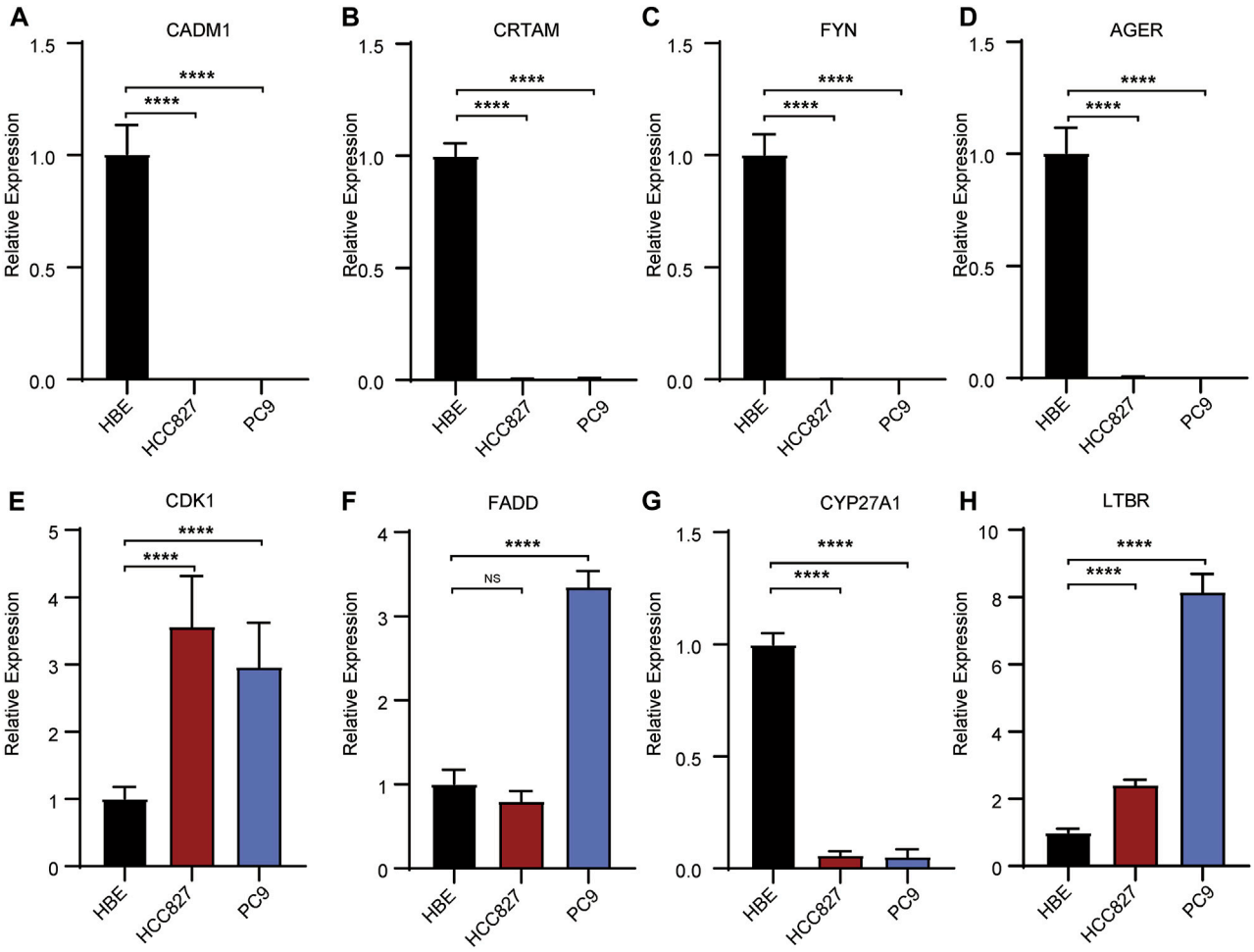


FIGURE 13
Validation of the expression of risk signatures in lung cancer cell lines (PC9 and HCC827) and normal lung cell (HBE) by RT-PCR analysis. **(A–D,G)** The expression of CADM1, CRTAM, FYN, AGER **(A–D)** and CYP27A1 **(H)** were decreased in normal lung cell lines. **(E, F, H)** The expression of CDK1, FADD **(E, F)** and LTBR **(H)** were unregulated in normal lung cell lines. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.001$.

exhibited variable efficacy in the treatment of lung cancers, including LUAD (Kwak et al., 2018; O'Donnell et al., 2019). T cell functions are usually inhibited in cancers because of transcriptional and translational modifications introduced by other cell populations in the TME. The cytokines, chemokines, and nutrients in the TME enable cancer cells to escape from antitumor T cells (Speiser et al., 2016). Thus, secondary immunosuppression contributes to multiple biological processes involved in tumor progression and initiation. Recently, numerous positive TPRs have been identified *via* genome-scale screening, providing new insights into T cell therapy (Legut et al., 2022). TPRs are optimal targets for immunotherapy and may be closely related to the prognosis of LUAD patients.

In this study, nine TPRs in patients with LUAD were identified, using LASSO-Cox regression analysis. Most of these TPRs are differentially expressed in LUAD and are correlated with prognosis. For example, *FADD* overexpression affects NF-κB activity and cell cycle progression and is correlated with poor clinical outcomes in LUAD (Chen et al., 2005). *CADM1* is downregulated by miR-423-5p in LUAD tissues and cell lines, contributing to proliferation and metastasis (Huang and Feng, 2021). *CYP27A1* downregulation enhances the effect of cholesterol on LUAD cell proliferation and invasion and reduces high cholesterol-induced LUAD metastasis *in vivo* (Li et al., 2022). *FYN* expression in LUAD correlates with a poor prognosis and is downregulated in LUAD tissues and cell lines (Xue et al., 2020). *CDK1* upregulation correlates with poor prognosis, poor survival until first progression, and poor post-progression survival in patients with LUAD (Li et al., 2020). Downregulation of *AGER* (Zhang et al., 2018) and upregulation of *LTBR* (Zhang et al., 2019) are also correlated with the prognosis of LUAD, as demonstrated by multiple bioinformatics analyses. Our RT-PCR results were consistent with the documented downregulation of CRTAM. However, the role of ADA in LUAD has not yet been elucidated, and further studies are needed.

Three distinct TPC subgroups were identified by unsupervised consensus clustering. These three clusters exhibited different TME immune cell infiltration levels, biological pathway enrichment, and drug sensitivities. Cluster 1 was characterized by moderate T cell immune infiltration and activation of carcinogenic pathways, while cluster 2 was characterized by robust T cell immune filtration and the enrichment of pathways associated with carcinogenic gene sets and tumor immunity. Furthermore, cluster 2 also exhibited a robust positive correlation with immune activation-related biological processes, including cytokine-cytokine receptor interaction, interferon-gamma response, and inflammatory response. Additionally, cluster 2 had the highest expression scores for four immune checkpoint genes (PDCD1, PDCD1LG2, CTLA4, and LAG3) involved in immune blockage (Masugi et al., 2017; Solinas et al., 2019; Ali et al., 2020). The cluster 2 also exhibited moderate T cell exhaust and higher T cell activation score calculated by ssGSEA algorithm (Figures 4J,K). Therefore, this cluster may be the most suitable for immunotherapy. However, cluster 2 did not show a matching survival advantage (Figure 2). Although cluster 2 is a T-cell activated state ($CD8^+$ T cell), the higher stromal cell activation indicates that high levels of CD4(+) T-cell-mediated, Treg infiltration and dendritic cell (poor antigen presentation capacity) may be induced immunosuppressive effect (Supplementary Figure S1) (Wang et al., 2012).

These results also indicate that cluster 2 can be classified as an immune-inflamed phenotype. Cluster 1 represents the intermediate phenotype, and the prognostic analysis results indicated that this cluster was associated with the best prognosis of the three. We attributed this result to the optimal localization and migration of T cells, which is essential for immune surveillance and the inhibition of tumor initiation (Smyth et al., 2016). In contrast, cluster 3 was negatively associated with the pathways linked to cluster 2 and featured low levels of immune cell infiltration, which is associated with immune tolerance and ignorance. Our analyses indicated that the dense stromal status in cluster 3 might influence the migration and activation of T cells, resulting in an immune-desert phenotype (Kim and Chen, 2016). To confirm the above findings, we performed a validation study in an independent TCGA-LUAD cohort. By analyzing TME immune cell infiltration and conducting enrichment analyses, we demonstrated the reliability of the TPCs for the identification and classification of immune phenotypes.

We have shown that TPRs are crucial mediators of multiple T cell functions and adaptive immune responses. However, we were unable to apply TPC analysis to individuals, as TPCs are a population-based tool. To account for the individual heterogeneity and complexity, a TPR risk model was established as a scoring system to evaluate and quantify the TPR patterns of individual LUAD patients. The low-risk group was classified as the enhanced T cell infiltration phenotype. This group was enriched in immune-related signaling pathways and was associated with a better prognosis. Conversely, the high-risk group was classified as the immune-excluded phenotype and was enriched in stromal cell-associated pathways, which restrict T cell entry into tumor islets by inhibiting their migration and penetration (Salmon et al., 2012). In addition, cluster 2, characterized by an immune-inflamed phenotype, was associated with a higher risk and a poor prognosis, whereas cluster 1 exhibited a lower risk and a better prognosis. These results demonstrated the feasibility and reliability of the risk model for assessing TPR patterns and prognosis in individuals with LUAD.

Checkpoint blockade therapy has shown surprising efficacy in the treatment of multiple cancers, especially in patients with an immune-inflamed TME (Kim and Chen, 2016; Cao et al., 2021). However, immune escape remains a major obstacle to achieving an extended OS in patients with solid tumors, including LUAD (Anichini et al., 2020). Many factors contribute to immune escape in LUAD, such as impaired antigen presentation, loss of heterozygosity in the human leukocyte antigen region, neoantigen silencing, and activation of immune checkpoints (Gajewski et al., 2013; Anichini et al., 2020). We successfully employed TPC analysis to distinguish among the immune phenotypes of the LUAD patients. We hypothesized that TPCs are associated with TMB and that TPCs can be used to predict the clinical response to checkpoint blockade immunotherapies. Consistent with previous reports, patients in cluster 2 with a high TMB (>10 Mb) and low TIDE score had a better response to PD-1/PD-L1 blockade therapy (Figures 4G–J) (Chan et al., 2019). In addition, the low-risk group was more susceptible to ipifarnib and nilotinib, both of which inhibit PD-1/PD-L1 directly or indirectly (Jackson et al., 1986; Tracy et al., 2022). Altogether, our results confirmed that TPCs are a valuable tool for predicting drug sensitivity and immunotherapy responses in patients with LUAD.

In summary, the TPR-related risk model exhibited reliability when used to evaluate the mutation features, degree of immune infiltration, and clinicopathological characteristics of individuals with LUAD. Moreover, the risk score served as a prognostic factor for predicting the prognosis of patients with LUAD and as a predictive factor for drug sensitivity. By developing a TPR-related risk model, our study provides novel insights into immunotherapy strategies.

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

YaL designed the study, interpreted the results, and revised the manuscript. GP and CQ analyzed the data and participated in

bioinformatics analysis. XW and YuL acquired the data and revised the manuscript. YuL drafted the manuscript and supervised the study. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Hunan Provincial Natural Science Committee Project (No. 2020JJ5859).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1003754/full#supplementary-material

SUPPLEMENTARY FIGURE S1
The immune infiltration of TPS subtype.

References

Ali, M. A., Abdelaziz, A., Ali, M., Abonar, A., Hanafy, M., Hussein, H., et al. (2020). PADI4 (rs2240340), PDCD1 (rs10204525), and CTLA4 (231775) gene polymorphisms and polyarticular juvenile idiopathic arthritis. *Br. J. Biomed. Sci.* 77 (3), 123–128. doi:10.1080/09674845.2020.1730626

Altorki, N. K., Markowitz, G. J., Gao, D., Port, J. L., Saxena, A., Stiles, B., et al. (2019). The lung microenvironment: An important regulator of tumour growth and metastasis. *Nat. Rev. Cancer* 19 (1), 9–31. doi:10.1038/s41568-018-0081-9

Anderson, N. M., and Simon, M. C. (2020). The tumor microenvironment. *Curr. Biol.* 30 (16), R921–R925. doi:10.1016/j.cub.2020.06.081

Anichini, A., Perotti, V. E., Sgambelluri, F., and Mortarini, R. (2020). Immune escape mechanisms in non small cell lung cancer. *Cancers (Basel)* 12 (12), E3605. doi:10.3390/cancers12123605

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1

Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D. R., Steins, M., Ready, N. E., et al. (2015). Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N. Engl. J. Med.* 373 (17), 1627–1639. doi:10.1056/NEJMoa1507643

Brustugun, O. T., Gronberg, B. H., Fjellbirkeland, L., Helbekkmo, N., Aanerud, M., Grimsrud, T. K., et al. (2018). Substantial nation-wide improvement in lung cancer relative survival in Norway from 2000 to 2016. *Lung Cancer* 122, 138–145. doi:10.1016/j.lungcan.2018.06.003

Cao, R., Ma, B., Wang, G., Xiong, Y., Tian, Y., and Yuan, L. (2021). Characterization of hypoxia response patterns identified prognosis and immunotherapy response in bladder cancer. *Mol. Ther. Oncolytics* 22, 277–293. doi:10.1016/j.omto.2021.06.011

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol.* 30 (1), 44–56. doi:10.1093/annonc/mdy495

Chen, G., Bhojani, M. S., Heaford, A. C., Chang, D. C., Laxman, B., Thomas, D. G., et al. (2005). Phosphorylated FADD induces NF-kappaB, perturbs cell cycle, and is associated with poor outcome in lung adenocarcinomas. *Proc. Natl. Acad. Sci. U. S. A.* 102 (35), 12507–12512. doi:10.1073/pnas.0500397102

Chen, J., Yang, H., Teo, A. S. M., Amer, L. B., Sherbaf, F. G., Tan, C. Q., et al. (2020). Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* 52 (2), 177–186. doi:10.1038/s41588-019-0569-6

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

Gajewski, T. F., Woo, S. R., Zha, Y., Spaapen, R., Zheng, Y., Corrales, L., et al. (2013). Cancer immunotherapy strategies based on overcoming barriers within the tumor microenvironment. *Curr. Opin. Immunol.* 25 (2), 268–276. doi:10.1016/j.coi.2013.02.009

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20 (3), 307–315. doi:10.1093/bioinformatics/btg405

Gettinger, S., Rizvi, N. A., Chow, L. Q., Borghaei, H., Brahmer, J., Ready, N., et al. (2016). Nivolumab monotherapy for first-line treatment of advanced non-small-cell lung cancer. *J. Clin. Oncol.* 34 (25), 2980–2987. doi:10.1200/JCO.2016.66.9929

Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., et al. (2017). Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* 16 (11), 2598–2608. doi:10.1158/1535-7163.MCT-17-0386

Hanahan, D., and Coussens, L. M. (2012). Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell* 21 (3), 309–322. doi:10.1016/j.ccr.2012.02.022

Harrison, P. T., Vyse, S., and Huang, P. H. (2020). Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. *Semin. Cancer Biol.* 61, 167–179. doi:10.1016/j.semcancer.2019.09.015

Hellmann, M. D., Ciuleanu, T. E., Pluzanski, A., Lee, J. S., Otterson, G. A., Audigier-Valette, C., et al. (2018). Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N. Engl. J. Med.* 378 (22), 2093–2104. doi:10.1056/NEJMoa1801946

Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553 (7689), 446–454. doi:10.1038/nature25183

Huang, Y., and Feng, G. (2021). MiR-423-5p aggravates lung adenocarcinoma via targeting CADM1. *Thorac. Cancer* 12 (2), 210–217. doi:10.1111/1759-7714.13745

Jackson, M. J., Seder, R. H., and Katz, A. E. (1986). Oral complications of experimental plasma exchange in head and neck cancer. *J. Prosthet. Dent.* 55 (6), 718–722. doi:10.1016/0022-3913(86)90449-x

Kang, S., Seo, S. S., Chang, H. J., Yoo, C. W., Park, S. Y., and Dong, S. M. (2008). Mutual exclusiveness between PIK3CA and KRAS mutations in endometrial carcinoma. *Int. J. Gynecol. Cancer* 18 (6), 1339–1343. doi:10.1111/j.1525-1438.2007.01172.x

Kim, J. M., and Chen, D. S. (2016). Immune escape to PD-L1/PD-1 blockade: Seven steps to success (or failure). *Ann. Oncol.* 27 (8), 1492–1504. doi:10.1093/annonc/mdw217

Kim, Y. A., Madan, S., and Przytycka, T. M. (2017). WeSME: Uncovering mutual exclusivity of cancer drivers and beyond. *Bioinformatics* 33 (6), 814–821. doi:10.1093/bioinformatics/btw242

Kwak, J. W., Laskowski, J., Li, H. Y., McSharry, M. V., Sippel, T. R., Bullock, B. L., et al. (2018). Complement activation via a C3a receptor pathway alters CD4(+) T lymphocytes and mediates lung cancer progression. *Cancer Res.* 78 (1), 143–156. doi:10.1158/0008-5472.CAN-17-0240

Legut, M., Gajic, Z., Guarino, M., Daniloski, Z., Rahman, J. A., Xue, X., et al. (2022). A genome-scale screen for synthetic drivers of T cell proliferation. *Nature* 603 (7902), 728–735. doi:10.1038/s41586-022-04494-7

Li, M., He, F., Zhang, Z., Xiang, Z., and Hu, D. (2020). CDK1 serves as a potential prognostic biomarker and target for lung cancer. *J. Int. Med. Res.* 48 (2), 300060519897508. doi:10.1177/0300060519897508

Li, X., Chen, H., Zhang, L., Chen, L., Wei, W., Gao, S., et al. (2022). 27-hydroxycholesterol linked high cholesterol diet to lung adenocarcinoma metastasis. *Oncogene* 41 (19), 2685–2695. doi:10.1038/s41388-022-02285-y

Masugi, Y., Nishihara, R., Hamada, T., Song, M., da Silva, A., Kosumi, K., et al. (2017). Tumor PDCD1LG2 (PD-L2) expression and the lymphocytic reaction to colorectal cancer. *Cancer Immunol. Res.* 5 (11), 1046–1055. doi:10.1158/2326-6066.CIR-17-0122

Mayekar, M. K., and Bivona, T. G. (2017). Current landscape of targeted therapy in lung cancer. *Clin. Pharmacol. Ther.* 102 (5), 757–764. doi:10.1002/cpt.810

Munshi, N. C., Anderson, L. D., Jr., Shah, N., Madduri, D., Berdeja, J., Lonial, S., et al. (2021). Idecabtagene vicleucel in relapsed and refractory multiple myeloma. *N. Engl. J. Med.* 384 (8), 705–716. doi:10.1056/NEJMoa2024850

O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* 16 (3), 151–167. doi:10.1038/s41571-018-0142-8

Petitprez, F., Levy, S., Sun, C. M., Meylan, M., Linhard, C., Becht, E., et al. (2020). The murine Microenvironment Cell Population counter method to estimate abundance of tissue-infiltrating immune and stromal cell populations in murine samples using gene expression. *Genome Med.* 12 (1), 86. doi:10.1186/s13073-020-00783-w

Powers, R. K., Goodspeed, A., Pielke-Lombardo, H., Tan, A. C., and Costello, J. C. (2018). GSEA-InContext: Identifying novel and common patterns in expression experiments. *Bioinformatics* 34 (13), i555–i564. doi:10.1093/bioinformatics/bty271

Saito, M., Suzuki, H., Kono, K., Takenoshita, S., and Kohno, T. (2018). Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. *Surg. Today* 48 (1), 1–8. doi:10.1007/s00595-017-1497-7

Saka, H., Nishio, M., Hida, T., Nakagawa, K., Sakai, H., Nogami, N., et al. (2021). Five-year follow-up results from phase II studies of nivolumab in Japanese patients with previously treated advanced non-small cell lung cancer: Pooled analysis of the ONO-4538-05 and ONO-4538-06 studies. *Jpn. J. Clin. Oncol.* 51 (1), 106–113. doi:10.1093/jjco/hyaa157

Salmon, H., Franciszkiewicz, K., Damotte, D., Dieu-Nosjean, M. C., Validire, P., Trautmann, A., et al. (2012). Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J. Clin. Invest.* 122 (3), 899–910. doi:10.1172/JCI45817

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2017. *Ca. Cancer J. Clin.* 71 (1), 7–30. doi:10.3322/caac.21387

Smyth, M. J., Ngiow, S. F., Ribas, A., and Teng, M. W. (2016). Combination cancer immunotherapies tailored to the tumour microenvironment. *Nat. Rev. Clin. Oncol.* 13 (3), 143–158. doi:10.1038/nrclinonc.2015.209

Solinas, C., Migliori, E., De Silva, P., and Willard-Gallo, K. (2019). LAG3: The biological processes that motivate targeting this immune checkpoint molecule in human cancer. *Cancers (Basel)* 11 (8), E1213. doi:10.3390/cancers11081213

Speiser, D. E., Ho, P. C., and Verdeil, G. (2016). Regulatory circuits of T cell function in cancer. *Nat. Rev. Immunol.* 16 (10), 599–611. doi:10.1038/nri.2016.80

Tracy, S. I., Venkatesh, H., Hekim, C., Heltemes Harris, L. M., Knutson, T. P., Bachanova, V., et al. (2022). Combining nilotinib and PD-L1 blockade reverses CD4+ T-cell dysfunction and prevents relapse in acute B-cell leukemia. *Blood* 140, 335–348. doi:10.1182/blood.2021015341

Travis, W. D. (2011). Pathology of lung cancer. *Clin. Chest Med.* 32 (4), 669–692. doi:10.1016/j.ccm.2011.08.005

Waldman, A. D., Fritz, J. M., and Lenardo, M. J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* 20 (11), 651–668. doi:10.1038/s41577-020-0306-5

Wang, W., Hodkinson, P., McLaren, F., MacKinnon, A., Wallace, W., Howie, S., et al. (2012). Small cell lung cancer tumour cells induce regulatory T lymphocytes, and patient survival correlates negatively with FOXP3+ cells in tumour infiltrate. *Int. J. Cancer* 131 (6), E928–E937. doi:10.1002/ijc.27613

Weiden, P. L., Flournoy, N., Thomas, E. D., Prentice, R., Fefer, A., Buckner, C. D., et al. (1979). Antileukemic effect of graft-versus-host disease in human recipients of allogeneic-marrow grafts. *N. Engl. J. Med.* 300 (19), 1068–1073. doi:10.1056/NEJM197905103001902

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Xue, F., Jia, Y., and Zhao, J. (2020). Overexpression of FYN suppresses the epithelial-to-mesenchymal transition through down-regulating PI3K/AKT pathway in lung adenocarcinoma. *Surg. Oncol.* 33, 108–117. doi:10.1016/j.suronc.2020.02.002

Yan, T. D., Black, D., Bannon, P. G., and McCaughan, B. C. (2009). Systematic review and meta-analysis of randomized and nonrandomized trials on safety and efficacy of video-assisted thoracic surgery lobectomy for early-stage non-small-cell lung cancer. *J. Clin. Oncol.* 27 (15), 2553–2562. doi:10.1200/JCO.2008.18.2733

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

Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., et al. (2021). Iobr: Multi-Omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front. Immunol.* 12, 687975. doi:10.3389/fimmu.2021.687975

Zhang, M., Zhu, K., Pu, H., Wang, Z., Zhao, H., Zhang, J., et al. (2019). An immune-related signature predicts survival in patients with lung adenocarcinoma. *Front. Oncol.* 9, 1314. doi:10.3389/fonc.2019.01314

Zhang, W., Fan, J., Chen, Q., Lei, C., Qiao, B., and Liu, Q. (2018). SPP1 and AGER as potential prognostic biomarkers for lung adenocarcinoma. *Oncol. Lett.* 15 (5), 7028–7036. doi:10.3892/ol.2018.8235

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 05 December 2022
DOI 10.3389/fgene.2022.1058207

OPEN ACCESS

EDITED BY
Jian Li,
University Hospital Bonn, Germany

REVIEWED BY
Dingwei Ye,
Fudan University, China
Rajesh Kumar,
National Institutes of Health (NIH),
United States

*CORRESPONDENCE
Qi Chen,
chenqi@fjnu.edu.cn

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

RECEIVED 30 September 2022
ACCEPTED 21 November 2022
PUBLISHED 05 December 2022

CITATION
Zhao G, Zheng J, Tang K and Chen Q (2022), EMILIN2 is associated with prognosis and immunotherapy in clear cell renal cell carcinoma.
Front. Genet. 13:1058207.
doi: 10.3389/fgene.2022.1058207

EMILIN2 is associated with prognosis and immunotherapy in clear cell renal cell carcinoma

Guangjian Zhao, Jianpei Zheng, Kai Tang and Qi Chen*

Fujian Key Laboratory of Innate Immune Biology, Biomedical Research Center of South China, Fujian Normal University Qishan Campus, Fuzhou, China

Background: EMILIN2 is a platelet-associated elastin that regulates angiogenesis. It has recently been found to play an essential role in various tumors. Nevertheless, the mechanism of action of EMILIN2 in clear cell renal cell carcinoma (ccRCC) remains unclear.

Methods: Samples from 33 cancers were obtained from UCSC Xena and The Cancer Genome Atlas (TCGA) database. The relationship between EMILIN2 expression and the clinicopathological characteristics and immune infiltration of ccRCC was investigated. Nonnegative matrix factorization (NMF) was used to classify ccRCC patients. A multigene risk prediction model of ccRCC was constructed using LASSO regression and multivariate regression analysis. A nomogram survival probability prediction map and calibration curve were constructed based on clinical information.

Results: EMILIN2 is significantly overexpressed in ccRCC, a phenomenon that is associated with poor prognosis. Meanwhile, EMILIN2 expression is closely related to tumor immune infiltration in ccRCC. Patients with clear cell renal cell carcinoma were divided into two subtypes using NMF, with subtype 2 showed poor prognosis. Next, we established a risk score model for ccRCC based on the common differentially expressed genes (DEGs) between subtypes and groups based on EMILIN2 expression. The results indicated poor prognosis in the high-risk group in the training set and were confirmed in the validation set.

Conclusion: Our findings suggest that EMILIN2 expression is closely associated with immune infiltration in ccRCC. EMILIN2 expression is negatively correlated with the prognosis of ccRCC patients. Here, we developed a tool that could predict the prognosis of ccRCC patients.

KEYWORDS

EMILIN2, clear cell renal cell carcinoma (ccRCC), prognosis, tumor microenvironment, nonnegative matrix factorization (NMF)

Introduction

Clear cell renal cell carcinoma is the most common form of renal cancer. The occurrence of ccRCC is associated with multiple factors, such as smoking, drugs, obesity and others (Koul et al., 2011). Patients with clear cell renal cell carcinoma have a high probability of tumor metastasis. At present, there are two main treatment methods for ccRCC: surgical treatment and immunotherapy. If the disease is confined to the kidney in the early stage, surgical resection may be considered (Petejova and Martinek, 2016). If the tumor exhibits metastases, whether local or systemic, the survival rate is dramatically reduced (Siegel et al., 2019). However, the form of cytotoxic chemotherapy currently used is not very effective for the treatment of ccRCC. Molecular targeted drugs such as tyrosine kinase inhibitors (TKIs) and vascular endothelial growth factor receptor (VEGFR) inhibitors are often used to treat ccRCC patients because of the strong dependence of the tumor on angiogenesis (Vuong et al., 2019).

The degree of immune infiltration in ccRCC is also high among various cancers. Because the complex tumor microenvironment plays a critical role in the treatment of patients, the study of the tumor microenvironment is closely related to the identification of relevant factors and the development of drugs. Commonly used immune checkpoint inhibitors (ICI) that block the inhibitory receptor on PD-1/PD-L1 or CTLA-4 T cells have been shown to be very effective against this cancer type (Motzer et al., 2015; Motzer et al., 2018).

EMILIN2 is located on the short arm of human chromosome 18, encoding an extracellular matrix glycoprotein with a relative molecular mass of 116 kD and five protein domains: C-terminal C1q domain, proline-rich domain, collagenous domain, coiled-coil domain, and N-terminal cysteine-rich domain (EMI domain) (Bressan et al., 1993; Colombatti et al., 2000). EMILIN2 has been shown to bind to EMILIN1. Both are elastin microfiber interface proteins that play an important role in imparting elasticity to tissue and blood vessels (Colombatti et al., 2000). Several studies have shown that EMILIN2 is associated with anti-PD1 therapy in melanoma (Fejza et al., 2021). In gastric cancer, EMILIN2 regulates the proliferation of cancer cells through apoptosis (Andreuzzi et al., 2020). The loss of EMILIN2 expression leads to defective angiogenesis (Paulitti et al., 2018).

Although large body of evidence indicates that EMILIN2 is associated with tumor immunity and angiogenesis in various cancers, the role of EMILIN2 in ccRCC remains elusive. Investigation into the role of EMILIN2 in ccRCC may help improve our understanding of the occurrence, progression, and metastasis of ccRCC and aid the development of new therapeutic strategies. In this study, we identified EMILIN2 as an independent prognostic factor in ccRCC patients. High EMILIN2 expression was found to predict poor prognosis in ccRCC patients. EMILIN2 expression is strongly associated with cancer progression in ccRCC patients. Immune checkpoint, immune infiltration and enrichment analyses revealed that EMILIN2 expression was significantly associated with the immunity of ccRCC patients. And we developed a comprehensive prognostic model of ccRCC with good performance.

Methods and materials

Data access

We obtained the RNA-sequencing data and clinical data of 33 types of cancer patients from UCSC Xena (https://xenabrowser.net/datapages/) and TCGA database (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga), including age, gender, tumor stage and survival data. The gene expression dataset format is TPM. The data set of ccRCC contained 531 cancer tissue samples and 72 adjacent tissue samples. The immune, matrix and ESTIMATE scores of patients were calculated by ESTIMATE (Yoshihara et al., 2013) (https://bioinformatics.mdanderson.org/estimate/rpackage.html).

Determination of tumor-infiltrating immune cells in clear cell renal cell carcinoma

We calculated the respective scores of 27 immune cells in ccRCC patients by single sample gene set enrichment analysis. These cell types are activated $CD4^+$ T cell, activated $CD8^+$ T cell, activated dendritic cell, CD56 bright natural killer (NK) cell, central memory $CD4^+$ T cell, central memory $CD8^+$ T cell, NK cell, NK T cell, type 1 T helper cell, type 17 T helper cell, CD56 dim NK cell, immature dendritic cell, macrophage, myeloid derived suppressive cell (MDSC), neutrophil, plasmacytoid dendritic cell, regulatory T cell (Treg), type 2 T helper cell, activated B cell, eosinophil, gamma delta T cell, immature B cell, mast cell, memory B cell, monocyte and T follicular helper cell (Foroutan et al., 2018).

Gene set enrichment analysis

In this analysis, the clusterProfiler package was used to analyze the pathway changes between patients. Patients were divided into high and low groups according to the median EMILIN2 expression by limma, and the DEGs were obtained (Ritchie et al., 2015). Then the clusterProfiler package was used to perform KEGG and GO enrichment analysis for these DEGs (Wu et al., 2021).

Molecular subtyping was performed using a non-negative matrix factorization (NMF) algorithm

We performed Cox regression analysis using the CancerSubtype package to identify the gene sets of 27 immune cell types obtained from the TISIDB (Xu et al., 2017) (http://cis.hku.hk/TISIDB/data/download/CellReports.txt). Following this, the NMF clustering algorithm was used to cluster ccRCC samples based on the screened genes (Brunet et al., 2004). The NMF package was used for NMF analysis. In the NMF method, the standard "Brunet" option was selected, and data were subjected to 50 iterations. The number of clusters "k" ranged from 2 to 10. Finally, the optimal number of clusters was found to be 2. The DEGs between molecular subtypes were identified using the limma package.

Development and validation of the risk scoring model

After the common DEGs between the molecular subtypes of ccRCC and between high- and low-EMILIN2 expression groups were identified, ccRCC samples were divided into the training and validation sets in an 8:2 ratio. Following this, we calculated the risk score by LASSO regression and multivariate regression analysis based on "risk scores = $\sum$coef* Exp(genes)." Based on the findings, a six-gene prognostic model was established. We then used the ROC curve to assess the ability of this feature to predict patient survival in the high-risk and low-risk groups.

We selected the immunohistochemical staining images of these model genes in normal tissues and ccRCC tissues from the Human Protein Atlas database (http://www.proteinatlas.org).

Construction of nomogram

To assess the predictive accuracy and importance of risk models and clinical characteristics, we analyzed the predictive relationships among age, sex, and TNM stage. And the importance of examining the RiskScore factors from one or more perspectives. Then the nomogram model including age, gender, TNM stage and risk score was constructed. Nomogram calibration plots were used to compare predicted survival events at 1, 3, and 5 years with actual observations.

Cell lines and quantitative polymerase chain reaction

The human ccRCC cell line (786-O) and human renal epithelial cell line (293T) used for reverse-transcription-quantitative polymerase chain reaction (RT-qPCR) in this study were purchased from Procell (Wuhan, China). Total RNA was isolated using Trizol (Takara, Japan). HiScript III RT SuperMix for qPCR (+gDNA Wiper) (Nanjing, China) was used for cDNA synthesis. qPCR was performed using the SYBR Green assay according to the manufacturer's protocol. Then the mRNA expressions of TNNT1, SAA1, IL20RB, COL22A1, B3GALT5, and C10orf99 were analyzed.

The following primers were used for RT-qPCR:TNNT1, forward:5′-TGATCCCGCCAAAGATCCC-3′; TNNT1, reverse: 5′- TCTTCCGCTGCTCGAAATGTA-3′; SAA1, forward:5′-GATCACCGATGCCAGAGAGA-3′; SAA1, reverse:5′- TTT GTATCCCTGCCCTGAGG-3′; IL20RB, forward:5′- GGCCAC TGTGCCATACAAC-3′; IL20RB, reverse:5′- TCTTTGGTG ATCTCCATCCCA -3′; COL22A1, forward:5′- CCTAGCGTT CGTGTAGAAGGA-3′; COL22A1, reverse:5′- CCCATCCGT ACATAGGAACTCT-3′; B3GALT5, forward:5′- AAGCTC CCAGATACAGACTGC-3′; B3GALT5, reverse:5′- TGGTCC ACCTCTTTCGTTTCC-3′; C10orf99, forward:5′- CCGGTC ACAGCTACAAATCC-3′; C10orf99, reverse:5′- TCAGGA GGCTAGGAAGGGAT-3′; GAPDH, forward: 5′-GGA GCGAGATCCCTCCAAAAT-3′; and GAPDH, reverse:5′-GGCTGTTGTCATACTTCTCATGG-3′.

Statistical analysis

All data analyses were based on R software (software version 4.0.2). We then used Student's *t*-test to calculate the significance of the difference between the two groups. *p*-value greater than 0.05 was considered significant.

Result

Elastin microfibril interfacer 2 is expressed at high levels in clear cell renal cell carcinoma and indicates a poor prognosis

To the effect of EMILIN2 on various cancer types, we analyzed the EMILIN2 expression in cancer tissues and adjacent tissues in 33 types of cancer using data from TCGA. EMILIN2 was found to be expressed differentially in various cancer types ($p < 0.05$). For example, EMILIN2 expression was significantly up-regulated in cholangiocarcinoma, colon adenocarcinoma, and ccRCC (Figure 1A). Concurrently, we analyzed the prognosis of EMILIN2 in these 33 cancers. EMILIN2 was found to have prognostic significance in five cancers. For example, high EMILIN2 expression led to poor prognosis in adrenocortical carcinoma, ccRCC, brain bower grade glioma, testicular germ cell tumors and uveal melanoma ($p < 0.05$) (Figure 1B). Interestingly, EMILIN2 was differentially expressed only in ccRCC ($p < 0.0001$) and

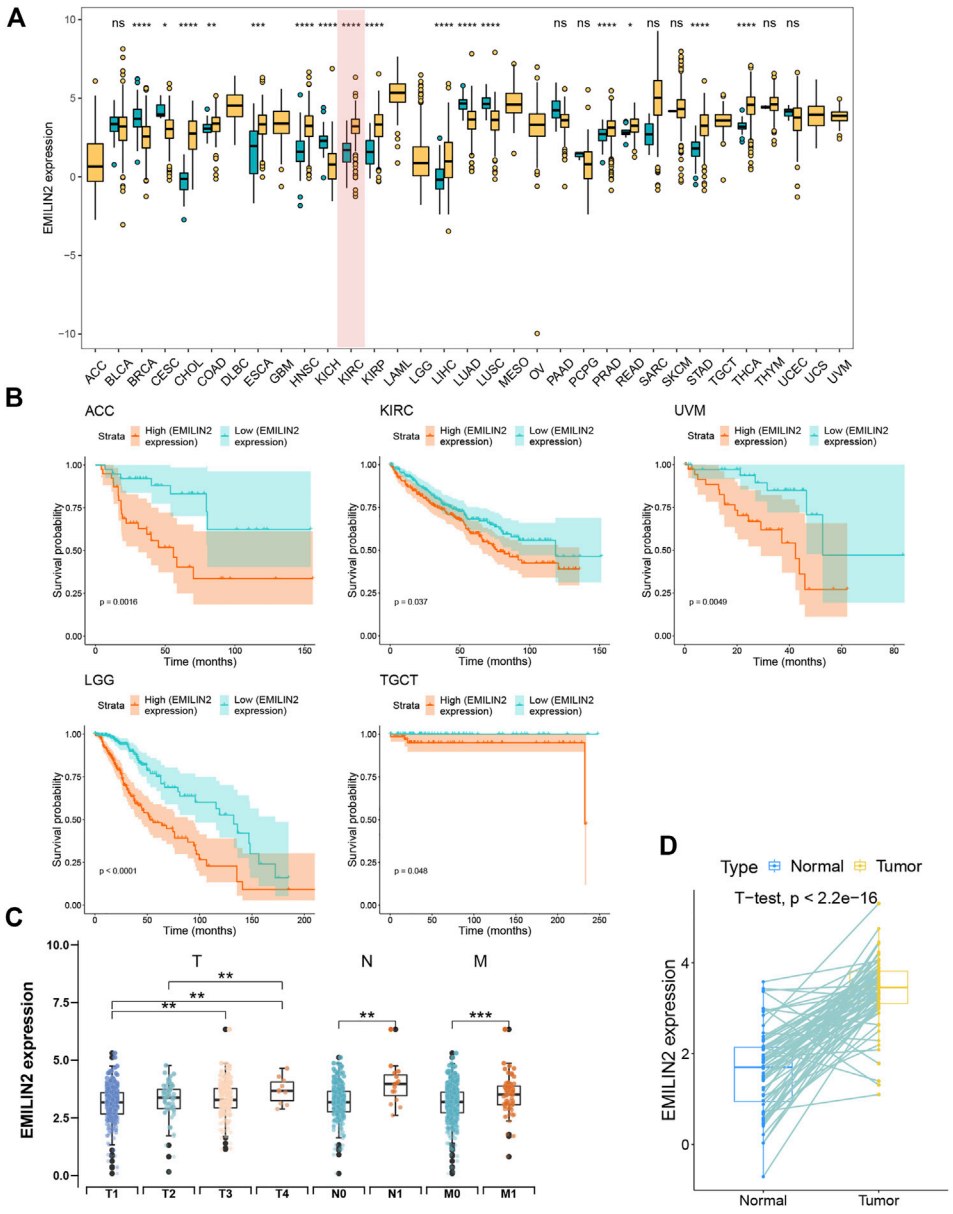


FIGURE 1
Pan-cancer analysis and expression analysis of EMILIN2 **(A)** EMILIN2 expression in different cancers, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns refers to not significant. **(B)** The survival map of EMILIN2 in different cancers. **(C)** EMILIN2 expression in different tumor stages, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns refers to not significant. **(D)** EMILIN2 expression in tumor tissues and its paired adjacent tissues, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns refers to not significant.

showed prognostic power. In order to study the relationship between ELIMIN2 expression and tumor stage, we analyzed the EMILIN2 expression in ccRCC patients using TNM stage information. The results showed that the EMILIN2 expression was significantly different in different tumor pathological stages. For example, EMILIN2 expression is significantly higher in patients with distant metastasis (n.s., not significant; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$) (Figure 1C). These results indicate that the EMILIN2 expression is closely related to tumor progression. In addition, we analyzed the EMILIN2 expression in paired ccRCC and adjacent normal tissues, and found that the EMILIN2 expression was significantly higher in tumor tissue samples (Figure 1D).

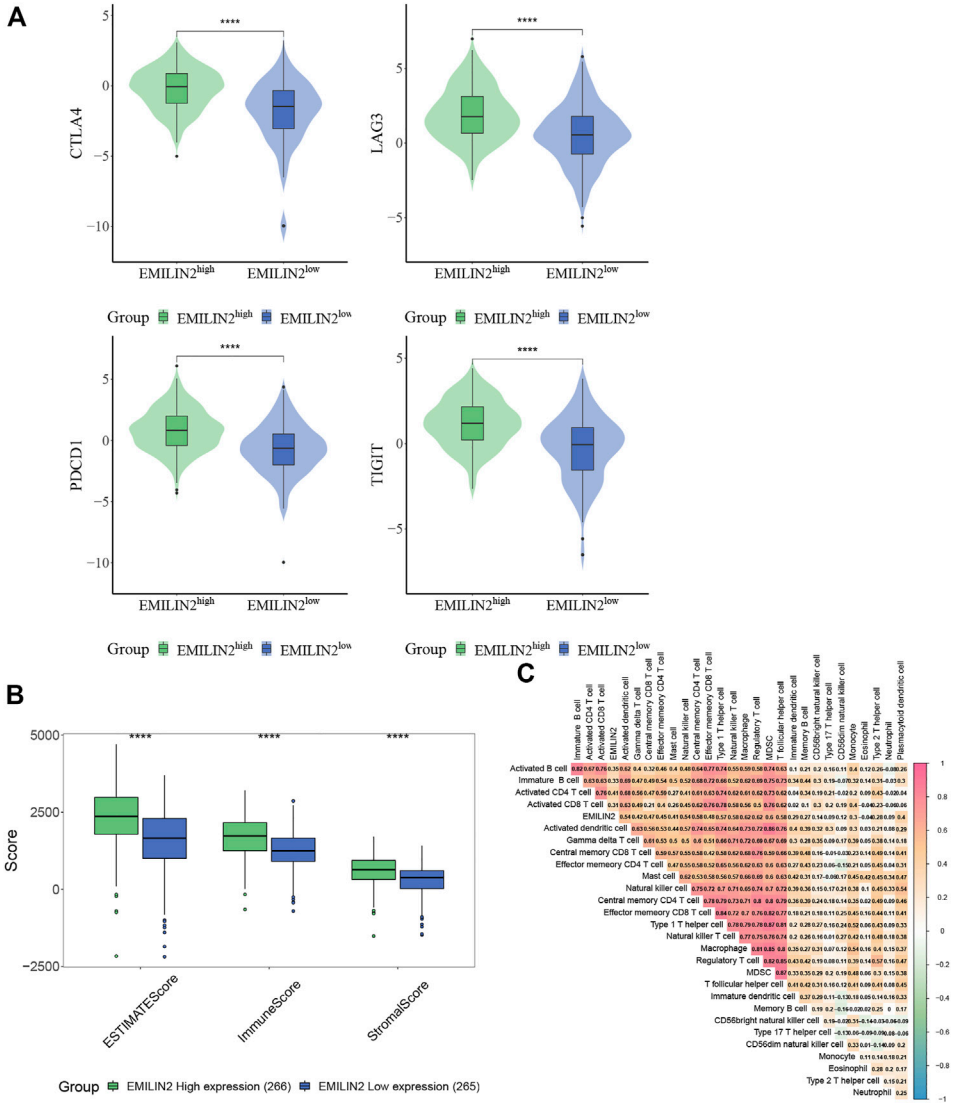


FIGURE 2
EMILIN2 is closely associated with immune infiltration in ccRCC **(A)** The EMILIN2 expression in high and low expression groups of different immune checkpoints, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns refers to not significant. **(B)** Immune, Stromal, and ESTIMATE Scores of EMILIN2 high and low expression groups, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns refers to not significant. **(C)** Correlation analysis of EMILIN2 and 27 kinds of immune cells.

Elastin microfibril interfacer 2 is associated with multiple immune checkpoints

The study of immune checkpoint is crucial for the treatment of cancer. To analyze the relationship between EMILIN2 and immune checkpoints, ccRCC patients were divided into high and low groups according to the median EMILIN2 expression. The results showed that several checkpoints including CTLA-2, PDCD1, LAG3, and TIGIT were significantly overexpressed in the EMILIN2 high expression group (Figure 2A). This indicates that the group with high EMILIN2 expression may have a better effect on immunotherapy.

Analysis of tumor immune infiltration with elastin microfibril interfacer 2

The tumor microenvironment is composed of immune cells and stromal cells in addition to tumor cells. The complexity of tumor microenvironment has a significant impact on tumor treatment. To analyze the relationship between ELIMIN2 and

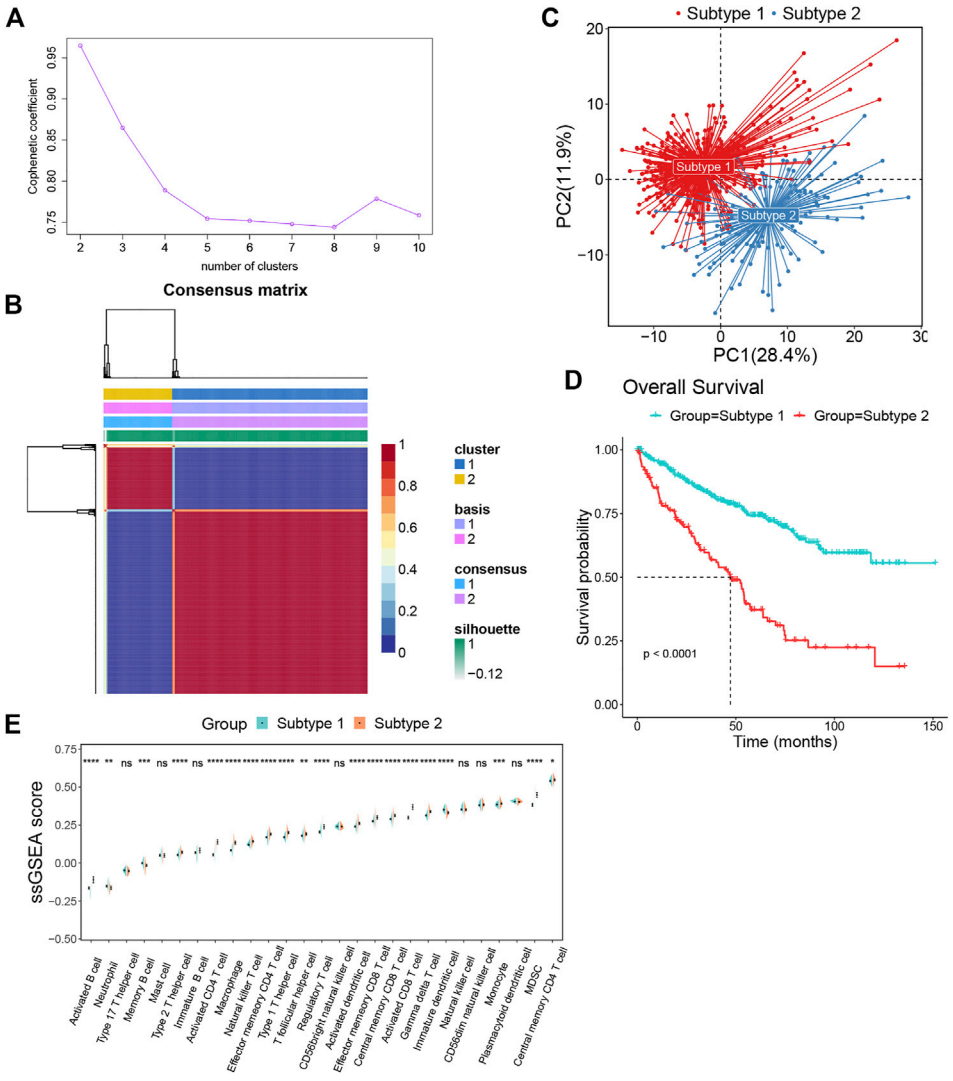


FIGURE 3
The ccRCC patients were divided into two subtypes by NMF algorithm **(A)** The best rank value was selected according to the Cophenetic value. **(B)** Cluster heatmap of two clusters (k = 2). **(C)** Principal component analysis of two subtypes. **(D)** The survival map of the two subtypes. **(E)**The differences in immune cell fractions among subtypes, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns refers to not significant.

tumor microenvironment, we calculated the immune score, stromal score, and ESTIMATE score of ccRCC patients by ESTIMATE. The results showed that the patients with high ELIMIN2 expression had immune score, stromal score, and ESTIMATE score ($p < 0.0001$) (Figure 2B). These evidences indicate that EMILIN2 is closely related to the tumor microenvironment.

To further analyze which immune cells were associated with EMILIN2, we calculated the scores of 27 types of immune cells in ccRCC patients by ssGSEA and found that EMILIN2 was significantly positively correlated with all 27 types of immune cells ($p < 0.05$) (Figure 2C). These evidences suggest that EMILIN2 is closely related to immune infiltration in ccRCC patients.

Patients with clear cell renal cell carcinoma can be categorized into two subtypes based on the findings of the nonnegative matrix factorization algorithm

We obtained the gene sets of 27 immune cell types (749 genes in total) that showed a significant positive correlation with EMILIN2. Next, we conducted multivariate regression analysis using the CancerSubtypes package for feature screening and identified 285 genes. Based on TCGA data, ccRCC patients were classified according to the genes expressed using the NMF package. The best value of K was 2 (Figure 3A). Finally,

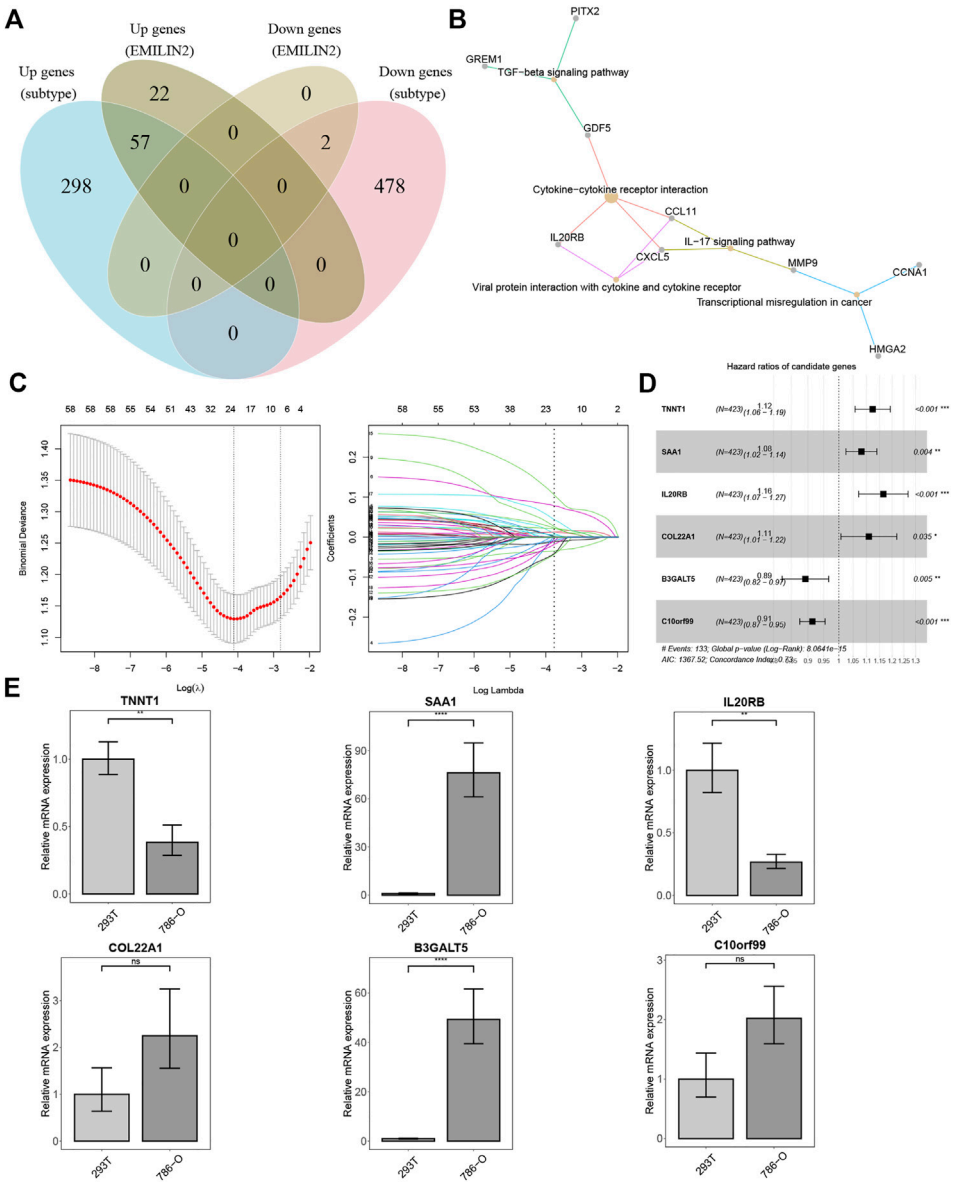


FIGURE 4
Establishment of risk score model and qPCR results **(A)** VEEN diagram of different DEGs. **(B)** KEGG enrichment analysis of 59 intersection genes. **(C)** The best λ value is obtained by ten-fold cross validation. **(D)** The forest map of risk model genes. **(E)** qPCR results of risk model genes, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns refers to not significant.

ccRCC patients were divided into two subtypes (Figure 3B). PCA was performed to evaluate the results of NMF typing, results indicated that the two subtypes showed significant differences (Figure 3C). Moreover, a significant difference was observed between in the survival of patients with the two subtypes of cancer. Patients with subtype 1 ccRCC had better overall survival, whereas patients with subtype 2 ccRCC had poor overall survival (Figure 3D).

To investigate the differences between subtypes 1 and 2, we analyzed immune cell infiltration in these two subtypes. Most immune cells, such as activated B cells, type 2 T helper cells, activated $CD4^+$ T cells, macrophages, natural killer T cells, effector memory $CD4^+$ T cells, type 1 T helper cells, t follicular helper cells, regulatory T cells, activated dendritic cells, effector memory $CD8^+$ T cells, central memory $CD8^+$ T cells, activated $CD8^+$ T cells, gamma delta T cells, monocytes, MDSCs and central memory $CD4^+$ T cells were more enriched in subtype 2 than subtype 1. Meanwhile, only a few immune cells were enriched in subtype 1, such as neutrophils, memory B cells and immature dendritic cells (Figure 3E).

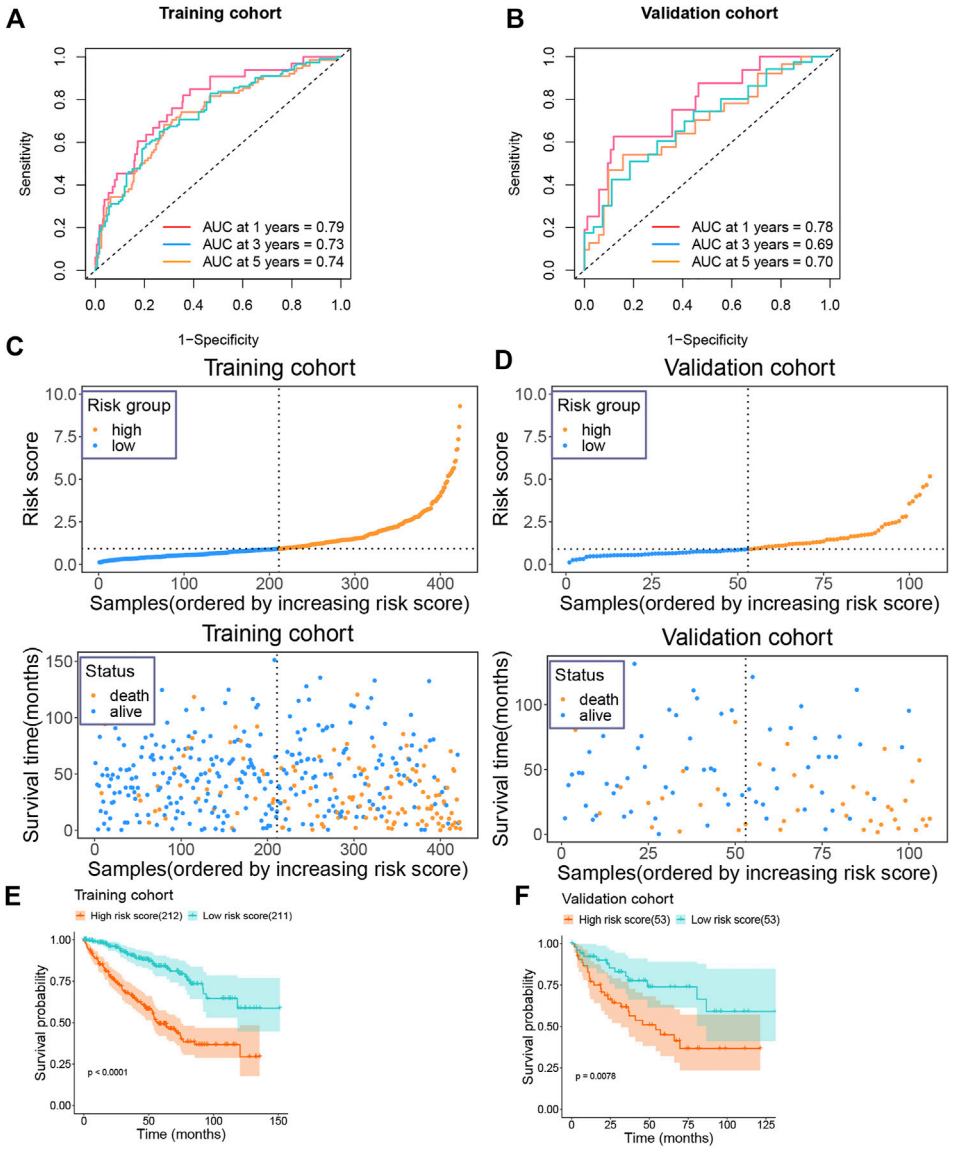


FIGURE 5
Prognostic value of risk model in training and validation groups **(A)** The AUC value of the risk model in the training group. **(B)** The AUC value of the risk model in the validation group. **(C)** Risk score and survival status of the patients in the training group. **(D)** Risk score and survival status of the patients in the validation group. **(E)** Survival plot of the high and low risk groups in the training group. **(F)** Survival plot of the high and low risk groups in the validation group.

Construction of a risk-score model for clear cell renal carcinoma

We analyzed the DEGs between the two subtypes and identified 355 upregulated and 478 downregulated genes. In addition, we identified the DEGs between patients in high and low expression groups according to the median EMILIN2 expression. Seventy-nine genes were upregulated and two genes were downregulated. Next, we evaluated the intersection of DEGs in the high and low EMILIN2 expression groups and the DEGs in the two subtypes of ccRCC to obtain a total of 59 genes (Figure 4A). These genes are associated with TGF-beta signaling pathway, cytokine-cytokine receptor interaction, viral protein interaction with cytokine and cytokine receptor, IL-17 signaling pathway and transcriptional misregulation in cancer (Figure 4B). Next, after excluding patient samples without survival data, 529 ccRCC patients were categorized in a training set and a validation set in an 8:2 ratio.

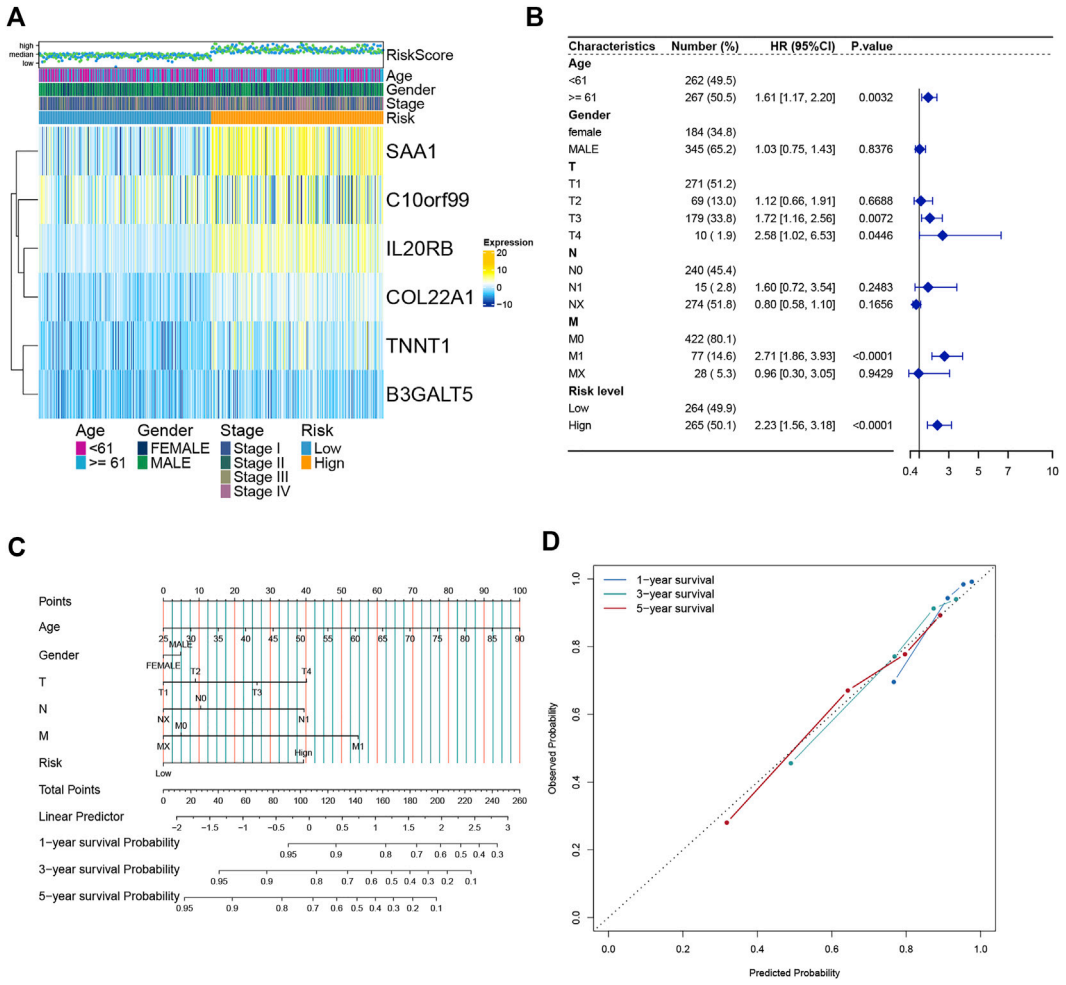


FIGURE 6
Establishment and verification of the Nomogram model **(A)** Heatmap of the expression of six model genes in all ccRCC patients. **(B)** Forest map of Risk score and clinical features. **(C)** Nomograms to predict 1 -, 3 -, and 5-year OS in ccRCC patients. **(D)** Calibration curves of the nomogram.

In the training set, we used LASSO regression to screen genes closely associated with patient prognosis among the 59 genes. The overfitting effect was overcome by ten-fold cross-validation, and the best λ value obtained was 0.016 (Figure 4C). The hazard ratios (HR) of TNNT1, SAA1, IL20RB, and COL22A1 were greater than 1, suggesting that these were risk genes. The HR values of B3GALT5 and C10orf99 were less than 1, suggesting these were protective genes (Figure 4D). We calculated the risk score for each ccRCC patient using the following formula:Risk score=(0.11597*TNNT1)+(0.07741*SAA1)+(0.15255*IL20RB)+(0.10302*COL22A1)+(−0.11426*B3GALT5)+(−0.09019*C10orf99). We used RT-qPCR to analyze the mRNA expression levels of the six genes in 786-O cells (derived from the renal epithelial cells of ccRCC patients) and 293T cells (human renal epithelial cells). TNNT1 and IL20RB had higher mRNA expression levels in 293T cells, SAA1 and B3GALT5 had higher mRNA expression levels in 786-O cells, and C10orf99 and COL22A1 showed no significant difference in mRNA expression levels in the 2 cell types (Figure 4E).

We analyzed the protein expression levels of the six model genes using the Human Protein Atlas database. TNNT1, COL22A1, and C10orf99 showed no significant difference in expression between cancer tissues and normal tissues, B3GALT5 expression was relatively high in cancer tissues, whereas SAA1 and IL20RB did not show protein expression (Supplementary Figures S1A–D).

Prognostic value of the risk score model

Following this, we calculated the ROC value of the model in the training set. The model showed good risk prediction potential

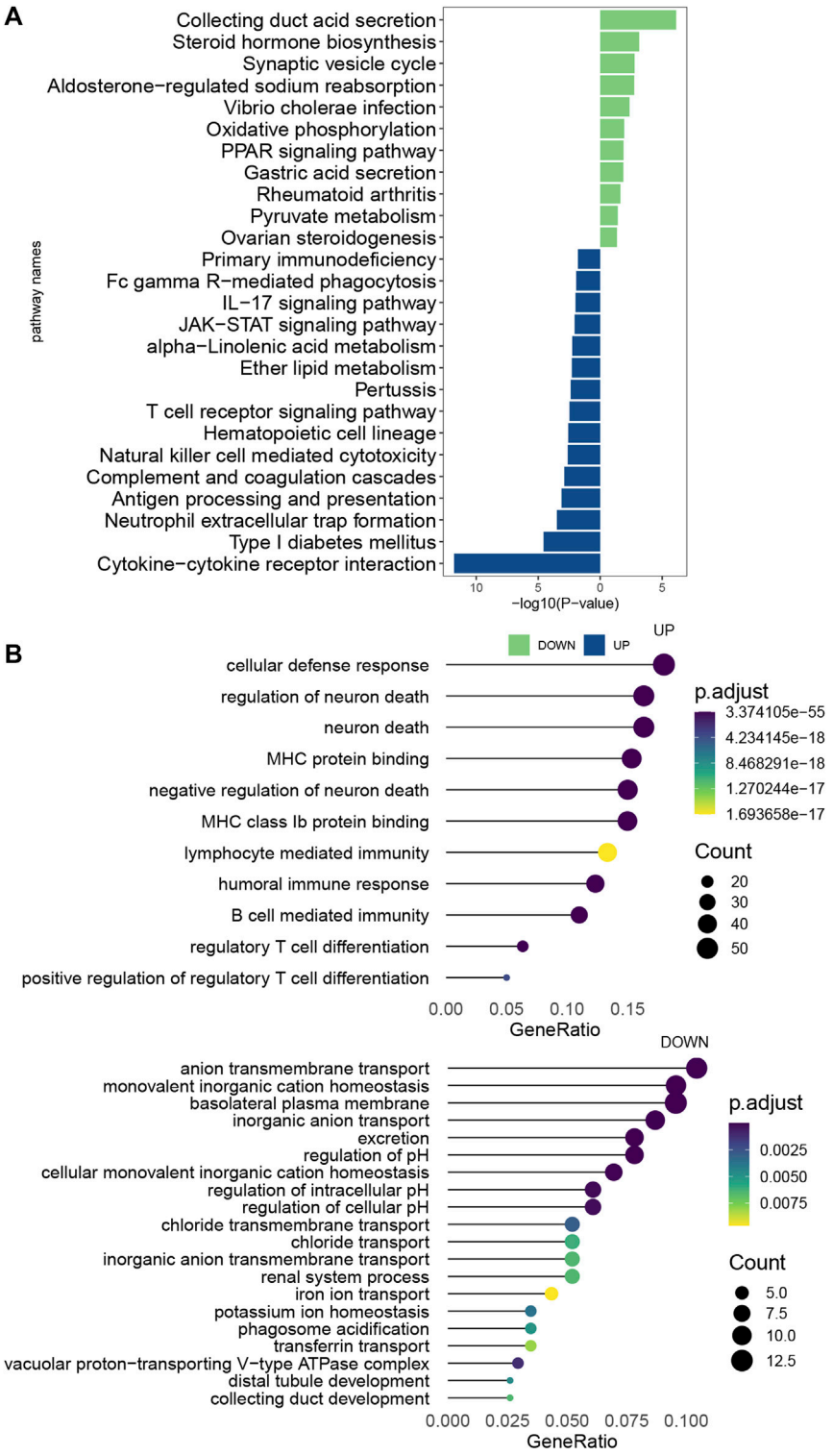


FIGURE 7
Pathway enrichment analysis of EMILIN2 **(A)** Differential pathways between KEGG-enriched EMILIN2 high and low expression groups. **(B)** Differential pathways between GO-enriched EMILIN2 high and low expression groups.

at 1, 3, and 5 years. The AUC value of the 1-year prediction was 0.79, that of the 3-year prediction was 0.73, and that of the 5-year prediction was 0.74 (Figure 5A). Similarly, in the validation set, the AUC of the model was 0.78 for the 1-year prediction, 0.69 for the 3-year prediction, and 0.7 for the 5-year prediction (Figure 5B).

We then categorized the patients in the training set into high-risk and low-risk groups according to their median risk score. As the risk score increased, the number of patients who died also increased (Figure 5C). Similarly, the data in the validation set showed consistently changes (Figure 5D). In addition, we also analyzed the survival of patients in the high-risk and low-risk groups. Patients in the high-risk group showed a very poor prognosis in the training set (Figure 5E). In the validation set, patients in the high-risk group also showed a poor prognosis (Figure 5F).

The nomogram was constructed and verified by combining clinical data

We analyzed the expression of six genes from the risk score model in all ccRCC patients. These six genes were found to be expressed differently in patients with different risk scores (Figure 6A). We then performed multivariate regression analysis by combining the risk scores of patients with clinical data to explore the association between the risk scores and clinical characteristics. Older age, advanced T stage, distant metastasis of the tumor, and a high risk score were identified as significant risk factors (Figure 6B). Previous analyses showed that the risk score value was significantly associated with patient survival. To develop a more accurate clinical prediction method for ccRCC patients, we constructed a comprehensive nomogram model combining the patient's risk score, gender, TNM stage, and age, which could be used to predict the 1-year, 3-year, and 5-year survival of patients (Figure 6C). Moreover, according to the results of the calibration map for patient survival prediction, the prediction results of the nomogram were in good agreement with the actual observation (Figure 6D). In conclusion, our analysis revealed that the constructed nomogram showed good performance in predicting the survival of ccRCC patients.

Pathway enrichment analysis

The ccRCC samples (cancer tissues and adjacent tumor samples) were divided into high and low expression groups according to the median expression of EMILIN2. Using the limma package, 366 DEGs were obtained, of which 235 genes were upregulated and 131 genes were downregulated. To explore the potential functions of these DEGs, we analyzed the biological pathways that these genes are associated with using the clusterProfiler package. In KEGG pathway analysis, cytokine-Cytokine_receptor_interaction, TCR signaling pathway, IL-17 signaling pathway, and natural killer cell-mediated cytotoxicity were significantly upregulated (Figure 7A). GO analysis revealed that these DEGs were significantly enriched in cellular defense response, regulatory T-cell differentiation, and B-cell-mediated immunity in biological processes. Moreover, these DEGs were significantly enriched in the basolateral plasma membrane in cellular components and significantly enriched in MHC class IB protein binding and MHC protein binding in molecular function (Figure 7B). This finding suggests that DEGs are associated with many important immune processes associated with T cells and B cells and may play an important role in tumor immunity in patients. The finding also indicates that EMILIN2 expression is closely associated with the immunity of patients. In addition, IL-17 signaling was found to affect angiogenesis, suggesting that EMILIN2 expression may be associated with angiogenesis in ccRCC.

Discussion

Clear cell renal cell carcinoma, the most common type of renal cancer, is a tumor highly dependent on vascular survival. At present, surgical treatment and immunotherapy are the major approaches used to treat patients with this cancer type (Siegel et al., 2019). Several immunotherapeutic drugs have been developed for treating ccRCC (Motzer et al., 2018) (Motzer et al., 2015). However, owing to the complexity of the tumor microenvironment, killing and removing tumor cells by immunotherapy is a complex process, and achieving good effectiveness is not feasible. Therefore, continuous exploration of the tumor microenvironment and the search for better immunotherapeutic agents remains a major challenge. Identifying new novel biomarkers that are closely related to patient immunity and survival is also of great significance.

In this study, bioinformatics analysis showed that EMILIN2, a gene that regulates angiogenesis, was significantly upregulated in ccRCC patients and had prognostic potential. Concurrently, EMILIN2 expression was found to be closely related to the immune status of ccRCC patients. EMILIN2 expression was found to be significantly correlated with multiple immune markers and immune cells. We then obtained the gene sets of these immune cells and performed the molecular typing of ccRCC patients using the NMF algorithm. Patients with clear cell renal cell carcinoma were grouped according to EMILIN2 expression, and the intersection of genes showing differential expression in different groups and subtypes was obtained. Based on this, a six-gene model was constructed and verified by LASSO regression and multivariate regression analysis. We then developed and validated a comprehensive nomogram model by combining risk scores and clinical data.

We first analyzed the EMILIN2 expression in 33 cancer types and its ability to predict patient survival. EMILIN2 showed a significant differential expression in ccRCC and indicated poor

prognosis. Thus, EMILIN2 expression was closely related to the prognosis of ccRCC. Following this, we investigated the association between EMILIN2 and immune checkpoints, and found that the group with high EMILIN2 expression showed the up-regulation of immune checkpoints. This implies that EMILIN2 expression has some potential effect on the immunotherapy of patients. We then calculated the immune, stromal and ESTIMATE scores of ccRCC patients using ESTIMATE, which is often used to study the tumor microenvironment and explore immune and stromal cells in the tumor microenvironment of different patients. Patients with high EMILIN2 expression also showed a significant increase in immune and stromal scores. This indicates that EMILIN2 expression affects the tumor microenvironment in ccRCC. The study of the complex tumor microenvironment is of great significance in the treatment of patients with tumor. These findings suggest that EMILIN2 may play an important role in tumor development, metastasis and immune response.

To explore the immune cells EMILIN2 might be associated with, we used ssGSEA to calculate the scores for 27 types of immune cells in ccRCC patients. Interestingly, EMILIN2 expression was positively correlated with the scores of 27 immune cell types, especially T cells. Thus, EMILIN2 expression affects the extent of infiltration of various immune cells in tumors. The tumor infiltration level of immune cells is an important index for prognostic judgment and evaluating therapeutic effects. Therefore, EMILIN2 has extremely important research value in the study of immune invasion of tumors. Next, we identified DEGs between the high and low EMILIN2 expression groups. KEGG pathway enrichment results revealed that these DEGs were significantly enriched in immune and vascular pathways, such as the TCR signaling pathway and IL-17 signaling pathway. Many studies have shown that the TCR signaling pathway plays an important role in tumor immunity, and the downstream signaling pathway mediated by the TCR signaling pathway plays a key role in promoting the anti-tumor immunity of $CD8^+$ T cells. The interaction of T cell receptors with MHC antigenic peptide complexes leads to changes in T cells at the molecular and cellular levels and mediates the activation of various genes (Shah et al., 2021). The IL-17 signaling pathway is associated with angiogenesis, and IL17 is an important pro-inflammatory factor. IL-17 can promote the activation of the STAT3 signal transduction pathway through the intermediate mediator IL-6 in tumor cells. This leads to the upregulation of angiogenic factors, thereby promoting tumor angiogenesis (Yu et al., 2007; Lavecchia et al., 2011; Zhang et al., 2012). Previous studies have shown that EMILIN2 regulates angiogenesis. For example, EMILIN2 expression in gastric cancer was found to be related to angiogenesis (Andreuzzi et al., 2020). Collectively, these results suggest that the DEGs are involved in various immune processes and may affect angiogenesis. This indicates that EMILIN2 expression is significantly associated with tumor immunity.

We divided ccRCC patients into two subtypes by screening signature genes from the gene set of 27 immune cell types. We identified 835 DEGs while analyzing differences in the expression between subtypes.

To further investigate the factors affecting the survival of ccRCC patients, we studied the intersection of the DEGs with the DEGs in the high and low EMILIN2 expression groups. These DEGs was associated with the TGF−beta signaling pathway, Cytokine−cytokine receptor interaction, Viral protein interaction with cytokine and cytokine receptor, IL−17 signaling pathway and Transcriptional misregulation in cancer were significantly correlated. In the training set, we constructed a 6-gene risk score model based on the 59 genes by combining lasso regression and multivariate regression. We then confirmed the predictive ability in the validation set. TNNT1, SAA1, IL20RB, and COL22A1 are risk factors for ccRCC, whereas B3GALT5 and C10orf99 act as protective factors when these six genes are expressed at high levels. TNNT1 was shown to be associated with various cancers, such as colorectal cancer (Chen et al., 2020). It was shown to promote the progression of colorectal cancer (Hao et al., 2020). In addition, TNNT1 may also promote the proliferation of breast cancer cells by promoting G1/S phase transition (Shi et al., 2018). SAA1 is transcriptionally activated by STAT3 and accelerates renal interstitial fibrosis by inducing ER stress (Zhang et al., 2021). Concurrently, high SAA1 expression is associated with poor prognosis in advanced renal cell carcinoma (Li et al., 2021). IL20RB overexpression can promote cell proliferation, invasion and migration of papillary renal cell carcinoma. And its overexpression can lead to a poor prognosis in patients (Cui et al., 2019). COL22A1 plays an important role in maintaining vascular homeostasis. In addition, COL22A1 mutation may be closely associated with the occurrence of intracranial aneurysms (Ton et al., 2018). High B3GALT5 expression is associated with tumor progression and the metastasis of breast cancer, and may lead to a poor prognosis for patients with breast cancer (Liao et al., 2021). C10orf99 is associated with the proinflammatory response of skin keratinocytes and affects skin barrier formation (Dainichi et al., 2022). Meanwhile, in colon cancer, C10orf99 expression may induce G1 arrest, leading to the inhibition of colon cancer cell growth (Pan et al., 2014). RT-qPCR results revealed that the mRNA expression levels of TNNT1 and IL20RB in 293T cells were higher than those in 786-O cells, indicating that TNNT1 and IL20RB expression in 786-O cells were limited compared with 293T cells. Also, RT-qPCR analysis showed that the mRNA expression levels of SAA1 and B3GALT5 in 786-O cells were higher than those in 293T cells, which may be an important factor affecting the survival of ccRCC patients. In addition, the mRNA expression levels of C10orf99 and COL22A1 in these 2 cells were not significantly different. In conclusion, these expression of these six genes is closely related to the occurrence, and progression of cancer and angiogenesis in tumors. Thus, these genes have

significant research value. The findings also indicate the reliability of the risk score model. Next, a more comprehensive nomogram model was established and verified by combining the risk score with clinical information.

However, our study may have had certain limitations, primarily because of the limited number of samples. Thus, further experiments are needed to verify the role of these six model genes. Although EMILIN2 was shown to affect angiogenesis, further experimental investigation is needed to confirm whether it affects tumor immunity by regulating angiogenesis in ccRCC.

Conclusion

The strengths of our study are the detailed analysis of the association of EMILIN2 in tumor immunity, the molecular typing of ccRCC, and the development of a 6-gene prognostic model with a high AUC value. Then a comprehensive nomogram model was established to predict the 1/3/5-year survival rate of ccRCC patients combined with clinical information. In conclusion, our study suggests that EMILIN2, a gene that regulates angiogenesis, may be a potential target and candidate marker for the treatment of ccRCC patients.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

Author contributions

GZ and QC designed the study. GZ, JZ, and KT completed the data sorting and analysis of this research. GZ finished writing the manuscript. GZ, JZ, KT, and QC completed the revision of the manuscript.

Funding

This study was supported in part by the Natural Science Foundation of Fujian Province, China (Grant No. 2021J01206).

Acknowledgments

We thank Dr. Junzhong Lai, Dr. Heng Zhao, Dr. Yangfan Zhang and the members of the Chen laboratory for technical assistance and helpful discussions.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1058207/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Expression of TNNT1, COL22A1, B3GALT5, and C10orf99 in the Human Protein Atlas protein database.

References

Andreuzzi, E., Fejza, A., Capuano, A., Poletto, E., Pivetta, E., Doliana, R., et al. (2020). Deregulated expression of Elastin Microfibril Interfacer 2 (EMILIN2) in gastric cancer affects tumor growth and angiogenesis. *Matrix Biol. Plus* 6-7, 6–7. doi:10.1016/j.mbplus.2020.100029

Bressan, G. M., Daga-Gordini, D., Colombatti, A., Castellani, I., Marigo, V., and Volpin, D. (1993). Emilin, a component of elastic fibers preferentially located at the elastin-microfibrils interface. *J. Cell Biol.* 121 (1), 201–212. doi:10.1083/jcb.121.1.201

Brunet, J. P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* 101 (12), 4164–4169. doi:10.1073/pnas.0308531101

Chen, Y., Wang, J., Wang, D., Kang, T., Du, J., Yan, Z., et al. (2020). TNNT1, negatively regulated by miR-873, promotes the progression of colorectal cancer. *J. Gene Med.* 22 (2), e3152. doi:10.1002/jgm.3152

Colombatti, A., Doliana, R., Bot, S., Canton, A., Mongiat, M., Mungiguerra, G., et al. (2000). The EMILIN protein family. *Matrix Biol.* 19 (4), 289–301. doi:10.1016/s0945-053x(00)00074-3

Cui, X. F., Cui, X. G., and Leng, N. (2019). Overexpression of interleukin-20 receptor subunit beta (IL20RB) correlates with cell proliferation, invasion and migration enhancement and poor prognosis in papillary renal cell carcinoma. *J. Toxicol. Pathol.* 32 (4), 245–251. doi:10.1293/tox.2019-0017

Dainichi, T., Nakano, Y., Doi, H., Nakamizo, S., Nakajima, S., Matsumoto, R., et al. (2022). C10orf99/GPR15L regulates proinflammatory response of keratinocytes and barrier formation of the skin. *Front. Immunol.* 13, 825032. doi:10.3389/fimmu.2022.825032

Fejza, A., Polano, M., Camicia, L., Poletto, E., Carobolante, G., Toffoli, G., et al. (2021). The efficacy of anti-PD-L1 treatment in melanoma is associated with the

expression of the ECM molecule EMILIN2. *Int. J. Mol. Sci.* 22 (14), 7511. doi:10.3390/ijms22147511

Foroutan, M., Bhuva, D. D., Lyu, R., Horan, K., Cursons, J., and Davis, M. J. (2018). Single sample scoring of molecular phenotypes. *BMC Bioinforma.* 19 (1), 404. doi:10.1186/s12859-018-2435-4

Hao, Y. H., Yu, S. Y., Tu, R. S., and Cai, Y. Q. (2020). TNNT1, a prognostic indicator in colon adenocarcinoma, regulates cell behaviors and mediates EMT process. *Biosci. Biotechnol. Biochem.* 84 (1), 111–117. doi:10.1080/09168451.2019.1664891

Koul, H., Huh, J. S., Rove, K. O., Crompton, L., Koul, S., Meacham, R. B., et al. (2011). Molecular aspects of renal cell carcinoma: A review. *Am. J. Cancer Res.* 1 (2), 240–254.

Lavecchia, A., Di Giovanni, C., and Novellino, E. (2011). STAT-3 inhibitors: State of the art and new horizons for cancer treatment. *Curr. Med. Chem.* 18 (16), 2359–2375. doi:10.2174/092986711795843218

Li, S., Cheng, Y., Cheng, G., Xu, T., Ye, Y., Miu, Q., et al. (2021). High SAA1 expression predicts advanced tumors in renal cancer. *Front. Oncol.* 11, 649761. doi:10.3389/fonc.2021.649761

Liao, Y. M., Wang, Y. H., Hung, J. T., Lin, Y. J., Huang, Y. L., Liao, G. S., et al. (2021). High B3GALT5 expression confers poor clinical outcome and contributes to tumor progression and metastasis in breast cancer. *Breast Cancer Res.* 23 (1), 5. doi:10.1186/s13058-020-01381-9

Motzer, R. J., Escudier, B., McDermott, D. F., George, S., Hammers, H. J., Srinivas, S., et al. (2015). Nivolumab versus everolimus in advanced renal-cell carcinoma. *N. Engl. J. Med.* 373 (19), 1803–1813. doi:10.1056/NEJMoa1510665

Motzer, R. J., Tannir, N. M., McDermott, D. F., Arén Frontera, O., Melichar, B., Choueiri, T. K., et al. (2018). Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N. Engl. J. Med.* 378 (14), 1277–1290. doi:10.1056/NEJMoa1712126

Pan, W., Cheng, Y., Zhang, H., Liu, B., Mo, X., Li, T., et al. (2014). CSBF/C10orf99, a novel potential cytokine, inhibits colon cancer cell growth through inducing G1 arrest. *Sci. Rep.* 4, 6812. doi:10.1038/srep06812

Paulitti, A., Andreuzzi, E., Bizzotto, D., Pellicani, R., Tarticchio, G., Marastoni, S., et al. (2018). The ablation of the matricellular protein EMILIN2 causes defective vascularization due to impaired EGFR-dependent IL-8 production affecting tumor growth. *Oncogene* 37 (25), 3399–3414. doi:10.1038/s41388-017-0107-x

Petejova, N., and Martinek, A. (2016)). Renal cell carcinoma: Review of etiology, pathophysiology and risk factors. *Biomed. Pap. Med. Fac. Univ. Palacky. Olomouc Czech. Repub.* 160 (2), 183–194. doi:10.5507/bp.2015.050

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

Shah, K., Al-Haidari, A., Sun, J., and Ju, K. (2021). T cell receptor (TCR) signaling in health and disease. *Signal Transduct. Target. Ther.* 6 (1), 412. doi:10.1038/s41392-021-00823-w

Shi, Y., Zhao, Y., Zhang, Y., AiErken, N., Shao, N., Ye, R., et al. (2018). TNNT1 facilitates proliferation of breast cancer cells by promoting G1/S phase transition. *Life Sci.* 208, 161–166. doi:10.1016/j.lfs.2018.07.034

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

Ton, Q. V., Leino, D., Mowery, S. A., Bredemeier, N. O., Lafontant, P. J., Lubert, A., et al. (2018). Collagen COL22A1 maintains vascular stability and mutations in COL22A1 are potentially associated with intracranial aneurysms. *Dis. Model. Mech.* 11 (12), dmm033654. doi:10.1242/dmm.033654

Vuong, L., Kotecha, R. R., Voss, M. H., and Hakimi, A. A. (2019). Tumor microenvironment dynamics in clear-cell renal cell carcinoma. *Cancer Discov.* 9 (10), 1349–1357. doi:10.1158/2159-8290.CD-19-0499

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation.* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Xu, T., Le, T. D., Liu, L., Su, N., Wang, R., Sun, B., et al. (2017). CancerSubtypes: An R/bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics* 33 (19), 3131–3133. doi:10.1093/bioinformatics/btx378

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

Yu, H., Kortylewski, M., and Pardoll, D. (2007). Crosstalk between cancer and immune cells: Role of STAT3 in the tumour microenvironment. *Nat. Rev. Immunol.* 7 (1), 41–51. doi:10.1038/nri1995

Zhang, F., Zhou, X., Zou, H., Liu, L., Li, X., Ruan, Y., et al. (2021). SAA1 is transcriptionally activated by STAT3 and accelerates renal interstitial fibrosis by inducing endoplasmic reticulum stress. *Exp. Cell Res.* 408 (1), 112856. doi:10.1016/j.yexcr.2021.112856

Zhang, X., Yue, P., Page, B. D., Li, T., Zhao, W., Namanja, A. T., et al. (2012). Orally bioavailable small-molecule inhibitor of transcription factor Stat3 regresses human breast and lung cancer xenografts. *Proc. Natl. Acad. Sci. U. S. A.* 109 (24), 9623–9628. doi:10.1073/pnas.1121606109

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 December 2022
DOI 10.3389/fgene.2022.1069921

OPEN ACCESS

EDITED BY
Yi Yao,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Dahmane Oukrif,
University College London, United Kingdom
Yanmei Zou,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Fei Ma,
wafsfd@sina.com
Baoliang Guo,
baoliangguo2013@163.com

†These authors have contributed equally to this work

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

RECEIVED 14 October 2022
ACCEPTED 28 November 2022
PUBLISHED 13 December 2022

CITATION
Li Y, Feng J, Wang T, Li M, Zhang H, Rong Z, Cheng W, Duan Y, Chen Z, Hu A, Yu T, Zhang J, Shang Y, Zou Y, Ma F and Guo B (2022), Construction of an immunogenic cell death-based risk score prognosis model in breast cancer. *Front. Genet.* 13:1069921. doi: 10.3389/fgene.2022.1069921

Construction of an immunogenic cell death-based risk score prognosis model in breast cancer

Yanling Li†, Jianyuan Feng†, Ting Wang†, Mingcui Li, Hanyu Zhang, Zhiyuan Rong, Weilun Cheng, Yunqiang Duan, Ziang Chen, Anbang Hu, Tianshui Yu, Jiarui Zhang, Yuhang Shang, Yiyun Zou, Fei Ma* and Baoliang Guo*

Department of General Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

Immunogenic cell death (ICD) is a form of regulated cell death that elicits immune response. Common inducers of ICD include cancer chemotherapy and radiation therapy. A better understanding of ICD might contribute to modify the current regimens of anti-cancer therapy, especially immunotherapy. This study aimed to identify ICD-related prognostic gene signatures in breast cancer (BC). An ICD-based gene prognostic signature was developed using Lasso-cox regression and Kaplan-Meier survival analysis based on datasets acquired from the Cancer Genome Atlas and Gene Expression Omnibus. A nomogram model was developed to predict the prognosis of BC patients. Gene Set Enrichment Analysis (GESA) and Gene Set Variation Analysis (GSVA) were used to explore the differentially expressed signaling pathways in high and low-risk groups. CIBERSORT and ESTIMATE algorithms were performed to investigate the difference of immune status in tumor microenvironment of different risk groups. Six genes (*CALR*, *CLEC9A*, *BAX*, *TLR4*, *CXCR3*, and *PIK3CA*) were selected for construction and validation of the prognosis model of BC based on public data. GSEA and GSVA analysis found that immune-related gene sets were enriched in low-risk group. Moreover, immune cell infiltration analysis showed that the immune features of the high-risk group were characterized by higher infiltration of tumor-associated macrophages and a lower proportion of $CD8^+$ T cells, suggesting an immune evasive tumor microenvironment. We constructed and validated an ICD-based gene signature for predicting prognosis of breast cancer patients. Our model provides a tool with good discrimination and calibration abilities to predict the prognosis of BC, especially triple-negative breast cancer (TNBC).

KEYWORDS

immunogenic cell death, immunotherapy, breast cancer, prognosis model, RNA-seq

Introduction

Breast cancer (BC) is the most prevalent cancer worldwide, causing 685,000 deaths in 2020, approximately 17% of cancer deaths in females (Sung et al., 2021). BC is a heterogeneous disease characterized by molecular and histological evidence. Treatment approaches and outcomes differ between subtypes. Hormone receptor [estrogen receptor (ER), progesterone receptor (PR)] and human epidermal receptor 2 (HER2) categorize BC into molecular subtypes, and also serve as prominent prognostic biomarkers (Pashayan et al., 2020). Other frequently utilized prognosis predictors in clinical practice include tumor size, tumor grade, the presence and number of axillary node metastases and ki-67 index (Donegan, 1997). In recent years, high-throughput sequencing technologies have made identifying novel biomarkers more achievable. The PAM50 assay, developed on the expression levels of selected gene signatures, aids to risk stratification strategies and treatment decisions (Ellis et al., 2011). Oncotype DX, another validated multigene test, contributes to screening patients with high risk of recurrence and can potentially benefit from adjuvant chemotherapy (Mariotto et al., 2020).

Cancer cells constantly interact with their microenvironment, especially immune cells. Immune cell-associated parameters, such as Immunoscore, have shown promising value for predicting clinical outcomes (Pagès et al., 2018; Galon and Bruni, 2020). Immunogenic cell death (ICD) refers to a cell death process that elicits immune response, which has been widely explored *in vivo* and *in vitro*, and is reviewed in detail by Kroemer et al. (2022). Cancer cells that undergo ICD generate tumor-specific immunity and long-term immunological memory (Krysko et al., 2012). Anti-cancer treatments, mainly conventional chemotherapeutics and radiation therapy can act as cellular stressors, inducing the emission of damage-associated molecular patterns (DAMPs) by cancer cells and activating downstream danger signaling (Galluzzi et al., 2017). ICD-related DAMPs, including surface-exposed calreticulin (CRT), secreted ATP and high mobility group protein B1 (*HMGB1*) can be recognized by pattern recognition receptors (PRRs) that are expressed by immune cells, resulting the activation of tumor suppressing immune response [recruitment of antigen presenting cells (APCs) and T cells, etc.] (Galluzzi et al., 2020b). *HMGB1* is positively correlated with overall survival in BC patients received neo-adjuvant chemotherapy (Exner et al., 2016). Reciprocally, tumor cells can subvert ICD through loss or downregulation of essential components in danger signaling (Galluzzi et al., 2017). Harnessing ICD or targeting ICD subversion strategies may provide new solutions to cancer treatment.

In this study, we screened ICD-associated biomarkers and developed a risk model that predicts the immune microenvironment, and prognosis in BC patients.

Materials and methods

Datasets

The gene expression profiles and clinicopathological data of TCGA-BRCA (n = 1,218) were accessed through UNSC Xena (https://xena.ucsc.edu/). For external validation, raw gene expression and clinical data (n = 123) were directly accessed through the Gene Expression Omnibus (GEO; accession number: GSE37181; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37181). The immunotherapy dataset were downloaded from GEO (accession number: GSE194040; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194040).

Identification of differentially expressed genes

Differential expression analysis of cancer (n = 1,097) and normal (n = 121) samples was performed using the DESeq2 R package (1.35.0). The screening criteria for mRNAs differential expression were determined as p value < 0.05 and absolute fold-change >1.5.

Consensus clustering

The R package ConcensusClusterPlus (1.59.0) was utilized to conduct consensus clustering to identify molecular subtypes according to a selected list of ICD-related genes based on previous research. We performed the clustering using K-means algorithm, and assessed the ideal cluster numbers between k = 2–10. This process was repeated 1,000 times to ensure the results were stable.

Construction of the immunogenic cell death-related risk score

Among 1,218 breast cancer samples, 399 samples without overall survival (OS) information and 60 samples with an observation time of 0 month were excluded. The remaining 759 samples were included for subsequent analyses. Kaplan-Meier analysis was performed to identify ICD-related DEGs with an impact on OS, using R packages survival (3.3-1) and survminer (0.4.9). The ICD-related DEGs with statistical significance were exposed to a LASSO cox regression analysis, as implemented in the R package glmnet (4.1-4). The risk score was constructed by using the regression coefficients derived from Cox regression analysis:

$$RS = \sum_{i=1}^{6} Coef_i DEG_i$$

Statistical analysis

Patients (n = 759) were classified into high-risk group (n = 390) or low-risk group (n = 369) according to the risk score with the cutoff value (risk score = 7.319) generated by the surv_cutpoint function in the R package survminer (0.4.9). The Kaplan–Meier survival curves were constructed by the function "gsurvplot," and the log-rank test was performed between the two groups. Multivariable Cox regression analysis was used to assess whether the risk score was an independent prognostic indicator, and the features to be included in the prognostic model were selected using two-way stepwise regression. A nomogram was plotted based on the clinical features and the risk score. The nomogram's discrimination performances were quantitatively assessed by the area under curve (AUC) of the receiver operating characteristic (ROC) curve and calibration curve. The Wilcoxon rank sum test was conducted to examine whether the risk score distribution differs among BC molecular subtypes.

Functional enrichment

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were carried out between high-risk and low-risk groups. The R package clusterProfiler (4.3.4) (Yu et al., 2012) was employed to evaluate GO and KEGG pathways, and the threshold of p-value was set as <0.05.

Gene set variation analysis

Gene Set Variation Analysis (GSVA) were conducted to determine the gene-set activity score for each sample, utilizing the R package GSVA (1.43.1). The gene sets were the c2 curated signatures downloaded from the Molecular Signature Database (MSigDB) of Broad Institute. The differential analysis of gene-set activity scores between the high-risk and low-risk groups was carried out by the R package limma (3.51.8). GSVA performed on the I-SPY2 dataset used the five gene signature (*CALR*, *TLR4*, *CXCR3*, *PIK3CA*, and *BAX*), because *CLEC9A* expression was not profiled in the dataset.

Immunophenoscore score and tumor immune exclusion score

The IPS score is calculated based on representative cell-type gene expression z-scores, with higher scores indicating increased immunogenicity. The IPS scores of high-risk and low-risk patients were obtained from the Cancer Immunome Atlas (TCIA) (https://tcia.at/home).

The tumor immune exclusion score was generated using expression signatures from immunosuppressive cells, which correlated negatively with T cell infiltration level. The tumor immune exclusion scores were calculated by TIDE (http://tide.dfci.harvard.edu/) (Jiang et al., 2018).

Immune infiltration analysis

CIBERSORT was applied to estimate the proportions of tumor-infiltrating immune with a deconvolution algorithm by the R package CIBERSORT (0.1.0). Besides, the ESTIMATE R package (1.0.13) was used to calculate ESTIMATE immune score of each sample.

Somatic mutation analysis

Somatic mutation data of the high-risk group (n = 373) and the low-risk group (n = 334) were retrieved from TCGA GDC Data Portal (https://portal.gdc.cancer.gov/) in maf format. The waterfall plots were illustrated by the Maftools R package (2.12.0).

Results

Consensus clustering identified two immunogenic cell death-associated subtypes

We conducted extensive literature research and collected 56 ICD-associated genes from previous studies (Supplementary Table S1). Next, consensus clustering was performed according to the patients' expression levels of the ICD-associated genes. Unsupervised consensus clustering identified two major sample clusters that were clearly molecularly distinguishable among patients with BC (Figures 1A–C). To investigate the ICD status in different clusters, we illustrated a heatmap, and found that in contrast with cluster 2, cluster 1 had higher expression levels of ICD-related genes (Figure 1E). To screen out the potentially significant genes in BC, differential analysis was performed between all cancer samples and normal samples, and in 3 molecular subtypes, respectively (Supplementary Figure S3). In total, 18 ICD-related genes (*AIM2*, *ANXA1*, *BAX*, *CALR*, *CCL2*, *CLEC9A*, *CXCR2*, *CXCR3*, *DDX58*, *IL1B*, *IL1R1*, *LRP1*, *P2RY2*, *PIK3CA*, *TLR3*, *TLR4*, *YKT6*, and *ZBP1*) were differently expressed in cancer samples in comparison to normal samples (Supplementary Table S2). Among the 18 DEGs, *AIM2*, and *ANXA1* were the most upregulated and downregulated, respectively (Figure 1D).

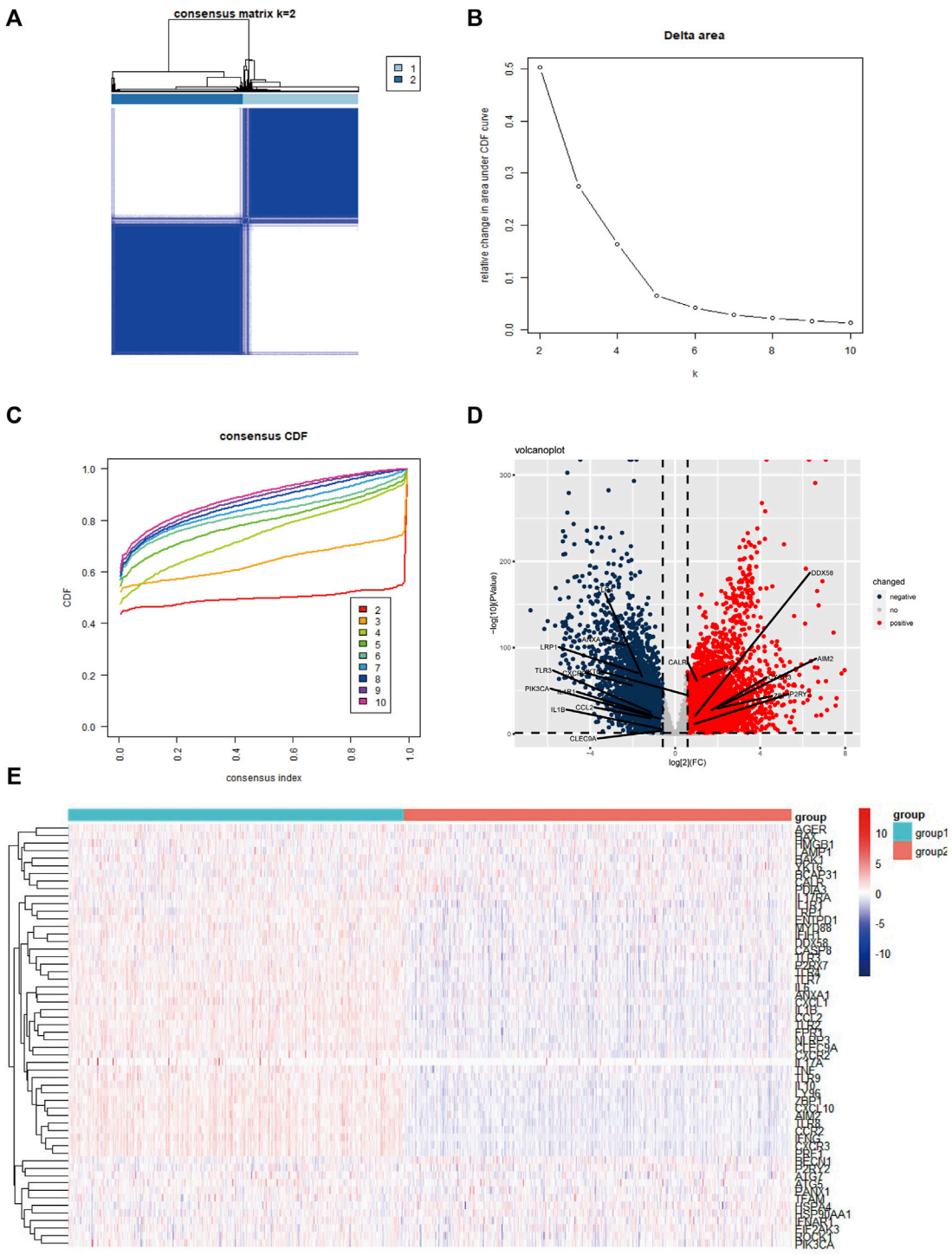


FIGURE 1
Identification of two ICD-based clusters and differential expressed genes **(A)** Heatmap of consensus clustering when $K = 2$ for 56 genes in breast cancer samples; **(B,C)** Delta area curve and the cumulative distribution function (CDF) curves for $k = 2–10$; **(D)** Volcano plot shows the differential expressed genes between cancer and normal samples; **(E)** Heatmap of 56 ICD-related genes' expression levels in two clusters. Red indicates high expression and blue indicates low expression; abbreviations: ICD, immunogenic cell death.

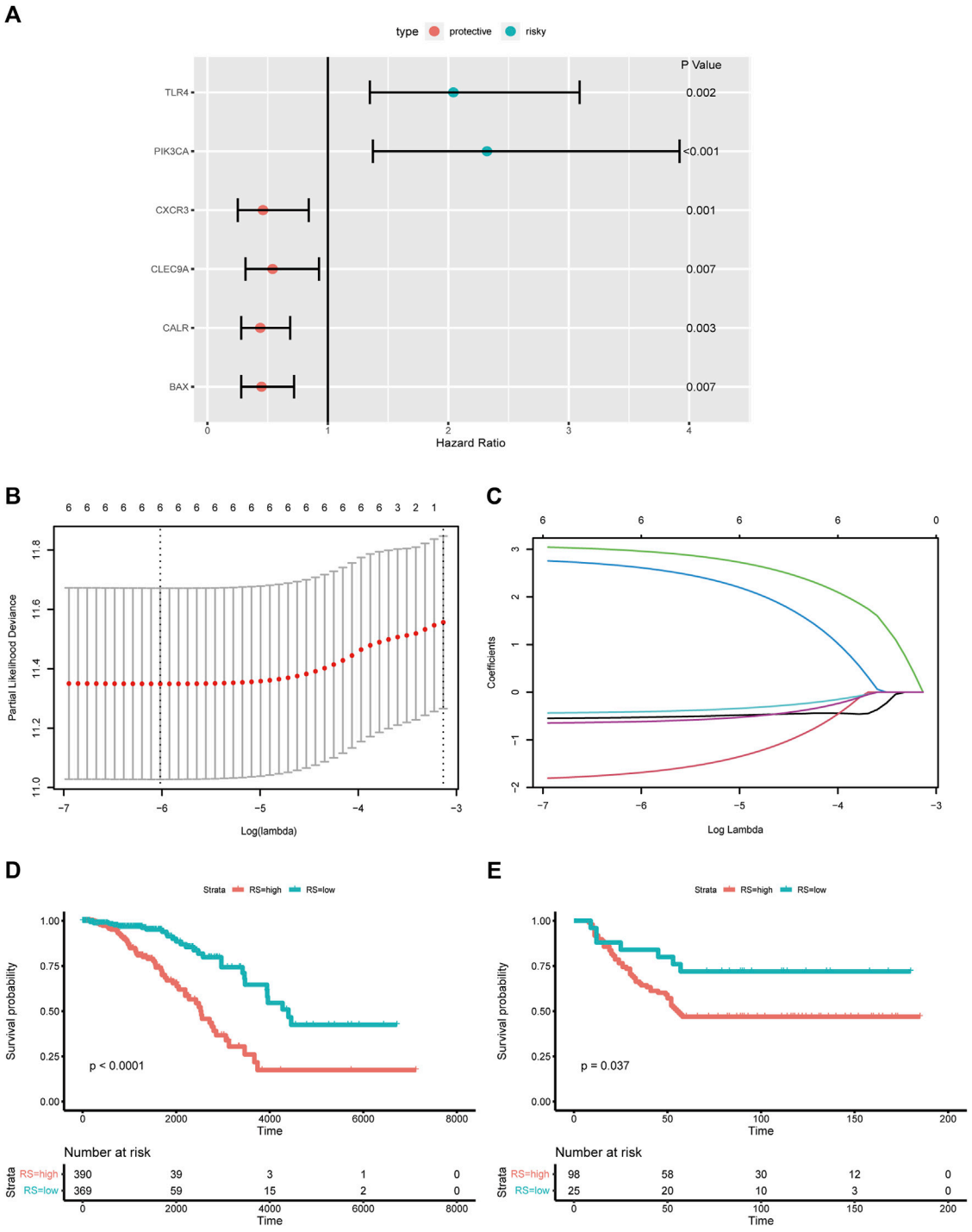


FIGURE 2
Construction and validation of the ICD-based risk score **(A)** Forest plot shows the HRs of six ICD-related DEGs; **(B,C)** Lasso regression of six ICD-related DEGs; **(D)** Kaplan-Meier analysis of the ICD-based risk score in training cohort; **(E)** Kaplan-Meier analysis of the ICD-based risk score in external validation cohort; abbreviations: HR, hazard ratio; ICD, immunogenic cell death; DEGs, differential expressed genes.

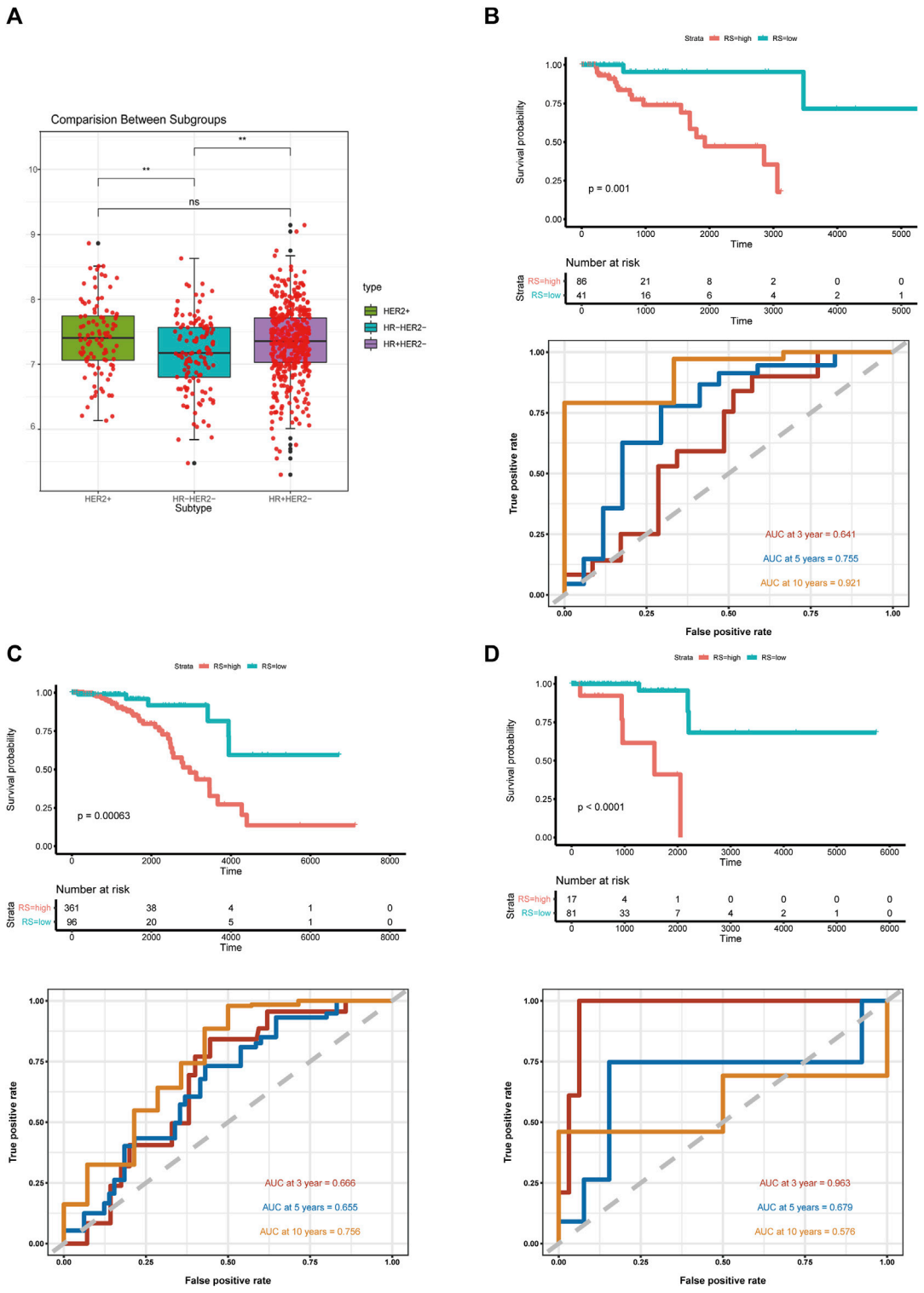


FIGURE 3
Risk score distribution in breast cancer molecular subtypes **(A)** Box plot of ICD-based risk score distribution among breast cancer molecular subtypes; **(B)** Kaplan-Meier analysis and ROC curves of TNBC; **(C)** Kaplan-Meier analysis and ROC curves of HR+ breast cancer samples; **(D)** Kaplan-Meier analysis and ROC curves of HER2+ breast cancer samples; abbreviations: ICD, immunogenic cell death; TNBC, triple-negative breast cancer; HR, hormone receptor; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

Construction and validation of the immunogenic cell death-based risk score

To assess the association of ICD- related DEGs with OS, we performed Kaplan-Meier analysis, and found that six DEGs were statistically significant (Figure 2A). All six ICD-related genes were tested and selected for constructing the ICD-based risk score in the LASSO regression analysis (Figures 2B,C). The risk-score model was developed premised on the regression coefficients derived from multivariate cox regression. The formula for the risk score was as below: Risk score = 3.0954369*PIK3CA + 2.8466656*TLR4 + (−0.5698641)*BAX + (−1.8812416)*CALR + (−0.4513673)*CLEC9A + (−0.6615107) *CXCR3. The prognostic significance of this risk score in BC was further examined by Kaplan-Meier analysis (Figure 2D). For external validation, data from GSE37181 were utilized, and the result was in concordant with the TCGA cohort (Figure 2E).

Risk score distribution in breast cancer molecular subtypes

BC is a heterogeneous disease. In the training set we used for model development, 682 patients had records of hormone receptor and HER2 receptor data, in which hormone receptor-positive (HR+ HER2−) patients accounted for 67% (n = 457), triple-negative BC (TNBC) patients and HER2+ (HR+ HER2−/HR− HER2+) patients accounted for 19% (n = 127) and 14% (n = 98), respectively. Since the number of HER2 positive patients is relatively small, we defined the HER2+ group irrespective of the hormone receptor status. Wilcoxon test indicated that the risk score distribution was statistically different between TNBC and HR+ or HER2+ groups (Figure 3A). To confirm the predictive ability of the risk score in different subtypes, we carried out Kaplan-Meier analysis and generated ROC curves in each subtype (Figures 3B–D). The ICD-based risk score was effective in three molecular subtypes, especially in the TNBC group, for the area under the ROC curve (AUC) reached 0.921 for 10-year OS.

Immunogenic cell death-based risk score is an independent prognostic factor in breast cancer

To further validate the prediction power of the risk score, we evaluated the prognostic effect of ICD-based risk score with age, nodule status, ER status, HER2 status and T stage in univariate cox regression analysis and multivariate cox regression analysis (Figures 4A,B). The risk score and age were independent prognostic factors according to the results, and two-way stepwise regression used in multivariate cox regression selected ER status, age and risk score to develop the final prognostic model. A nomogram was constructed based on multivariate cox regression results (Figure 4C), and the ROC curves and calibration curves were generated for 3-, 5- and 10-year survival (Figure 4D). The AUCs of the nomogram were 0.768, 0.737, and 0.729 for 3-, 5- and 10-year survival. We further validated the model in an external validation cohort, and the ROC curves and calibration curves were illustrated in Supplementary Figures S1A–C. In general, higher OS rates were associated with a lower risk score, younger age and ER-positive status.

Identification of differentially expressed signaling pathways in different risk groups

For better understanding of the pathogenic molecular mechanism underlying the disparity of prognosis in two risk groups, we performed GO and KEGG analyses. The DEGs in low-risk group were enriched in gene sets associated with immunity, including regulation of immune effector process in GO analysis, and antigen processing and presentation, natural killer cell-mediated cytotoxicity, Th1, Th2 and Th17 cell differentiation, T cell receptor signaling pathways, B cell receptor signaling pathways and PD-L1 expression pathways in KEGG analysis (Figures 5A–C). GSVA was used to compare the expression of immune-related signatures across the training datasets, using REACTOME pathway gene sets (Figure 5D). Compared with the high-risk group, most of the pathways were enriched in the low-risk group. In contrast, interleukin-16-associated pathways expressed higher in the high-risk group, which are pro-tumorigenesis, according to previous studies (Grivennikov and Karin, 2011). Furthermore, the GSVA analysis showed that regulation of innate immune response to cytosolic DNA was the most enriched pathway in low-risk TNBC in comparison to its high-risk counterpart, which is crucial for ICD danger signaling (Figure 5E).

Since high TMB is associated with more neoantigens that could be recognized by the immune system, we analyzed somatic mutation profiles between the two risk groups (Maleki Vareki, 2018). TP53 and PIK3CA were the most frequent mutations in both groups, and the median TMB of the low-risk group was slightly higher than the high-risk group (Supplementary Figures S2A,B).

High risk score is associated with immune suppressive tumor microenvironment

Given that the ICD-based risk score was related to tumor immunity, we next assessed the immune status of tumor microenvironment in different groups. We performed CIBERSORT algorithm to calculate the immune cells in the two risk groups. Low-risk group was associated with

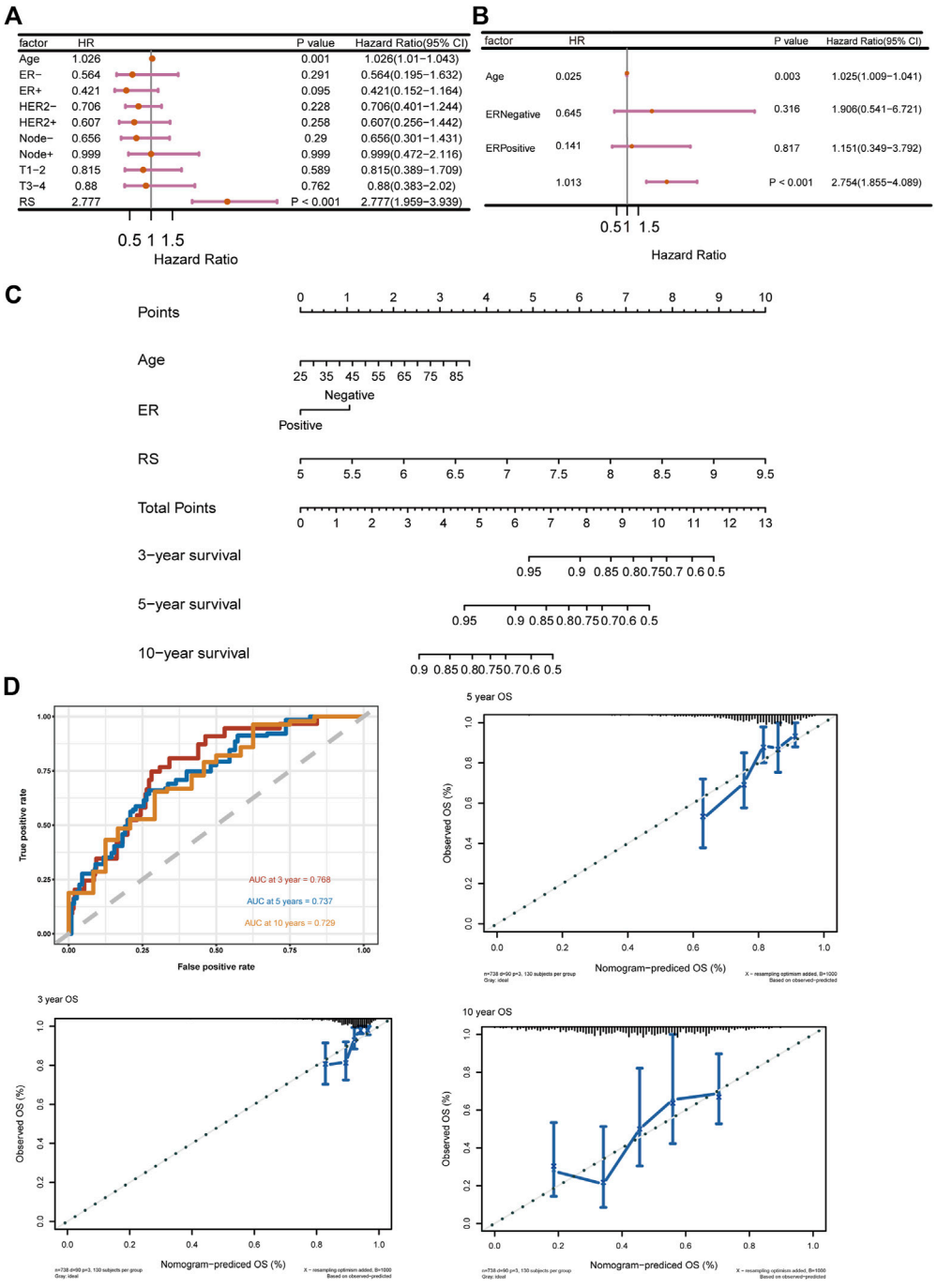


FIGURE 4
ICD-based risk score is an Independent prognostic factor **(A)** Univariate cox regression of ICD-based risk score and other risk factors in breast cancer; **(B)** Multivariate cox regression of ICD-based risk score, age, and ER status in breast cancer; **(C)** Nomogram based on multivariate cox analysis results; **(D)** ROC curves and calibration curves for 3-, 5- and 10-year survival; abbreviations: ICD, immunogenic cell death; ER, estrogen receptor.

considerably more $CD8^+$ T cells and fewer M2 macrophages than the high-risk group (Figure 6A). Immunophenoscore (IPS) was constructed using the expression of immune-related gene signatures, including MHC molecules, immunomodulators, effector cells and suppressor cells. Higher immunophenoscore represents higher tumor immunogenicity (Charoentong et al., 2017). Both IPS and the immune score calculated by the ESTIMATE algorithm in low-risk group were statistically

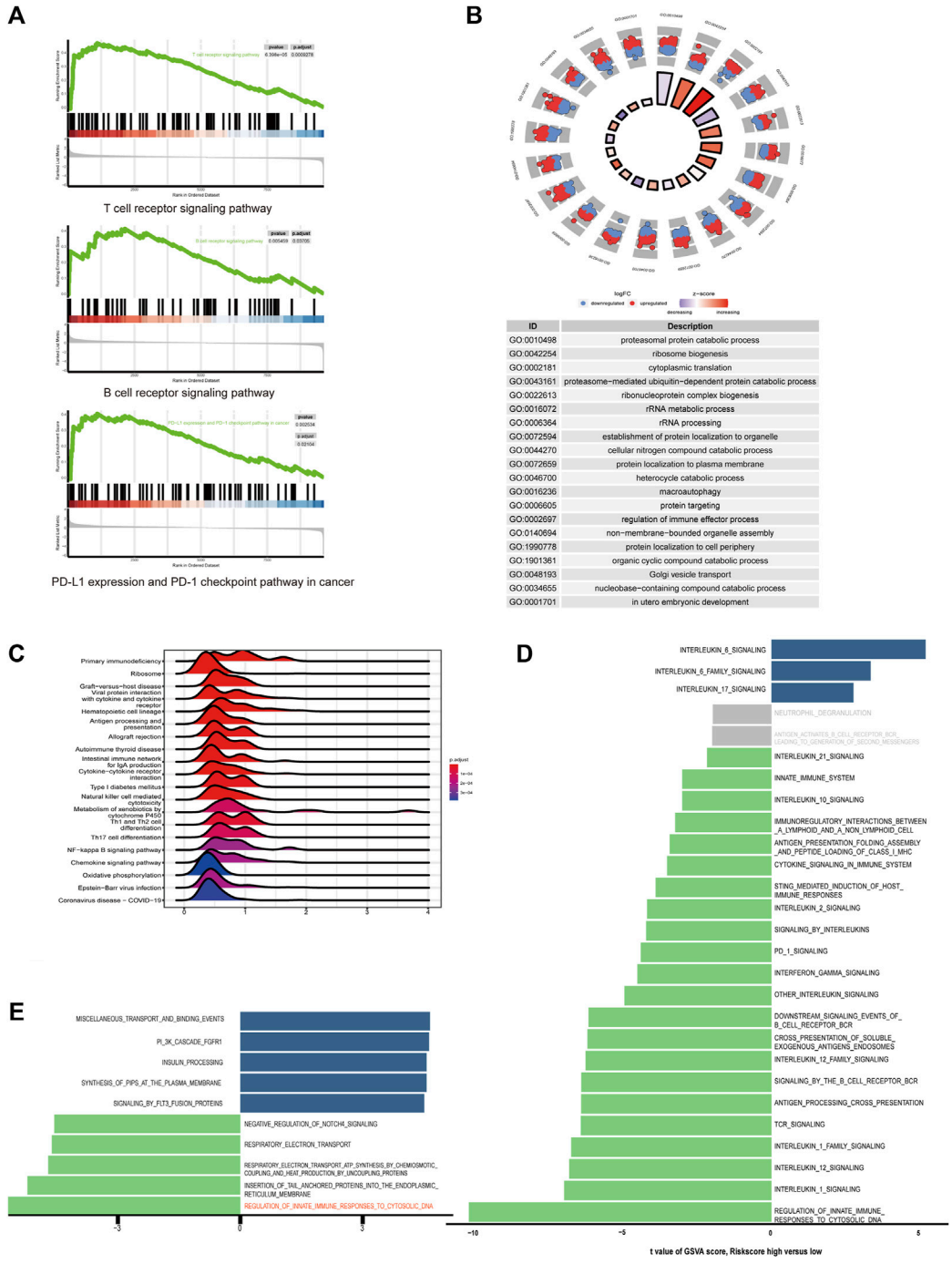


FIGURE 5
Differentially expressed signaling pathways in high and low-risk groups **(A)** Lymphocyte and immune check point related pathways enriched in low-risk group in GSEA analysis. **(B)** Ridge plot of KEGG analysis between high-risk and low-risk groups. The color of the ridges represents adjust *p*-value. **(C)** Circle plot of GO analysis. The color of the dots indicates log2 fold change. **(D)** Bar plot of GSVA analysis of immune-related pathways in Reactome database ordered by t score. T scores between −2 and 2 are colored in grey. **(E)** Bar plot of GSVA analysis in triple-negative breast cancer.

higher than high-risk group, indicating immune hot tumors (Figure 6B). The tumor immune exclusion score, which was generated based on signatures of immune suppressive cells, namely cancer-associated fibroblasts (CAFs), myeloid-derived suppressor cells (MDSCs) and the M2 subtype of tumor-associated macrophages (TAMs), showed a higher median

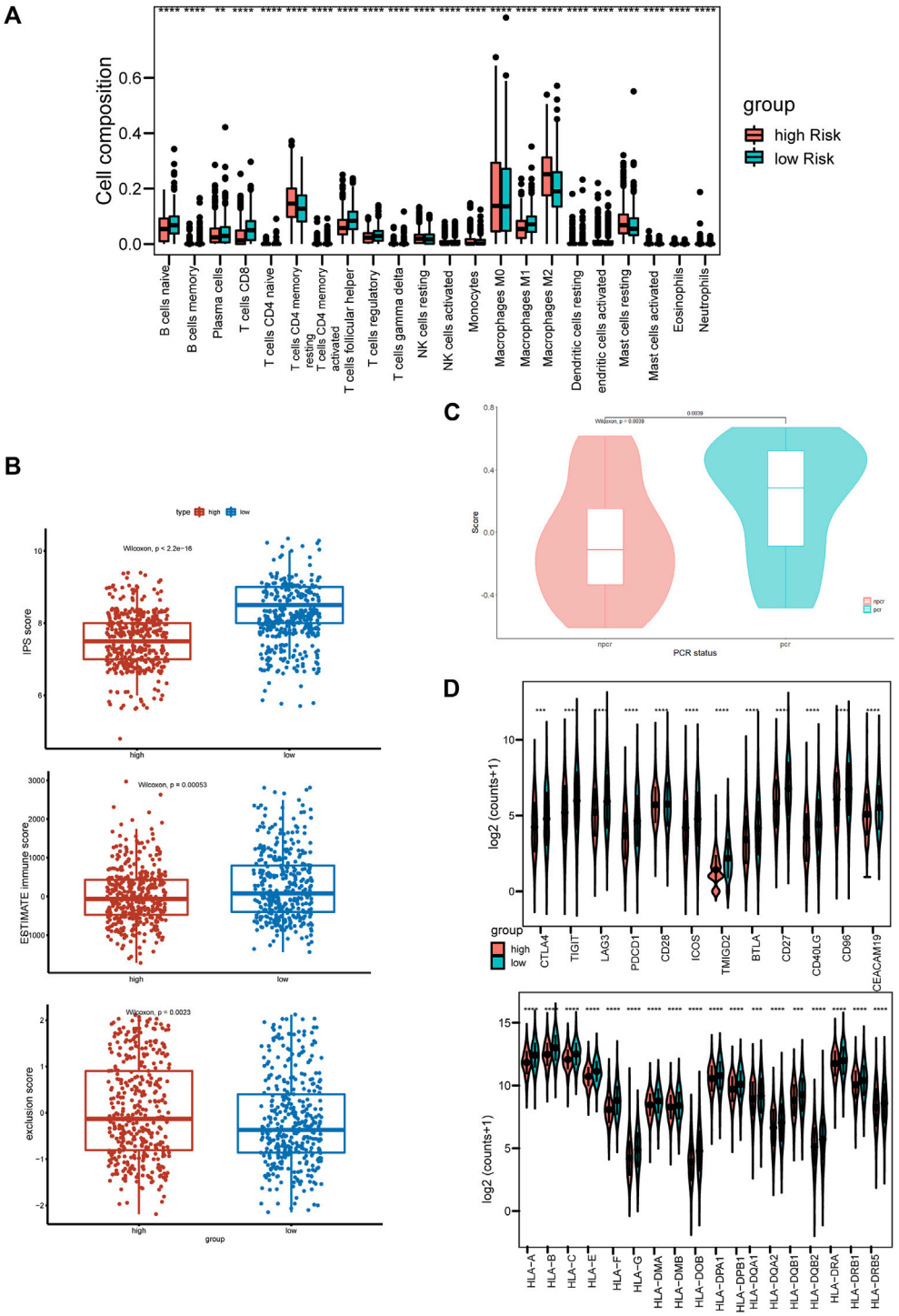


FIGURE 6
High-risk score is associated with immune-suppressive tumor microenvironment **(A)** Boxplot of immune cell compositions calculated by CIBERSORT algorithm of high-risk and low-risk groups. **(B)** Boxplots of IPS score (top), ESTIMATE immune score (middle), and TIDE exclusion score (bottom). Red and blue represent the high-risk group, and low-risk group, respectively. **(C)** Violin plot of the gene signature scores in PCR and non-PCR groups. **(D)** The expression levels of co-stimulation and HLA molecules. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

score in high-risk group, indicating an immune suppressive microenvironment (Figure 6B). Furthermore, in the differential analysis, most of the human leukocyte antigen (HLA) genes and immune checkpoints had significantly higher expression levels in the low-risk group (Figure 6D). To further validate the clinical importance of the ICD-related gene signature, GSVA analysis was performed on the I-SPY2 trial dataset. For patients treated with Pembrolizumab and achieved Pathologic Complete Response (PCR), the GSVA scores were significantly higher than the non-PCR group, indicating the gene signature was positively correlated with immunotherapy responses (Figure 6C).

Discussion

One of the leading causes of ICD is anti-cancer therapy, including chemotherapy (Galluzzi et al., 2017). Multiple chemotherapy drugs commonly used in BC have been demonstrated to have immune- modulatory effect, including Anthracyclines and Taxanes (Ramakrishnan et al., 2010; Mattarollo et al., 2011). Immune checkpoint inhibitors (ICIs), which target PD-1 and PD-L1 improve therapeutic efficacy by enhancing immunogenicity, and the combination of ICIs with conventional chemotherapy drugs performs a synergetic effect (Galluzzi et al., 2020a). A Series of clinical trials have proven the efficacy of combination therapies in BC, and found the clinical benefits correlated with patients' immune status, such as the presence and abundance of tumor-infiltrating lymphocytes (TILs) (Nanda et al., 2020; Schmid et al., 2020). Therefore, it could be advantageous to identify ICD-related biomarkers that help with the risk stratification of BC patients.

In this study, we demonstrated that the ICD-related genes are closely associated with prognosis and tumor microenvironment of BC. We identified six differentially expressed genes that impacted overall survivals of BC patients and developed a prognosis model with external validation. Moreover, we found that the ICD-based risk score was closely associated with tumor immune microenvironment. Previous studies have confirmed that both immunotherapy and chemotherapy induce anti-tumor immune responses, including the expansion of $CD8^+$ T cells, etc. (Krysko et al., 2012; Philip and Schietinger, 2022). In our study, high-risk score indicates the immune exclude subtype, which can be potentially improved by immunotherapy and chemotherapy. Interestingly, our results showed better long-term prognostic power for HR+ HER2− subtype and TNBC instead of HER2+ subtype. This finding could suggest that immune status, which is closely related to the efficacy of immunotherapy and chemotherapy is more important in HER2− subtypes rather than in HER2+ subtype, for HER2− targeted therapy brings significant benefits to HER2+ patients.

The genes we selected for model construction have been proved to play essential roles in tumor growth, invasion, and metastasis. *CALR* and *BAX*, components of ICD-danger signaling pathways, are both independent prognosis predictors in BC (Binder et al., 1996; Lwin et al., 2010). *PIK3CA* and *TLR4* contribute to tumorigenesis through the phosphoinositide 3 (PI3)-kinase/Akt signaling pathway and IPS/TLR4 pathway, respectively (Verret et al., 2019; Afroz et al., 2022). *CLEC9A* and *CXCR3* are associated with intratumoral dendritic cells (DCs), which are necessary for anti-tumor immunity. *CLEC9A* is a biomarker for DCs, while chemoattractant receptor *CXCR3* influences the biological function of DCs (de Mingo Pulido et al., 2018; Hammerl et al., 2021). Furthermore, Xu et al. (2022) classified ICD-associated DAMPs into three subtypes in TNBC patients, among which the inflammatory DAMPs was featured with high expression of *CALR*, higher anti-tumor immune cell infiltration, and better prognosis. This finding is in concordant with our results, for the low-risk group had considerably higher expression of *CALR* (Xu et al., 2022).

Conclusion

In conclusion, our study addressed the importance of ICD in the modulation of tumor immune microenvironment in breast cancer. Besides, we constructed and validated an ICD-based prognostic signature, which served significant value in predicting OS of breast cancer 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

YL and ZR wrote the main manuscript. YL, ML, HZ, TW, and WC performed bioinformatic analysis. YD, JF, ZC, AH, TY, YZ, YS, and TW collected datasets and prepared figures. FM and BG designed 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/fgene.2022.1069921/full#supplementary-material

References

Afroz, R., Tanvir, E. M., Tania, M., Fu, J., Kamal, M. A., and Khan, M. A. (2022). LPS/TLR4 pathways in breast cancer: Insights into cell signalling. *Curr. Med. Chem.* 29, 2274–2289. doi:10.2174/0929867328666210811145043

Binder, C., Marx, D., Binder, L., Schauer, A., and Hiddemann, W. (1996). Expression of Bax in relation to Bcl-2 and other predictive parameters in breast cancer. *Ann. Oncol.* 7, 129–133. doi:10.1093/oxfordjournals.annonc.a010538

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18, 248–262. doi:10.1016/j.celrep.2016.12.019

De Mingo Pulido, Á., Gardner, A., Hiebler, S., Soliman, H., Rugo, H. S., Krummel, M. F., et al. (2018). TIM-3 regulates CD103+ dendritic cell function and response to chemotherapy in breast cancer. *Cancer Cell* 33, 60–74. doi:10.1016/j.ccell.2017.11.019

Donegan, W. L. (1997). Tumor-related prognostic factors for breast cancer. *Ca. Cancer J. Clin.* 47, 28–51. doi:10.3322/canjclin.47.1.28

Ellis, M. J., Suman, V. J., Hoog, J., Lin, L., Snider, J., Prat, A., et al. (2011). Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: Clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype--ACOSOG Z1031. *J. Clin. Oncol.* 29, 2342–2349. doi:10.1200/JCO.2010.31.6950

Exner, R., Sachet, M., Arnold, T., Zinn-Zinnenburg, M., Michlmayr, A., Dubsky, P., et al. (2016). Prognostic value of HMGB1 in early breast cancer patients under neoadjuvant chemotherapy. *Cancer Med.* 5, 2350–2358. doi:10.1002/cam4.827

Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2017). Immunogenic cell death in cancer and infectious disease. *Nat. Rev. Immunol.* 17, 97–111. doi:10.1038/nri.2016.107

Galluzzi, L., Humeau, J., Buqué, A., Zitvogel, L., and Kroemer, G. (2020a). Immunostimulation with chemotherapy in the era of immune checkpoint inhibitors. *Nat. Rev. Clin. Oncol.* 17, 725–741. doi:10.1038/s41571-020-0413-z

Galluzzi, L., Vitale, I., Warren, S., Adjemian, S., Agostinis, P., Martinez, A. B., et al. (2020b). Consensus guidelines for the definition, detection and interpretation of immunogenic cell death. *J. Immunother. Cancer* 8, e000337. doi:10.1136/jitc-2019-000337

Galon, J., and Bruni, D. (2020). Tumor immunology and tumor evolution: Intertwined histories. *Immunity* 52, 55–81. doi:10.1016/j.immuni.2019.12.018

Grivennikov, S. I., and Karin, M. (2011). Inflammatory cytokines in cancer: Tumour necrosis factor and interleukin 6 take the stage. *Ann. Rheum. Dis.* 70 (1), i104–i108. doi:10.1136/ard.2010.140145

Hammerl, D., Martens, J. W. M., Timmermans, M., Smid, M., Trapman-Jansen, A. M., Foekens, R., et al. (2021). Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. *Nat. Commun.* 12, 5668. doi:10.1038/s41467-021-25962-0

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558. doi:10.1038/s41591-018-0136-1

Kroemer, G., Galassi, C., Zitvogel, L., and Galluzzi, L. (2022). Immunogenic cell stress and death. *Nat. Immunol.* 23, 487–500. doi:10.1038/s41590-022-01132-2

Krysko, D. V., Garg, A. D., Kaczmarek, A., Krysko, O., Agostinis, P., and Vandenabeele, P. (2012). Immunogenic cell death and DAMPs in cancer therapy. *Nat. Rev. Cancer* 12, 860–875. doi:10.1038/nrc3380

Lwin, Z.-M., Guo, C., Salim, A., Yip, G. W.-C., Chew, F.-T., Nan, J., et al. (2010). Clinicopathological significance of calreticulin in breast invasive ductal carcinoma. *Mod. Pathol.* 23, 1559–1566. doi:10.1038/modpathol.2010.173

Maleki Vareki, S. (2018). High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J. Immunother. Cancer* 6, 157. doi:10.1186/s40425-018-0479-7

Mariotto, A., Jayasekerea, J., Petkov, V., Schechter, C. B., Enewold, L., Helzlsouer, K. J., et al. (2020). Expected monetary impact of oncotype DX score-concordant systemic breast cancer therapy based on the TAILORx trial. *J. Natl. Cancer Inst.* 112, 154–160. doi:10.1093/jnci/djz068

Mattarollo, S. R., Loi, S., Duret, H., Ma, Y., Zitvogel, L., and Smyth, M. J. (2011). Pivotal role of innate and adaptive immunity in anthracycline chemotherapy of established tumors. *Cancer Res.* 71, 4809–4820. doi:10.1158/0008-5472.CAN-11-0753

Nanda, R., Liu, M. C., Yau, C., Shatsky, R., Pusztai, L., Wallace, A., et al. (2020). Effect of Pembrolizumab plus neoadjuvant chemotherapy on pathologic Complete response in women with early-stage breast cancer: An analysis of the ongoing phase 2 adaptively randomized I-SPY2 trial. *JAMA Oncol.* 6, 676–684. doi:10.1001/jamaoncol.2019.6650

PagèS, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.-S., Bifulco, C., et al. (2018). International validation of the consensus Immunoscore for the classification of colon cancer: A prognostic and accuracy study. *Lancet (London, Engl.* 391, 2128–2139. doi:10.1016/S0140-6736(18)30789-X

Pashayan, N., Antoniou, A. C., Ivanus, U., Esserman, L. J., Easton, D. F., French, D., et al. (2020). Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat. Rev. Clin. Oncol.* 17, 687–705. doi:10.1038/s41571-020-0388-9

Philip, M., and Schietinger, A. (2022). $CD8^+$ T cell differentiation and dysfunction in cancer. *Nat. Rev. Immunol.* 22, 209–223. doi:10.1038/s41577-021-00574-3

Ramakrishnan, R., Assudani, D., Nagaraj, S., Hunter, T., Cho, H.-I., Antonia, S., et al. (2010). Chemotherapy enhances tumor cell susceptibility to CTL-mediated killing during cancer immunotherapy in mice. *J. Clin. Invest.* 120, 1111–1124. doi:10.1172/JCI40269

Schmid, P., Salgado, R., Park, Y. H., MuñOZ-Couselo, E., Kim, S. B., Sohn, J., et al. (2020). Pembrolizumab plus chemotherapy as neoadjuvant treatment of high-risk, early-stage triple-negative breast cancer: Results from the phase 1b open-label, multicohort KEYNOTE-173 study. *Ann. Oncol.* 31, 569–581. doi:10.1016/j.annonc.2020.01.072

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Verret, B., Cortes, J., Bachelot, T., Andre, F., and Arnedos, M. (2019). Efficacy of PI3K inhibitors in advanced breast cancer. *Ann. Oncol.* 30, x12–x20. doi:10.1093/annonc/mdz381

Xu, M., Lu, J.-H., Zhong, Y.-Z., Jiang, J., Shen, Y.-Z., Su, J.-Y., et al. (2022). Immunogenic cell death-relevant damage-associated molecular patterns and sensing receptors in triple-negative breast cancer molecular subtypes and implications for immunotherapy. *Front. Oncol.* 12, 870914. doi:10.3389/fonc.2022.870914

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics a J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1070048

OPEN ACCESS

EDITED BY
Yi Yao,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Wendy Wu,
Independent researcher, China
Dan Li,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Yukun Zu,
zuyukun@163.com

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

RECEIVED 14 October 2022
ACCEPTED 05 December 2022
PUBLISHED 04 January 2023

CITATION
Wu X, Yin J, Deng Y and Zu Y (2023), Whole-genome characterization of large-cell lung carcinoma: A comparative analysis based on the histological classification.
Front. Genet. 13:1070048.
doi: 10.3389/fgene.2022.1070048

Whole-genome characterization of large-cell lung carcinoma: A comparative analysis based on the histological classification

Xiaowei Wu[1], Jin Yin[2], Yu Deng[1] and Yukun Zu[1]*

[1]Department of Thoracic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Departments of Hematology, Tongji Hospital, Tongji Medical Collage, Huazhong University of Science and Technology, Wuhan, China

Background: According to the 2015 World Health Organization classification, large cell neuroendocrine carcinoma (LCNEC) was isolated from Large-cell lung cancer (LCLC) tumors, which constitutes 2%–3% of non-small cell lung cancer (NSCLC). However, LCLC tumors are still fairly vaguely defined at the molecular level compared to other subgroups.

Materials and Methods: In this study, whole-genome sequencing (WGS) was performed on 23 LCLC and 15 LCNEC tumor specimens. Meanwhile, data from the TCGA (586 LUADs and 511 LUSCs) and U Cologne (120 SCLCs) were analyzed and compared.

Results: The most common driver mutations were found in *TP53* (13/23, 57%), *FAM135B* (8/23, 35%) and *FAT3* (7/23, 30%) in LCLC, while their counterparts in LCNEC were *TP53* (13/15, 87%), *LRP1B* (6/15, 40%) and *FAT1* (6/15, 40%). Notably, *FAM135B* mutations only occurred in LCLC (P = 0.013). Cosmic signature analysis revealed widespread defective DNA mismatch repair and tobacco-induced mutations in both LCLC and LCNEC. Additionally, LCNEC had a higher incidence of chromosomal copy number variations (CNVs) and structural variations (SVs) compared with LCLC, although the differences were not statistically significant. Particularly, chromothripsis SVs was significantly associated with CNVs. Furthermore, mutational landscape of different subtypes indicated differences between subtypes, and there seems to be more commonalty between our cohort and SCLC than with other subtypes. *SMARCA4* mutations may be specific driver gene alteration in our cohort.

Conclusion: Our results support that LCLC and LCNEC tumors follow distinct tumorigenic pathways. To our knowledge, this is the first genome-wide profiling comparison of LCLC and LCNEC.

KEYWORDS

non-small cell lung cancer (NSCLC), Large-cell lung cancer (LCLC), large cell neuroendocrine carcinoma (LCNEC), whole-genome sequencing, histological classification

Introduction

Lung cancer, commonly divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), is the leading cause of cancer-related mortality worldwide (Bray et al., 2018). Large-cell lung cancer (LCLC) is the third most common NSCLC subtype after lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC), representing 2%–3% of NSCLC (Howlader N et al., 2015). Compared to other NSCLCs, LCLC is more malignant due to faster growth and earlier metastasis (Asamura et al., 2006; Shi et al., 2020). Histopathologically, the diagnosis of LCLC is usually excluded from LUAD, LUSC, and SCLC. LCLC is defined as an undifferentiated NSCLC without glandular or squamous cell differentiation in the WHO2004 lung cancer classification, while LCNEC was defined as LCLC with neuroendocrine morphological characteristics and at least one positive neuroendocrine immunohistochemical (IHC) marker (Travis et al., 2004). However, the 2015 WHO classification protocol (Brambilla et al., 2015) now isolates large cell neuroendocrine carcinoma (LCNEC) from LCLC tumors. Furthermore, LCLC expressing previously histologically defined lung cell markers (TTF1, Napsin A) were reclassified as LUAD, and squamous marker positive LCLC (P40, CK5/6, P63) were classified as non-keratinized or basal cell LUSC. Tumors that are surgically removed without expression of these markers are defined as LCLC.

It is of clinical importance to accurately distinguish histological subtypes. Subtype-directed diagnosis and treatment have been widely established for LUAD, LUSC, and SCLC. However, LCLC tumors are still fairly vaguely defined at the molecular level compared to other subgroups, especially given the otherwise strong molecular efforts of the 2015 WHO classification scheme. Some studies of smaller gene sets found abnormal expression of *TP53* in LCLC and LCNEC tumors, with *KRAS* mutations predominating in LCLC (Iyoda et al., 2004; Rossi et al., 2014). There was also a difference in the frequency of oncogene mutations between WHO2004 LCLC tumors expressing LUAD or LUSC markers and those with invalid markers (Rekhtman et al., 2013; Karlsson et al., 2015; Pelosi et al., 2015; Driver et al., 2016). However, studies on the genome-wide altered landscape of LCLC are lacking. In this study, we aimed to investigate the whole-genome landscape of LCLC and LCNEC tumors in relation to other histological subgroups of lung cancer. Our results link recent lung cancer classification schemes to the genome-wide landscape of the disease, supporting that LCLC and LCNEC tumors follow distinct tumorigenic pathways. To our knowledge, this is the first genome-wide profiling comparison of LCLC and LCNEC.

Materials and methods

Patient and tissue selection

A total of 23 LCLC and 15 LCNEC patients who have undergone surgical resection from June 2017 to December 2020 were retrospectively included and analyzed in this study, including 36 males and 2 females. All patients provided written informed consents. This study was approved by the Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (TJ-IRB20220639). The clinical characteristics of the patients were summarized in Table 1. The diagnosis of LCLC and LCNEC were confirmed by two experienced pathologists.

Immunohistochemistry

As previously described, standardized institutional protocols were used for immunohistochemical staining. The whole-slide serial tissue sections from FFPE surgical resection specimens were used to determine the expression levels for PD-L1, P40, CK5/6, P63, TTF1, Napsin A, Ki-67 and other tumor biomarkers. The PD-L1 expression was evaluated by two methods, including the tumor proportion score (TPS), defined as the percentage of viable tumor cells showing partial or complete membrane staining at any intensity (A TPS≥1% was considered as positive), and combined positive score (CPS), defined as the number of PD-L1-positive cells (tumor cells, macrophages and lymphocytes) divided by the total number of tumor cells and multiplied by 100.

Whole genome sequencing and analysis

DNA was extracted from the tumors and paired para-cancer FFPE tissues using the QIAamp DNA FFPE tissue kit (Qiagen, United States). The resultant DNA was then quality-controlled using Nanodrop and Qubit (Thermo Fisher Scientific, United States) to ensure adequate purity and quality. Illumina paired-end libraries were prepared from extracted DNA and sequenced on Illumina HiSeq platforms (Illumina, San Diego, United States), with a mean average coverage of 50 × for both tumors and matched para-cancer tissues.

Burrows-Wheeler Aligner (BWA) was used to align the Paired-end sequencing reads to the human reference genome (hg19), and GATK 4.0 was used to sort and remove PCR duplicates. Somatic single nucleotide variants (SNVs), insertions, and deletions (indel) with default parameters were called with Strelka2 with default parameters (Kim et al., 2018). The ANNOVAR was used to annotate possible variant candidates. Germline mutation was called using best practices with the Genome Analysis Toolkit (GATK) HaplotypeCaller (version 3.6) as previously described (McKenna et al., 2010).Somatic copy number variations (CNVs) identified by FACETS and recurrently occurring CNVs were detected with GISTIC2.0. The GISTIC2.0 was used to identify regions of the genome that are significantly amplified or deleted across a set of samples (Mermel et al., 2011). CNV burden was calculated based on the identified copy number variants as previously described (Wolf

TABLE 1 The patient characteristics and clinicopathological data.

Characteristic	LCLC*	SCLC (U Cologne)<	LUAD (TCGA)	LUSC (TCGA)
Total number	38	120	586	511
Histology				
LCLC	23 (60.5%)			
LCNEC	15 (39.5%)			
Age (y/o)				
≥60	21 (55.3%)			
<60	14 (36.8%)			
unkown	3 (7.9%)			
Gender				
male	36 (94.7%)			
female	2 (5.3%)			
Smoking				
No	9 (23.7%)			
Yes	29 (76.3%)			
Drinking				
No	23 (60.5%)			
Yes	15 (39.5%)			
Family History				
No	28 (73.7%)			
Yes	10 (26.3%)			
PD-L1 expression				
negative	12 (31.6%)			
positive	26 (68.4%)			
Tumor grade				
T1/T2	17 (44.7%)			
T3/T4	21 (55.3%)			
Lymph node metastasis				
No	25 (65.8%)			
Yes	13 (34.2%)			
Radiation/chemotherapy				
No	14 (36.8%)			
Yes	24 (63.2%)			
Patients evaluable for				
mutations	38 (100%)	120	586	511
CNV	38 (100%)			
SV	38 (100%)			

LCLC* implies WHO2004 classification.

et al., 2019), and then the average CNV burden was estimated for each patient. Somatic structural variants were called with ShatterSeek (Cortes-Ciriano et al., 2020) and Manta (Chen et al., 2016).

DDR gene status analysis

DDR inactivation mutation status was determined by retrieving and combining DNA data copy number variation and single nucleotide variation of DDR genes (Tian et al., 2020). Alterations in the DDR pathway were defined as any non-synonymous somatic mutation in the protein-coding region, or homozygous deletions of at least one genes in DDR-related pathway.

Statistics

All analyses were performed using SPSS (version 25 for Windows, Armonk, NY: IBM Corp.). Patient characteristics

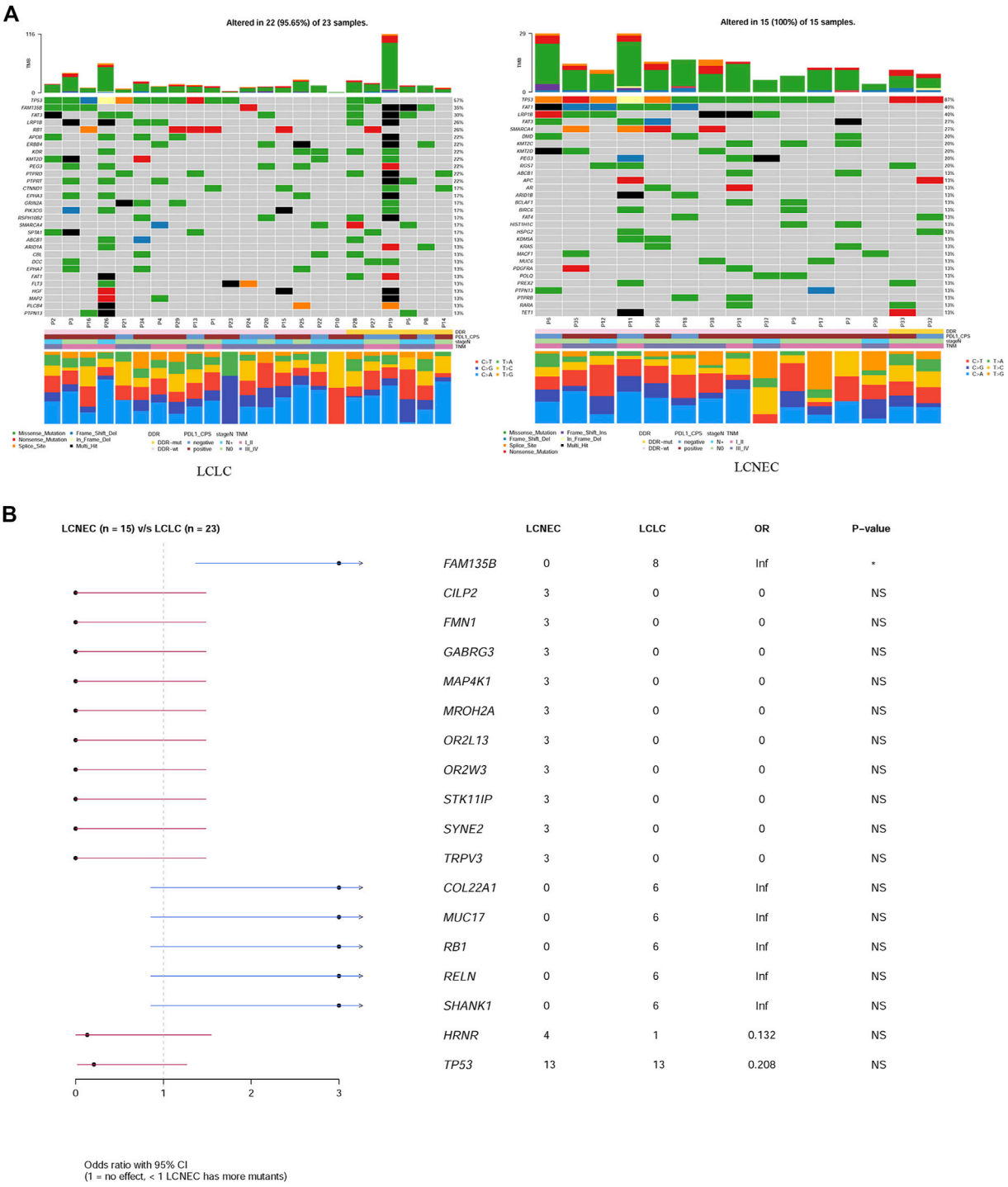


FIGURE 1
Comparison of mutation landscape between LCLC and LCNEC. **(A)** A comparison of the mutational landscapes of LCLC and LCNEC is provided, along with the most frequently mutated driver genes. The top panel represents the TMB and the middle panel represents the matrix of frequently mutated genes. Columns represent samples, and clinicopathological characteristics of individual patients are presented below. Bar plots in the lower panel shows the contribution of six substitutions. **(B)** Forestplot shows the significant differences of driver genes between the two groups.

were evaluated with descriptive statistics. Correlation of histological classification with DDR gene status, PD-L1 expression, age, gender, stage, smoking and drinking was investigated using the chi-square test. All reported p values were two-sided and considered statistically significant at $p < 0.05$, unless otherwise specified.

Results

Clinical characteristics

All participants were pathologically reviewed, and 23 LCLC patients and 15 LCNEC patients were included in the retrospective analysis. The clinical characteristics of patients are summarized in Supplementary Table S1. Twenty-six (68.4%) patients were positive and 12 patients were negative for programmed death ligand-1 (PD-L1) staining. All of the LCNEC patients were male, and 80% of them were current or former smokers. There were no significant differences in age of diagnosis, gender, smoking, drinking history, tumor stage, lymph node metastasis and treatment between LCLC and LCNEC patients (Supplementary Table S1). Interestingly, *TP53*/*RB1* co-mutations, an important molecular subtype of LCNEC, were not present in the 15 LCNEC patients in our cohort.

Mutational landscape of LCLC and LCNEC

Gene mutation profiles of 38 patients with LCLC and LCNEC were analyzed by whole genome sequencing. In total, 9931 non-synonymous somatic mutations were identified in 6013 genes. Thirty-seven of the 38 patients, including 22 LCLC patients and 15 LCNEC patients, showed at least one gene variant. *TP53* and *TTN* were the most common variants in LCLC and LCNEC (Supplementary Figure S1A). Specifically, the mutations of *CILP2*, *FMN1*, *GABRG3*, *MAP4K1*, *MROH2A*, *OR2L13*, *OR2W3*, *STK11IP* and *SYNE2* occurred only in LCNEC group (3/15, 20%, $p = 0.054$, respectively), while *FAM135B* (8/23, 35%, $p = 0.013$) mutations only occurred in LCLC group (Supplementary Figure S1B). *TP53* mutations were the dominant driver gene alteration in both LCLC (13/23, 57%) and LCNEC (13/15, 87%) tumors ($p = 0.077$) (Figure 1A). The frequency of other driver gene variants found was much lower in both subgroups. *FAM135B* (8/23, 35%) and *FAT3* (7/23, 30%) were the second and third most commonly mutated genes in LCLC, while their counterparts in LCNEC were *LRP1B* (6/15, 40%) and *FAT1* (6/15, 40%). These alterations highlight more general differences between the two subgroups. Interestingly, *RB1* (6/23, 26%, $p = 0.063$) mutations were also exclusively found in LCLC cases (Figure 1B). DDR inactivation mutation status was also identified, but in this relatively small retrospective cohort, we found no significant differences in DDR status, PD-L1 expression, lymph node metastasis, or tumor grade associated with histological classification.

In addition, a high frequency of C > A with accompanying C > G has been observed in both LCLC and LCNEC (Figure 1A), indicating a signature of tobacco exposure. To determine the association between the distribution of mutations and cosmic signatures in LCLC and LCNEC patients, mutation signature analyses were performed for all point mutations and the surrounding trinucleotide context. Mutational spectrum of six substitutions revealed a high frequency of C>A transversions and C>T transitions in LCLC and LCNEC. The median percentages of variants of C > A, C > G, C > T, T > A, T > C, and T > G were −40%, 13%, 22%, 11%, 11%, and 4% respectively in LCLC, while −32%, 14%, 25%, 9%, 13%, and 8% respectively in LCNEC ($p = 0.02$) (Figure 2A; Supplementary Figure 2A). The profiles of 96 substitutions exhibited similar results (Figure 2B; Supplementary Figure 2B).

Based on the proportion of mutation signatures in each sample and unsupervised hierarchical clustering, the patients were divided into 5 clusters (Supplementary Figure S2C). The mutation patterns of our cohort were similar to the characteristics of COSMIC "signature 4," "signature 5" and "signature 13." Unsupervised similarity analysis of tumor mutation spectrum of all published signature patterns confirmed that the maximum cosine similarity with these signatures was 0.964, 0.868, and 0.848 respectively (Supplementary Figure S2D). However, there was no significant difference in cosmic signatures between LCLC and LCNEC (Figure 2C). Signatures 3 and 4 were mainly identified in both the LCLC and LCNEC groups, with signature 3 being associated with deficiencies in DNA-double-strand break repair and signature 4 being linked to tobacco-induced mutations. Additionally, the differences between groups were analyzed according to the cluster groups, such as age, drinking, gender, smoking, grade and histopathology, but the results were not statistically significant (Supplementary Figure S2E). Moreover, the tumor mutational burden (TMB) of clustering samples (cluster 2) related to "signature 4" was generally higher (Supplementary Figure S3A). We also found that the mean value of TMB in the LCLC and LCNEC groups was 6.62 mutations per million base pairs (MB) and 4.16 mutations/MB, but the differences were not significant ($p = 0.2$) (Supplementary Figure S3B). The average weighted Genome Instability Index (wGII) score was 0.254 in the LCLC group and 0.309 in the LCNEC group, with no significant difference ($p = 0.65$) (Supplementary Figure S3C).

CNV profiles of LCLC and LCNEC

To characterize specific copy number variations (CNVs), we identified differential copy number variation genes between LCLC and LCNEC groups (Figure 3A). The LCNEC group

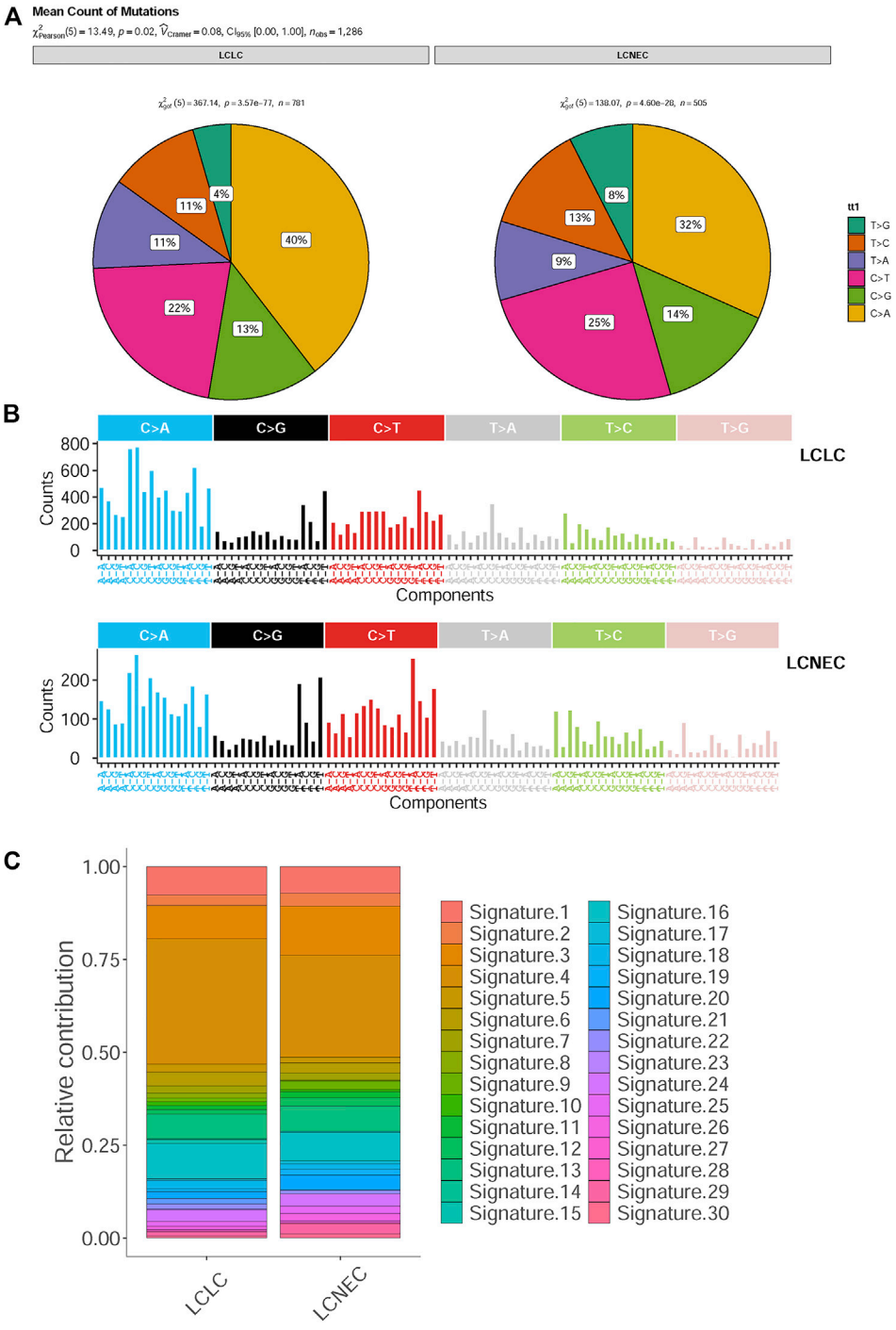


FIGURE 2
Mutational spectrum analysis for LCLC and LCNEC. **(A)** The pie chart shows the relative contribution of the six substitutions. **(B)** Relative contribution of 96 substitution subtypes SNV in each group. **(C)** Relative contributions of mutational signatures in each group.

exhibited a higher rate of chromosome CNV compared with the LCLC group, corresponding to higher CNV burden (22.87/case *vs* 17.15/case), but the difference was not significant (p = 0.33) (Figure 3B). GISTIC2.0 was used to identify significantly amplified or deleted regions of the genome across a set of samples. CNVs were found throughout the genome, with copy number gains being more prevalent than copy number losses. Chromosomes 12p13.31, 19p12, and 9q21.11 were lost and

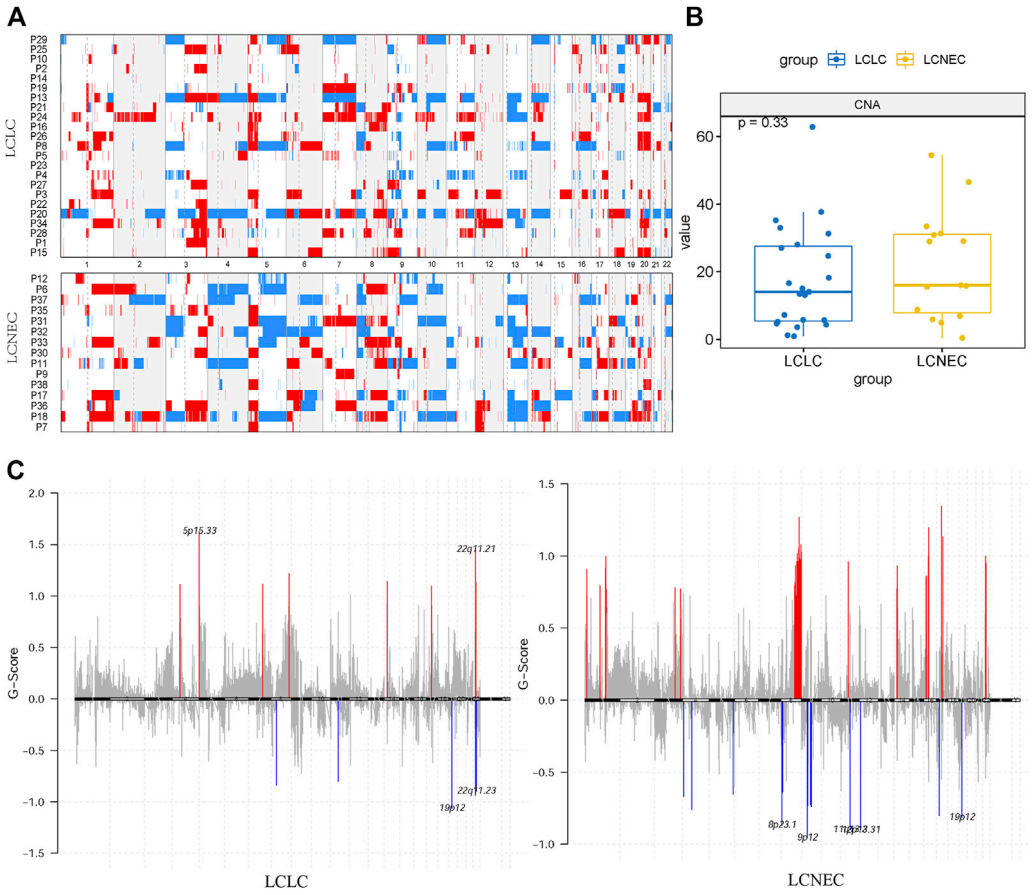


FIGURE 3
Distinct CNA landscape of LCLC and LCNEC. **(A)** Overall copy number variation (CNV) profile of LCLC and LCNEC. Red represented amplification and blue represented deletion. **(B)** Comparison of the CNA burden between LCLC and LCNEC. **(C)** Somatic copy number alterations in each group. Deletions and amplifications are represented on the y-axis by blue or red bars, respectively. Each peak region (cytoband) is displayed together with its known or potential cancer-related genes.

chromosomes 8q24.21, 14q11.2, 16p11.2, and 17q12 were amplified in both LCLC and LCNEC groups. Some chromosomes with CNVs were only identified in LCLC, such as 5P15.33, 7q22.1, and 22q11.23, while some were only identified in LCNEC, such as 9p12, 11q13.2, 8p23.1, 17p11.2, 4q13.2, and 9q12 (Figure 3C, Supplementary Table S2). Specifically, the copy number gains were found at chromosomes 11q11, 17q21.31, 8p11.22, and 1p36.21 in LCNEC, while the copy number losses occurred at these chromosomes in LCLC.

SV patterns of LCLC and LCNEC

To define the patterns of structural variation (SV), ShatterSeek (Cortes-Ciriano et al., 2020) and Manta (Chen et al., 2016) were integrated to implement our final SV catalog. We identified a median of 111 SVs per LCLC patient (range 21–591) and 151 SVs per LCNEC patient (range 31–833) (Figure 4A). Translocation (TRA) accounted for the greatest proportion of all categories (47%), followed by deletion (DEL) at 22%.The count of each SV class was not significantly different between LCLC and LCNEC patients (Figure 4B). In all patients, the different SV classes showed clear patterns of co-occurrence, mutual exclusion, and association with recurrent molecular alterations. For example, the burden of chromothripsis SVs per patient was significantly positively correlated with the number of single deletions (Spearman $p = 0.57$), and negative correlation with single tandem duplication (DUP) (Figure 4C). Furthermore, our results suggest that chromothripsis SVs may be significantly associated with CNVs. We present example of a chromothripsis event in chromosome 9 with CN oscillations that span 3 CN levels showing interspersed loss of heterozygosity and templated insertions, as evidenced by their size, and breakpoint orientations at their edges

10.3389/fgene.2022.1070048

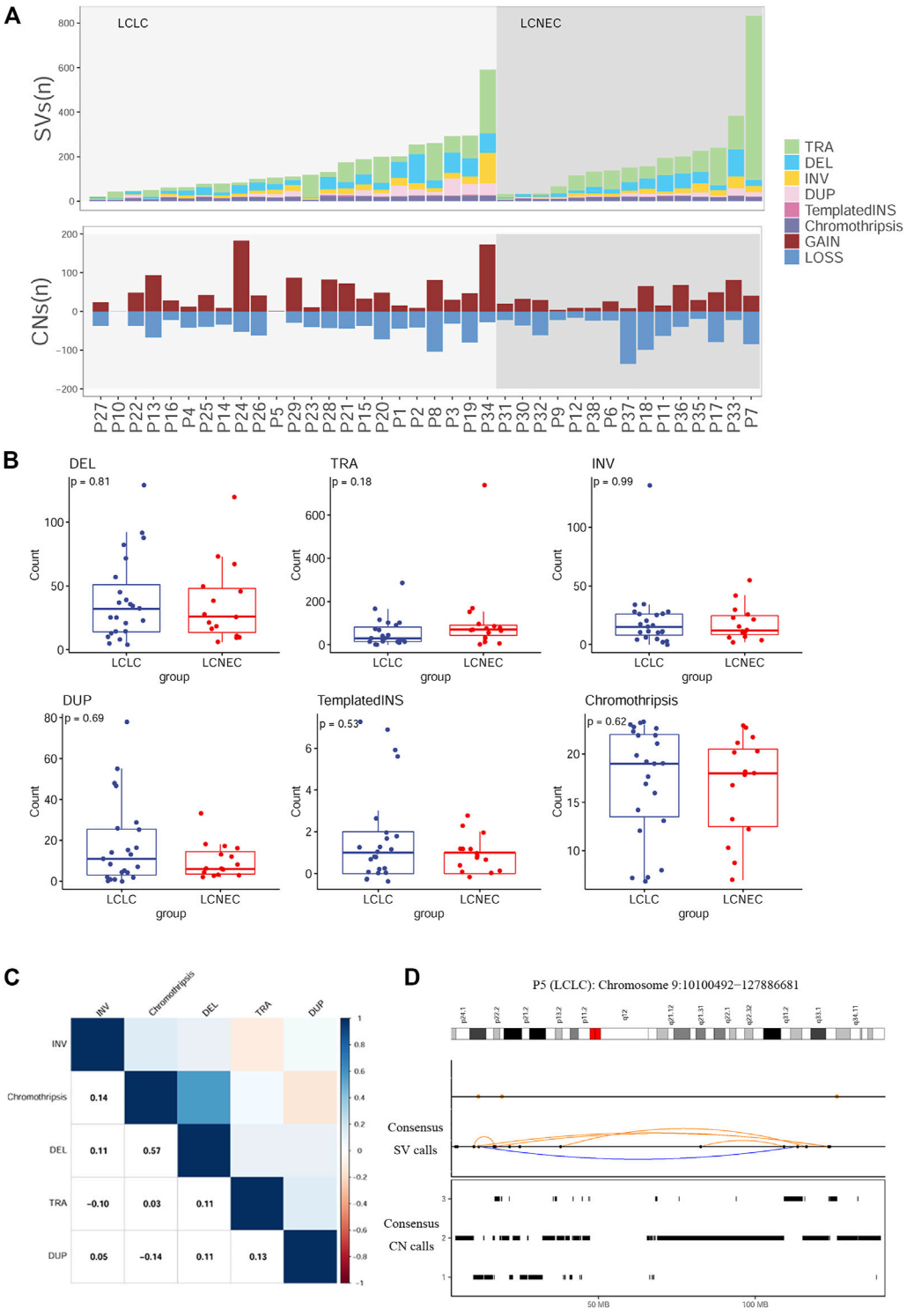


FIGURE 4
Distribution of SVs in LCLC and LCNEC **(A)** Stacked bars show the genome-wide burden of each SV class (color) in each patient (x-axis). Lower panel: SVs resulting in copy-number gain or loss **(B)** Comparison of each SV burden between LCLC and LCNEC. **(C)** Pairwise associations between the numbers of SVs across patients. Color was determined by the magnitude of positive (blue) and negative (red) Spearman correlation coefficients, plotted only where $q < 0.1$. **(D)** Example of a chromothripsis event in chromosome 9 involving CN oscillations with interspersed loss of heterozygosity and templated insertions. Breakpoints corresponding to interchromosomal SVs are depicted as colored dots in the SV profile, whereas intrachromosomal SVs are represented with black dots and colored arcs.

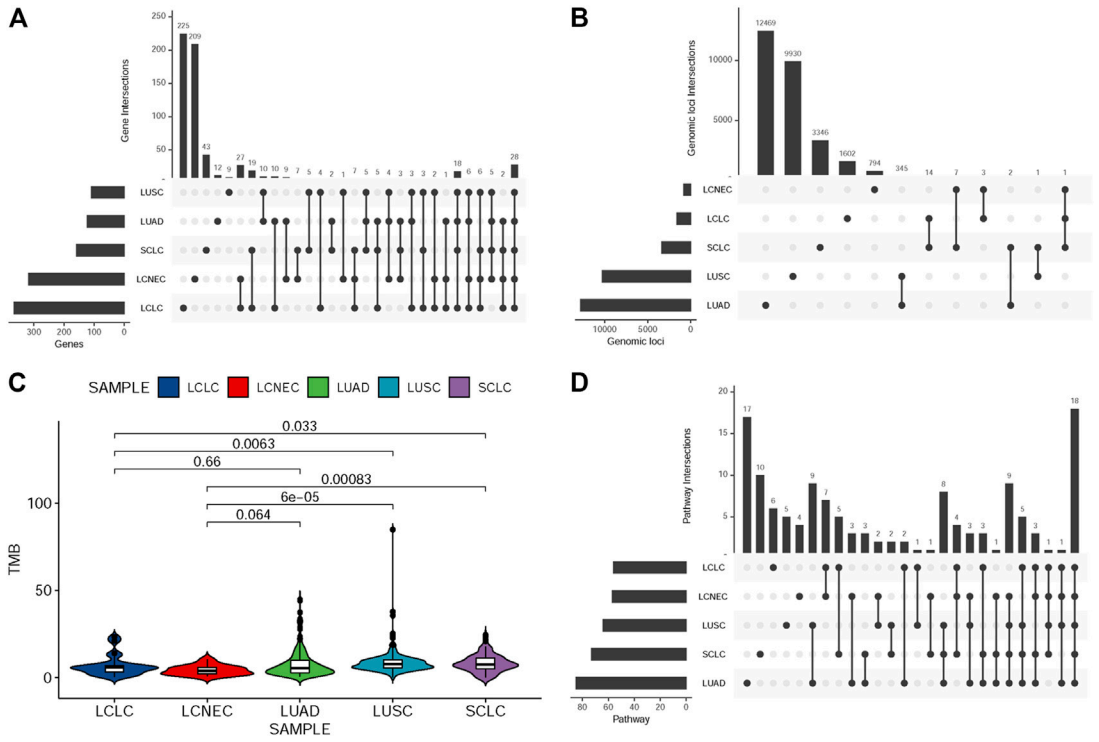


FIGURE 5
Comparison of mutation landscape between this cohort and other three subtypes. **(A)**UpSetR plot shows the overlap of germline-regulated genes identified in the present study for the five lung cancer subtypes. **(B)** UpSetR plot shows the overlap of independent genomic loci that represent the genes shown in **(A)**. **(C)** The TMB difference in the five lung cancer subtypes. **(D)**UpSetR plot shows the overlap of pathways from the Kyoto Encyclopedia of Genes and Genomes enriched with the germline-regulated genes.

(Figure 4D). Similar results were also presented in the chromosomes of LCNEC patients as illustrated in Supplementary Figure S4A.

Comparative analysis with other pathological subtypes

To explore the unique driver genes in LCLC, we compared typical somatic mutation profiles with other lung cancer subtypes (586 LUADs and 511 LUSCs in TCGA and 120 SCLCs in U Cologne). All subtypes share 28 genes and these genes are located in different regions of the genome (FIGURE 5A,B). In addition, we compared the mutation frequency of TOP20 driver genes with other lung cancer subtypes. Results showed that most of these genes had no significant difference in mutation frequency between LCLC* and SCLC (Table 2). Notably, the mutation frequencies of *TP53*, *SMARCA4*, and *RB1* differed significantly across subtypes, suggesting that these may be specific driver gene characteristics of LCLC*, especially *SMARCA4*. Furthermore, the mutation sites of *TP53* and *RB1* were compared with those of the other three subtypes. Most mutation sites of *TP53* are located in the *P53* DNA-bingding domain. Meanwhile, the same *RB1* mutation was found only in SCLC (Table 3, Supplementary FigureS4B). Regarding the TMB, we found significant differences between LCLC and LUSC (p = 0.0063) or SCLC (p = 0.033), but not between LCLC and LUAD, respectively (Figure 5C). To gain a deeper understanding of the biological characteristics driven by these germline regulatory genes, we performed a biopathway enrichment analysis of genes in each subtype. WebGestalt was used to identify pathways that were significantly enriched in each subtype using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. We found that all subtypes share 18 pathways, including small cell lung cancer pathways and non-small cell lung cancer pathways (Figure 5D).

Discussion

Diagnostic terms for LCLC have been applied inconsistently in the clinic, based solely on morphology and insufficient IHC markers. It is of great importance to generate more knowledge regarding the genetic alterations in LCLC to propose more

TABLE 2 Comparison of TOP20 driver genes in different lung cancer subtypes.

	LCLC* (n = 38)	LUAD (n = 586)	*p*-value	LUSC(n = 511)	*p*-value	SCLC (n = 120)	*p*-value
TP53	68.42%	111 (21.51%)	<0.001	146 (29.08%)	<0.001	103 (85.83%)	0.0282
LRP1B	31.58%	88 (17.05%)	0.0457	153 (30.48%)	0.8571	51 (42.50%)	0.2587
FAT3	28.95%	50 (9.69%)	0.0013	32 (6.37%)	<0.001	22 (18.33%)	0.1741
FAT1	23.68%	40 (7.75%)	0.0035	55 (10.96%)	0.0325	16 (13.33%)	0.1339
FAM135B	21.05%	64 (12.40%)	0.1339	70 (13.94%)	0.2324	26 (21.67%)	1
KMT2D	21.05%	22 (4.26%)	0.0004	40 (7.97%)	0.0133	22 (18.33%)	0.8126
PEG3	21.05%	26 (5.04%)	0.0011	35 (6.97%)	0.0067	14 (11.67%)	0.1785
SMARCA4	21.05%	22 (4.26%)	0.0004	18 (3.59%)	0.0002	5 (4.17%)	0.0029
ERBB4	15.79%	26 (5.04%)	0.0167	33 (6.57%)	0.0468	10 (8.33%)	0.2179
PTPRT	15.79%	33 (6.40%)	0.0420	33 (6.57%)	0.0468	11 (9.17%)	0.2448
RB1	15.79%	23 (4.46%)	0.0102	33 (6.57%)	0.0468	87 (72.50%)	<0.001
ABCB1	13.16%	25 (4.84%)	0.0465	34 (6.77%)	0.1810	7 (5.83%)	0.1620
APOB	13.16%	43 (8.33%)	0.3621	36 (7.17%)	0.1958	16 (13.33%)	1
DMD	13.16%	45 (8.72%)	0.3734	66 (13.15%)	1	21 (17.50%)	0.6229
EPHA3	13.16%	29 (5.62%)	0.0740	38 (7.57%)	0.2132	3 (2.50%)	0.0202
KDR	13.16%	29 (5.62%)	0.0740	41 (8.17%)	0.3579	6 (5.00%)	0.1354
KMT2C	13.16%	57 (11.05%)	0.6007	37 (7.37%)	0.2042	12 (10.00%)	0.5587
PTPN13	13.16%	9 (1.74%)	0.0015	18 (3.59%)	0.0174	6 (5.00%)	0.1354
PTPRD	13.16%	91 (17.64%)	0.6570	68 (13.55%)	1	12 (10.00%)	0.5587
SPTA1	13.16%	100 (19.38%)	0.5187	60 (11.95%)	0.7962	23 (19.17%)	0.4729

LCLC* implies WHO2004 classification.

effective diagnosis and new molecular markers of predisposition and prognosis. Studies on LCLC gene profiles mainly focus on small gene sets (Karlsson et al., 2015; Pelosi et al., 2015; Driver et al., 2016), and there are few literatures on whole-genome sequencing profiling. In this study, we performed a genome-wide analysis of 23 LCLC patients and 15 LCNEC patients with a comparative analysis based on the histological classification (Figure 6). To our knowledge, this is the first genome-wide profiling comparison of LCLC and LCNEC.

The role of genes in the therapeutic efficacy of LCLC is often limited. Therefore, in the current NCCN guidelines for NSCLC, LCLC is classified as adenocarcinoma for treatment and molecular detection. In the study, frequent mutations of *TP53* were predictably observed in both LCLC and LCNEC, as well as low frequency alterations in *EGFR*, *BRAF*, and *PIK3CA* genes, consistent with previous studies (Rekhtman et al., 2013; Rossi et al., 2014). Similarly, alterations in tumor suppressors *PTEN* and *STK11* are mainly observed together with *TP53* mutations in both LCLC and LCNEC. Compared with the literature, our cohort also showed differences. We did not observe *KRAS* mutations in LCLC patients, which is the most commonly reported mutation (Karlsson et al., 2015) and this difference may be explained by ethnic differences and limited cohort size. Studies have shown that LCNEC can be further divided into SCLC-like with *TP53*/*RB1* inactivation and NSCLC-like with retained *TP53*/*RB1* functions, with different chemotherapy treatment results (Derks et al., 2018). In our study, all *RB1* mutations occurred in LCLC rather than LCNEC, and 83% (5/6) of them were co-occurring with *TP53* mutations. Unfortunately, our cohort is insufficient for further classification. Additionally, the most common mutations were found in *TTN* and *TP53* genes in both LCLC and LCNEC, with a high total frequency (45%) of a *TTN*/*TP53* double mutation. It has been suggested that *TTN* mutation or *TTN*/*TP53* co-mutation is associated with the prognosis of LUSC (Cheng et al., 2019). However, whether *TTN* is involved in lung cancer development is controversial. The focus of controversy lies in its large and complex structure and the false positive results caused by the heterogeneity of the mutation process (Hofree et al., 2013; Kim et al., 2017). Interestingly, in our cohort, *FAM135B* mutations occurred only in LCLC, with a 35% mutation rate, suggesting that the mutation may be specific to LCLC compared to LCNEC. It has been reported to have high mutation rates in other lung cancers such as LUSC (Xie et al., 2021) and SCLC (Hu et al., 2019; Wang et al.,

TABLE 3 Distribution of mutations in different subtypes of lung cancer.

	LCLC* (n = 38)	LUAD (n = 586)	LUSC (n = 511)	SCLC (n = 120)
TP53	R280G (n = 1, 3%)	R280G (n = 1, 0.2%)	0	0
	R158L (n = 1, 3%)	R158L (n = 3, 0.5%)	0	0
	E271* (n = 1, 3%)	0	E271* (n = 1, 0.2%)	0
	R248W (n = 1, 3%)	0	R248W (n = 1, 0.2%)	0
	R249S (n = 1, 3%)	0	R249S (n = 1, 0.2%)	0
	X307_splice (n = 1, 3%)	0	X307_splice (n = 1, 0.2%)	0
	M237I (n = 1, 3%)	M237I (n = 1, 0.2%)	M237I (n = 1, 0.2%)	0
	R248L (n = 1, 3%)	R248L (n = 1, 0.2%)	R248L (n = 1, 0.2%)	0
	R283P (n = 1, 3%)	R283P (n = 1, 0.2%)	R283P (n = 1, 0.2%)	0
	X225_splice (n = 1, 3%)	X225_splice (n = 1, 0.2%)	X225_splice (n = 1, 0.2%)	0
	R181P (n = 1, 3%)	0	0	R181P (n = 1, 0.8%)
	E294* (n = 1, 5%)	0	0	E294* (n = 1, 0.8%)
	R158P (n = 1, 3%)	R158P (n = 1, 0.2%)	0	R158P (n = 1, 0.8%)
	V172F (n = 1, 3%)	0	V172F (n = 1, 0.2%)	V172F (n = 1, 0.8%)
	X125_splice (n = 1, 3%)	X125_splice (n = 1, 0.2%)	X125_splice (n = 2, 0.4%)	X125_splice (n = 1, 0.8%)
	E298* (n = 1, 3%)	E298* (n = 1, 0.2%)	E298* (n = 3, 0.6%)	E298* (n = 2, 1.7%)
RB1	R445* (n = 1, 3%)	0	0	R445* (n = 1, 0.8%)
	X702_splice (n = 1, 3%)	0	0	X702_splice (n = 1, 0.8%)

LCLC* implies WHO2004 classification.

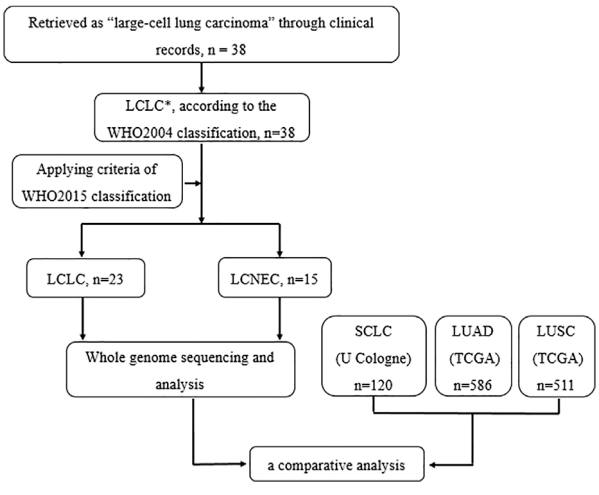


FIGURE 6
Flowchart illustrating patient enrollment and analysis.

2019). Esophageal squamous cell carcinoma has been shown to strongly express *FAM135B* with poor prognosis and silencing *FAM135B* increases radiosensitivity (Bi et al., 2021; Dong et al., 2021), but there is little evidence to support mutation as the underlying cause of elevated expression. In addition, our findings also showed differences in substitutions, copy number variations and structural variations between LCLC and LCNEC. Together, these results support that LCLC and LCNEC tumors follow different tumorigenic paths.

Furthermore, we performed a comparative analysis on the mutational profiles of histologically classified primary LUAD, LUSC, SCLC and above two subtypes. To the best of our knowledge, this is also the first study to compare and contrast these five subtypes. We used a range of different regulatory data to identify SNPs within regulatory regions of the genome that have a defined target gene. We found some overlap in SNPS, genes, and pathways among the five subtypes. LCLC and LCNEC had more private mutated genes than the other three subtypes. It is worth emphasizing that among the 28 genes shared by the 5 lung cancer subtypes, besides *TP53* and *TTN*, some other genes have been reported to be associated with lung cancer. Low-density lipoprotein (LDL) receptor-associated protein 1B (*LRP1B*), a member of the LDL receptor family, is often inactivated in lung cancer. Single gene mutations in *LRP1B* were found to be associated with high TMB in lung

cancer (Lan et al., 2019), which may be associated with favorable outcomes with immune checkpoint inhibitors (Brown et al., 2021). Similarly, *FAT3* mutations have been reported to be associated with NSCLC prognosis and elevated TMB levels (Qiu et al., 2020). Interestingly, LUAD subsets with co-mutations in *FAT3* and *LRP1B* showed significantly prolonged immunotherapy progression-free survival (PFS) (Zhu et al., 2021). Mutations in anti-matrix metalloproteinase mucin 16 (*MUC16*) have been reported to be potentially associated with air pollution, thus contributing to the development of air pollution-associated lung cancer (Chen et al., 2019). *MUC16* overexpression induced by gene mutations promotes lung cancer cell growth, metastasis and chemoresistance (Lakshmanan et al., 2017; Kanwal et al., 2018). In addition to the overlapping genes mentioned above, our pathway enrichment analysis revealed that 18 biological pathways were shared among the 5 subtypes. Most of these are cancer-related signaling pathways, including pathways in small cell lung cancer and non-small cell lung cancer. The genes were also significantly enriched in important tumor onset and metastasis pathways, such as "ECM-receptor interaction" and "Focal adhesion".

Our study has several limitations that should be noted. First, our sample size was not very large, however, due to the low incidence of LCLC, it took us 3 years and 6 months to collect these 38 samples from June 2017 to December 2020 (a total of 76 tissues, including 23 LCLCs and 15 LCNECs), making it difficult to collect more samples in the limited time available. Non-etheless, more patients and more complete clinical data (including regular follow-up) are needed in the future to validate the results of this study. Second, our study lacked other omics analyses that may provide more molecular characteristics for LCLC and LCNEC.

In this study, we aimed to investigate the whole-genome landscape of LCLC and LCNEC tumors in relation to other histological subgroups of lung cancer. Our results link recent lung cancer classification schemes to the genome-wide landscape of the disease, supporting that LCLC and LCNEC tumors follow distinct tumorigenic pathways. To our knowledge, this is the first genome-wide profiling comparison of LCLC and LCNEC.

Data availability statement

The data presented in the study are deposited in the Genome Sequence Archive repository, accession number HRA005093.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Tongji Hospital of Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XW: Designed the project, analysis, and wrote the article; JY and YD: Provided conceptual advice, software and analysis, performed the data curation, resources, and review. YZ: Contributed to resources, revised original draft and gave final approval of the version.

Funding

This work was supported by National Natural Science Foundation of China (81700174, 82102796).

Acknowledgments

We acknowledge the technical support of Shanghai Tongshu Biotechnology Co., Ltd.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1070048/full#supplementary-material

References

Asamura, H., Kameya, T., Matsuno, Y., Noguchi, M., Tada, H., Ishikawa, Y., et al. (2006). Neuroendocrine neoplasms of the lung: A prognostic spectrum. *J. Clin. Oncol.* 24, 70–76. doi:10.1200/JCO.2005.04.1202

Bi, L., Wang, H., and Tian, Y. (2021). Silencing FAM135B enhances radiosensitivity of esophageal carcinoma cell. *Gene* 772, 145358. doi:10.1016/j.gene.2020.145358

E. Brambilla, A. G. Nicholson, A. Burke, and W. D. Travis (Editors) (2015). *WHO classification of tumours of the lung, pleura, thymus and heart* (Lyon: IARC Press).

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

Brown, L. C., Tucker, M. D., Sedhom, R., Schwartz, E. B., Zhu, J., Kao, C., et al. (2021). LRP1B mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. *J. Immunother. Cancer* 9, e001792. doi:10.1136/jitc-2020-001792

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., et al. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222. doi:10.1093/bioinformatics/btv710

Chen, Y., Huang, Y., Kanwal, M., Li, G., Yang, J., Niu, H., et al. (2019). MUC16 in non-small cell lung cancer patients affected by familial lung cancer and indoor air pollution: Clinical characteristics and cell behaviors. *Transl. Lung Cancer Res.* 8, 476–488. doi:10.21037/tlcr.2019.07.10

Cheng, X., Yin, H., Fu, J., Chen, C., An, J., Guan, J., et al. (2019). Aggregate analysis based on TCGA: TTN missense mutation correlates with favorable prognosis in lung squamous cell carcinoma. *J. Cancer Res. Clin. Oncol.* 145, 1027–1035. doi:10.1007/s00432-019-02861-y

Cortes-Ciriano, I., Lee, J. J., Xi, R., Jain, D., Jung, Y. L., Yang, L., et al. (2020). Comprehensive analysis of chromothripsis in 2, 658 human cancers using whole-genome sequencing. *Nat. Genet.* 52, 331–341. doi:10.1038/s41588-019-0576-7

Derks, J. L., Leblay, N., Thunnissen, E., van Suylen, R. J., den Bakker, M., Groen, H. J. M., et al. (2018). Molecular subtypes of pulmonary large-cell neuroendocrine carcinoma predict chemotherapy treatment outcome. *Clin. Cancer Res.* 24, 33–42. doi:10.1158/1078-0432.CCR-17-1921

Dong, D., Zhang, W., Xiao, W., Wu, Q., Cao, Y., Gao, X., et al. (2021). A GRN autocrine-dependent FAM135B/AKT/mTOR feedforward loop promotes esophageal squamous cell carcinoma progression. *Cancer Res.* 81, 910–922. doi:10.1158/0008-5472.CAN-20-0912

Driver, B. R., Portier, B. P., Mody, D. R., Deavers, M., Bernicker, E. H., Kim, M. P., et al. (2016). Next-Generation sequencing of a cohort of pulmonary large cell carcinomas reclassified by World Health organization 2015 criteria. *Arch. Pathol. Lab. Med.* 140, 312–317. doi:10.5858/arpa.2015-0361-OA

Hofree, M., Shen, J. P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat. Methods* 10, 1108–1115. doi:10.1038/nmeth.2651

Howlader, N., Noone, A. M., Krapcho, M., Miller, D., Bishop, K., Altekruse, S. F., et al. (2015). *SEER cancer statistics review*. Bethesda (MD): National Cancer Institute (2016), 7-1.

Hu, J., Wang, Y., Zhang, Y., Yu, Y., Chen, H., Liu, K., et al. (2019). Comprehensive genomic profiling of small cell lung cancer in Chinese patients and the implications for therapeutic potential. *Cancer Med.* 8, 4338–4347. doi:10.1002/cam4.2199

Iyoda, A., Hiroshima, K., Moriya, Y., Mizobuchi, T., Otsuji, M., Sekine, Y., et al. (2004). Pulmonary large cell neuroendocrine carcinoma demonstrates high proliferative activity. *Ann. Thorac. Surg.* 77, 1891–1895. doi:10.1016/j.athoracsur.2003.10.119

Kanwal, M., Ding, X. J., Song, X., Zhou, G. B., and Cao, Y. (2018). MUC16 overexpression induced by gene mutations promotes lung cancer cell growth and invasion. *Oncotarget* 9, 12226–12239. doi:10.18632/oncotarget.24203

Karlsson, A., Brunnstrom, H., Lindquist, K. E., Jirstrom, K., Jonsson, M., Rosengren, F., et al. (2015). Mutational and gene fusion analyses of primary large cell and large cell neuroendocrine lung cancer. *Oncotarget* 6, 22028–22037. doi:10.18632/oncotarget.4314

Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Kallberg, M., et al. (2018). Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* 15, 591–594. doi:10.1038/s41592-018-0051-x

Kim, Y. A., Madan, S., and Przytycka, T. M. (2017). WeSME: Uncovering mutual exclusivity of cancer drivers and beyond. *Bioinformatics* 33, 814–821. doi:10.1093/bioinformatics/btw242

Lakshmanan, I., Salfity, S., Seshacharyulu, P., Rachagani, S., Thomas, A., Das, S., et al. (2017). MUC16 regulates TSPYL5 for lung cancer cell growth and chemoresistance by suppressing p53. *Clin. Cancer Res.* 23, 3906–3917. doi:10.1158/1078-0432.CCR-16-2530

Lan, S., Li, H., Liu, Y., Ma, L., Liu, X., Liu, Y., et al. (2019). Somatic mutation of LRP1B is associated with tumor mutational burden in patients with lung cancer. *Lung cancer* 132, 154–156. doi:10.1016/j.lungcan.2019.04.025

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, R41. doi:10.1186/gb-2011-12-4-r41

Pelosi, G., Fabbri, A., Papotti, M., Rossi, G., Cavazza, A., Righi, L., et al. (2015). Dissecting pulmonary large-cell carcinoma by targeted next generation sequencing of several cancer genes pushes genotypic-phenotypic correlations to emerge. *J. Thorac. Oncol.* 10, 1560–1569. doi:10.1097/JTO.0000000000000658

Qiu, Y., Liu, L., Yang, H., Chen, H., Deng, Q., Xiao, D., et al. (2020). Integrating histologic and genomic characteristics to predict tumor mutation burden of early-stage non-small-cell lung cancer. *Front. Oncol.* 10, 608989. doi:10.3389/fonc.2020.608989

Rekhtman, N., Tafe, L. J., Chaft, J. E., Wang, L., Arcila, M. E., Colanta, A., et al. (2013). Distinct profile of driver mutations and clinical features in immunomarker-defined subsets of pulmonary large-cell carcinoma. *Mod. pathology official J. U. S. Can. Acad. Pathology* 26, 511–522. doi:10.1038/modpathol.2012.195

Rossi, G., Mengoli, M. C., Cavazza, A., Nicoli, D., Barbareschi, M., Cantaloni, C., et al. (2014). Large cell carcinoma of the lung: Clinically oriented classification integrating immunohistochemistry and molecular biology. *Virchows Arch.* 464, 61–68. doi:10.1007/s00428-013-1501-6

Shi, Y., Chen, W., Li, C., Qi, S., Zhou, X., Zhang, Y., et al. (2020). Clinicopathological characteristics and prediction of cancer-specific survival in large cell lung cancer: A population-based study. *J. Thorac. Dis.* 12, 2261–2269. doi:10.21037/jtd.2020.04.24

Tian, W., Shan, B., Zhang, Y., Ren, Y., Liang, S., Zhao, J., et al. (2020). Association between DNA damage repair gene somatic mutations and immune-related gene expression in ovarian cancer. *Cancer Med.* 9, 2190–2200. doi:10.1002/cam4.2849

W. D. Travis, E. Brambilla, H. K. Müller-Hermelink, and C. C. Harris (Editors) (2004). "World Health organization classification of tumours," *Pathology and genetics of tumours of the lung, pleura, thymus and heart* (Lyon: IARC Press).

Wang, Y., Han, X., Wang, X., Sheng, W., Chen, Z., Shu, W., et al. (2019). Genomic based analyses reveal unique mutational profiling and identify prognostic biomarker for overall survival in Chinese small-cell lung cancer. *Jpn. J. Clin. Oncol.* 49, 1143–1150. doi:10.1093/jjco/hyz131

Wolf, Y., Bartok, O., Patkar, S., Eli, G. B., Cohen, S., Litchfield, K., et al. (2019). UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell.* 179, 219–235. e21. doi:10.1016/j.cell.2019.08.032

Xie, X., Tang, Y., Sheng, J., Shu, P., Zhu, X., Cai, X., et al. (2021). Titin mutation is associated with tumor mutation burden and promotes antitumor immunity in lung squamous cell carcinoma. *Front. Cell. Dev. Biol.* 9, 761758. doi:10.3389/fcell.2021.761758

Zhu, M., Zhang, L., Cui, H., Zhao, Q., Wang, H., Zhai, B., et al. (2021). Co-mutation of FAT3 and LRP1B in lung adenocarcinoma defines a unique subset correlated with the efficacy of immunotherapy. *Front. Immunol.* 12, 800951. doi:10.3389/fimmu.2021.800951

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 03 February 2023
DOI 10.3389/fgene.2023.1038924

OPEN ACCESS

EDITED BY
Shun Lu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Tianhang Li,
Nanjing University, China
Youtao Lu,
University of Pennsylvania, United States

*CORRESPONDENCE
Yuming Yu,
yuym72@163.com
Yanjun Liu,
yanjun@smu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 07 September 2022
ACCEPTED 20 January 2023
PUBLISHED 03 February 2023

CITATION
Tang Y, Ye C, Zeng J, Zhu P, Cheng S, Zeng W, Yang B, Liu Y and Yu Y (2023), Identification of a basement membrane-based risk scoring system for prognosis prediction and individualized therapy in clear cell renal cell carcinoma. *Front. Genet.* 14:1038924. doi: 10.3389/fgene.2023.1038924

Identification of a basement membrane-based risk scoring system for prognosis prediction and individualized therapy in clear cell renal cell carcinoma

Yanlin Tang[1,2†], Chujin Ye[2†], Jiayi Zeng[2,3], Ping Zhu[4], Shouyu Cheng[2,5], Weinan Zeng[1,2], Bowen Yang[2,6], Yanjun Liu[4*] and Yuming Yu[2,3*]

[1]Shantou University Medical College, Shantou, China, [2]Department of Urology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China, [3]The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China, [4]Department of Immunology, School of Basic Medical Science, Southern Medical University, Guangzhou, China, [5]School of Medicine, South China University of Technology, Guangzhou, China, [6]Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

Clear cell renal cell carcinoma (ccRCC) belongs to one of the 10 most frequently diagnosed cancers worldwide and has a poor prognosis at the advanced stage. Although multiple therapeutic agents have been proven to be curative in ccRCC, their clinical application was limited due to the lack of reliable biomarkers. Considering the important role of basement membrane (BM) in tumor metastasis and TME regulation, we investigated the expression of BM-related genes in ccRCC and identified prognostic BM genes through differentially expression analysis and univariate cox regression analysis. Then, BM-related ccRCC subtypes were recognized through consensus non-negative matrix factorization based on the prognostic BM genes and evaluated with regard to clinical and TME features. Next, utilizing the differentially expressed genes between the BM-related subtypes, a risk scoring system BMRS was established after serial analysis of univariate cox regression analysis, lasso regression analysis, and multivariate cox regression analysis. Time-dependent ROC curve revealed the satisfactory prognosis predictive capacity of BMRS with internal, and external validation. Multivariate analysis proved the independent predictive ability of BMRS and a BMRS-based nomogram was constructed for clinical application. Some featured mutants were discovered through genomic analysis of the BMRS risk groups. Meanwhile, the BMRS groups were found to have distinct immune scores, immune cell infiltration levels, and immune-related functions. Moreover, with the help of data from The Cancer Immunome Atlas (TCIA) and Genomics of Drug Sensitivity in Cancer (GDSC), the

Abbreviation: ccRCC, Clear cell renal cell carcinoma; ICIs, Immune checkpoint inhibitors; VEGF, Vascular endothelial growth factor; TKIs, Tyrosine kinase inhibitors; TME, Tumor microenvironment; BM, Basement membrane; ECM, Extracellular matrix; TCGA, The cancer genome atlas; TPM, Transcripts per million; CNV, Copy number variations; LASSO, Least absolute selection and shrinkage operator; BMRS, BM-related risk score system; CTA, Cancer testis antigens; HRD, Homologous recombination repair deficiency; ITH, Intratumoral heterogeneity; SNV, Single-nucleotide variation; TIDE, Tumor immune dysfunction and evasion; TMB, Tumor mutation burden; MDSC, Myeloid-derived suppressor cell; Treg, Regulatory T cell; WGCNA, Weight gene correlation network analysis; CNMF, Non-negative matrix factorization; ESTIMATE, Estimation of stromal and immune cells in malignant tumor tissues using expression data; ssGSEA, Single-sample gene set enrichment analysis.

potential of BMRS in predicting therapeutic response was evaluated and some possible therapeutic compounds were proposed through ConnectivityMap (CMap). For the practicability of BMRS, we validated the expression of BMRS-related genes in clinical samples. After all, we identified BM-related ccRCC subtypes with distinct clinical and TME features and constructed a risk scoring system for the prediction of prognosis, therapeutic responses, and potential therapeutic agents of ccRCC. As ccRCC systemic therapy continues to evolve, the risk scoring system BMRS we reported may assist in individualized medication administration.

KEYWORDS

clear cell renal cell carcinoma, basement membrane, gene expression, tumor microenvironment, individualized therapy

1 Introduction

Globally, more than 430,000 individuals suffering from kidney cancers were newly diagnosed in 2020 and approximately 180,000 people died from this type of cancer (Sung et al., 2021). As one of the major subtypes, clear cell renal cell carcinoma (ccRCC) could earn a good prognosis when treated at an early stage. However, around one-third of ccRCC patients were found to be in the metastatic stage, requiring systemic therapy other than radical surgery (Jonasch et al., 2014). Although novel treatments including immunotherapies and targeted therapies were demonstrated to be curative in this chemoresistant cancer type, their clinical effects were uncontrollable due to the lack of predictive biomarkers for the therapeutic response and adverse events (Jhaveri and Perazella, 2018; Motzer et al., 2020a). Besides, it was demonstrated that combined therapy of immune checkpoint inhibitors (ICIs) and vascular endothelial growth factor (VEGF) tyrosine kinase inhibitors (TKIs) could exert better curative effects than monotherapy, leading to a requirement for more individualized markers for treatment selection (Amin and Hammers, 2018). Therefore, reliable predictive biomarkers ought to be developed for the prognosis and therapeutic response of ccRCC.

The tumor microenvironment (TME) is a complex ecosystem including immune cells, stromal cells, and extracellular matrix, surrounding and interacting with tumor cells (Blankenstein et al., 2012). As a highly immune-infiltrated cancer type, ccRCC cells were able to modulate the TME including immune cells for evasion of anti-cancer immunity through multiple mechanisms (Díaz-Montero et al., 2020). The understanding of these mechanisms could facilitate the application of cancer-specific therapies, such as ICIs, which restored anti-cancer immunity through interrupting the suppressive signals from the ccRCC cells (Motzer et al., 2015). Besides, accumulating evidence indicated that molecular classification of ccRCC into groups with distinct TME features could distinguish their prognosis and therapeutic response (de Velasco et al., 2017). Thus, it would be valuable to investigate the TME in ccRCC for the discovery of novel predictive biomarkers.

Basement membrane (BM) is a thin sheet of extracellular matrix (ECM) lining beneath endothelial and epithelial tissues, mainly composed of collagen IV and laminin (Yurchenco, 2011). It serves as one of the barriers preventing cancer cells from invasion, but its remodeling and stiffness would contribute to the metastasis of tumor (Chang and Chaudhuri, 2019). Some BM-related genes were revealed to be associated with the prognosis of RCC. Wragg et al. (2016) demonstrated that the high expression of LAMA4, a laminin component, was correlated with the poor prognosis of RCC (Wragg et al., 2016; Ho et al., 2017). Moreover, BM could mediate the signal transduction between the microenvironment and cells. As a major component of BM, laminin was revealed to have the ability to modulate the migration, activation and functionality of T lymphocytes within tumors (Liu et al., 2022). With these concerns, investigating the BM-related genes in ccRCC may assist in understanding the relationship between ccRCC and TME and developing predictive biomarkers.

In the current study (Figure 1), we investigated the expression of BM-related genes in ccRCC and used them to classify ccRCC patients into distinct subtypes, based on which a risk scoring system, BMRS, was established. Comprehensive analyses were conducted to evaluate the capacity of BMRS in distinguishing the prognosis, TME features, and therapeutic response of ccRCC. In this way, we constructed a gene-based BMRS for prognosis and treatment prediction of ccRCC and provided molecular candidates as novel therapeutic targets.

2 Materials and methods

2.1 Data acquisition

A ccRCC cohort including 539 tumor samples and 72 normal samples from the KIRC project of The Cancer Genome Atlas (TCGA) was selected and its RNA expression data, somatic mutation data, and the corresponding clinical data were extracted from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov). ArrayExpress (https://www.ebi.ac.uk/arrayexpress) is another public database containing high-throughput genomic data from more than 75,000 experiments. Gene expression data of a ccRCC cohort, E-MTAB-1980, with 101 tumor samples were acquired from ArrayExpress and the survival information was obtained from previous research (Li et al., 2018). The expression data of both cohorts were transformed into a data format of transcripts per million (TPM) for better analysis. All the data were publicly available and no ethical consent was required.

2.2 Investigation of the basement membrane genes in ccRCC

Jayadev et al. discovered 222 protein-coding genes that were related to BM and human health (Jayadev et al., 2022). The expression of these genes was extracted from the TCGA cohort

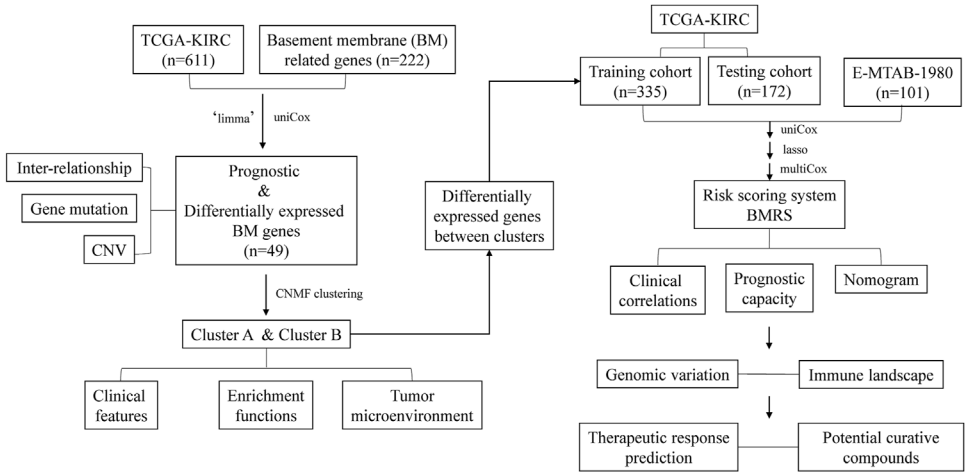


FIGURE 1
The flow chart of the current study.

and differentially analyzed between tumor and normal samples with the help of the R package "limma" to identify the differentially expressed BM genes (Ritchie et al., 2015). Thereafter, univariate cox regression analysis facilitated the selection of the prognostic BM genes from those differentially expressed genes. For a better understanding of the expression of the prognostic BM genes in ccRCC, correlation analyses were performed to reveal their interrelationship. Besides, utilizing the R package "maftool" (Mayakonda et al., 2018), the variation in these genes was depicted including both somatic mutation status and copy number variations (CNV).

2.3 Discovery and investigation of basement membrane-related clusters in ccRCC

Consensus non-negative matrix factorization (CNMF) is a powerful method for the dimension reduction of genomic data to discover distinguished molecular patterns. Through the R package 'CancerSubtypes' (Xu et al., 2017), the expression data of the prognostic BM genes were used to classify the ccRCC samples in the TCGA cohort into 2, 3, and 4 clusters. Silhouette width is a measurement for the evaluation of the classifications and a larger average silhouette width means more valuable subtypes. Clusters with the highest average silhouette width were selected for the following analyses. To further investigate the value of these BM-related clusters, Kaplan-Meier survival analysis and chi-square test were conducted to visualize the clinical difference between them. Meanwhile, though GSEA 4.1.0, gene sets enrichment analysis (GSEA) together with gene ontology (GO) gene sets assisted the understanding of the molecular functions enriched in each cluster. Furthermore, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) (Yoshihara et al., 2013) was utilized to quantify the immune and stromal status of ccRCC samples and the resulting scores were compared between each cluster to discover whether there were differences in their TME. Besides, we calculated the infiltrating levels of 23 tumor infiltrating immune cells (TIIC) through single sample GSEA (ssGSEA) and 14 stoma cells by xCell (https://xcell.ucsf.edu/). By comparing the infiltrating level of these cells between different clusters, their TME differences could be better understood.

2.4 Construction of a gene expression-based risk score system

After discovering the distinct features between the BM-related subtypes, we tended to construct a risk score system based on these clusters. First, the differentially expressed genes between the clusters were identified. Then, the TCGA cohort was randomly divided into a training cohort and a testing cohort in a ratio of 7 to 3. In the training cohort, expression of the differentially expressed genes was extracted and used for univariate cox regression analysis to select the prognostic genes. Through R package 'glmnet' (Friedman et al., 2010), Least Absolute Selection and Shrinkage Operator (LASSO) regression analysis facilitated identifying those genes with higher association with ccRCC prognosis, and multivariate cox regression analysis was utilized to choose the best genes for construction of the BM-related risk score system (BMRS): risk score = coefficient 1*gene expression 1 + . . . + coefficient n * gene expression n (1 to n represent each prognostic genes).

2.5 Analysis of the prognostic predictive capacity of BMRS

To investigate the predictive ability of BMRS, the TCGA training cohort, TCGA testing cohort, TCGA cohort, and E-MTAB-1980 cohort were respectively divided into high and low risk groups according to the median risk scores. Kaplan-Meier analysis was applied in each pair of high and low risk groups to reveal the relationship between risk score and the overall survival (OS) of ccRCC patients. Subsequently, with the help of time-dependent Receiver Operating Characteristic (tROC) curves, the predictive power of the risk scoring system for the 1-, 3-, and 5-year OS of each cohort were illustrated and the corresponding values of area under curve

(AUC) were calculated. Additionally, the clinical features (age, gender, Fuhrmann grade, and AJCC stage) in high and low risk groups of the TCGA cohort were compared through chi-square test. Meanwhile, the ability of BMRS in differentiating ccRCC OS was tested under different clinical statuses by Kaplan-Meier survival analysis.

2.6 Establishment of a clinical predictive nomogram

For better application in the clinic, BMRS (high risk *vs*. low risk) combined with age (<65-year-old *vs*. >=65-year-old), gender (female *vs*. male), Fuhrmann grade (Grade 1/2 *vs*. Grade 3/4), AJCC stage (Stage I/II *vs*. Stage III/IV), T stage (T1/2 *vs*. T3/4), N stage (N0 *vs*. N1), and M stage (M0 *vs*. M1) were incorporated into univariate and multivariate cox regression analyses to demonstrate its predictive value. Furthermore, the resulting values of the multivariate cox regression analysis were used to construct a nomogram for the prediction of ccRCC survival. The predictive capability of the nomogram for 3-and 5-year survival was assessed by calibration plot. Besides, decision curve analysis (DCA) was utilized to compare the net benefit of BMRS, AJCC stage and two previous BM-related gene signatures (Zhou et al., 2022a; Xiong et al., 2022) in predicting 3- and 5-year survival of ccRCC. TCGA cohort was separated into high and low risk groups according to the median of nomogram scores and was analyzed to identify their survival difference.

2.7 Investigation of the genomic variation in different BMRS groups

The genomic features in high and low BMRS groups were illustrated and compared through the R package 'maftool'. The variation status of the top 20 mutation genes in both groups was depicted and all the mutated genes were compared between groups to identify the group-specific mutation. At the same time, by applying pairwise Fisher's exact test between every two genes, we wanted to discover whether there were some exclusive or co-occurrence gene pairs in the high BMRS group. Furthermore, some tumor heterogeneity-related features, including single nucleotide variation (SNV), homologous recombination defects (HRD), cancer testis antigen (CTA), and intratumor heterogeneity (ITH) were introduced from previous research (Thorsson et al., 2018) and compared between different BMRS groups.

2.8 Analysis of the immune landscape related to BMRS

In order to investigate the BMRS-related biological processes, cellular components, and molecular functions, the genes that were differentially upregulated in the high BMRS group were chosen and incorporated into GO functional analysis through the R package 'clusterProfiler' (Wu et al., 2021). Concerning the immune microenvironment in ccRCC, the immune scores calculated from ESTIMATE and the infiltrating scores of 23 TIICs were compared between BMRS groups. Each TIIC was analyzed through Kaplan-Meier survival analysis to discover its relationship with ccRCC prognosis. Meanwhile, ssGSEA algorithm was utilized to induce scores representing immune suppression and some immune-related functions with the help of previously published gene signatures (Yi et al., 2020). In addition, a publicly accessible website called Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu) which provides a platform for estimation of T cells dysfunction scores based on gene expression data, was adopted to compare the T cells status between high and low BMRS clusters.

2.9 Exploration of the therapeutic predictive potential of BMRS

The gene expression of several popular inhibitory immune checkpoints was extracted and differentially analyzed between high and low BMRS groups. Additionally, the immunophenoscores, which are estimated scores of immunotherapeutic response to ICIs for TCGA samples, were obtained from The Cancer Immunome Atlas (TCIA, https://tcia.at/home) and compared between BMRS groups. Except for immunotherapeutic response, sensitivity toward some therapeutic drugs (Axitinib, Pazopanib, Sorafenib, Sunitinib, and Mitomycin C) clinically used for ccRCC patients were predicted through R package 'pRRophetic' (Barbour et al., 2014), which connects a large amount of gene expression and drug sensitivity data in Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org). Furthermore, the gene expression of immune chemokines, immune receptors, and major histocompatibility complex (MHC) molecules was correlated with BMRS with the concern of discovering potential therapeutic targets. Besides, we uploaded the top 150 upregulated genes in high and low BMRS groups respectively to ConnectivityMap (CMap, https://clue.io) for exploration of potential therapeutic compounds for patients in each group.

2.10 Transcriptome sequencing analysis

To further prove the applicability of BMRS, we collected clinical ccRCC and adjacent normal samples (18 ccRCC and 6 adjacent normal samples) for transcriptome sequencing analysis. This project was supported by the hospital ethics committee and consent was acquired from all the patients. In line with the protocol of the manufacturer, each ccRCC and adjacent normal samples underwent paired-end sequencing on the NovaSeq 6000 high-throughput sequencing platform (Illumina, United States) to remove sequencing reads containing aptamer sequences and low-quality reads as well as bases. Then, high-quality pairwise reads were aligned to the human genome GRCh38 through HISAT2 (v2.1.1), generating BAM files. BAM files were arranged by samtools (v1.15.1) and then counted with the help of Subread (v2.0.1). Raw counts of transcripts per gene were converted to the format of TPM, allowing better analysis of gene expression between samples. Thereafter, the genes enrolled in BMRS were differentially analyzed between the normal and ccRCC samples.

2.11 Statistical analysis

All the analyses process in the current study were achieved through the usage of R 4.1.0 and R studio Desktop 2022.07.1 + 554. The graphs displayed were drawn by R studio Desktop

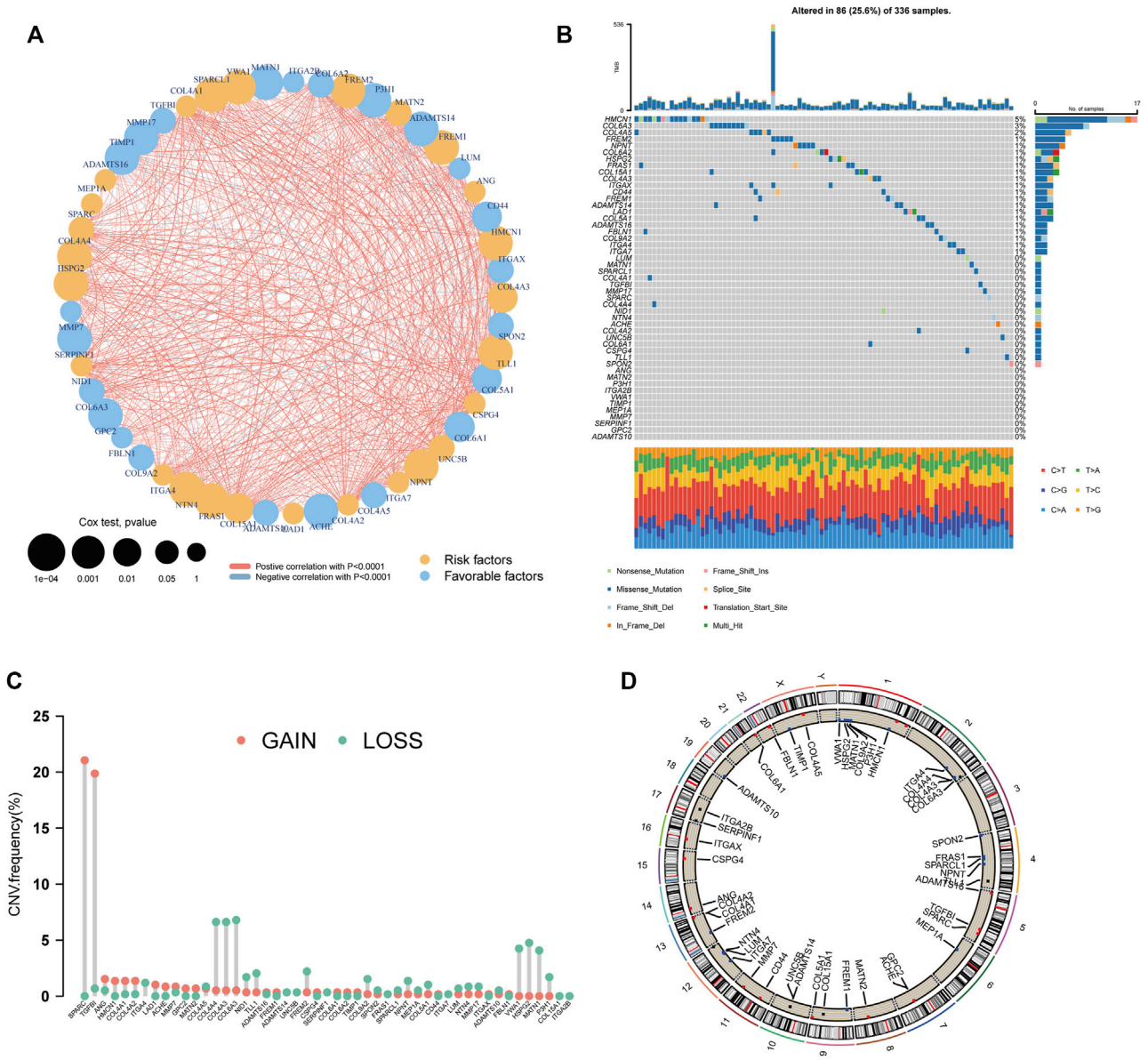


FIGURE 2
Genetic analyses of the BM-related genes. **(A)** The inter-relationship between each BM-related gene and the role of these genes in the prognosis of ccRCC. **(B)** The waterfall plot depicting the mutation of the prognostic BM genes. **(C)** The CNV frequency of the prognostic BM genes. **(D)** The distribution of the prognostic BM genes in each chromosome with CNV gain colored red, CNV loss colored blue and no CNV colored black.

2022.07.1 + 554 and Adobe Illustrator CS6 (64 Bit). During differentially expression analysis, genes with the absolute value of log fold change (logFC) more than were selected. *p*-value less than 0.05 was regarded as significant for all analyses.

3 Results

3.1 Identification of the differentially expressed and prognostic BM genes

Differential expression analysis of the 222 BM genes between 539 ccRCC samples and 72 normal samples revealed that there were 106 differentially expressed BM genes, in which 39 genes were downregulated and 67 genes were upregulated (Supplementary Table S1). Then, after univariate cox regression analysis, 49 BM genes were demonstrated to be prognostic including 26 protective genes and 23 risk genes (Supplementary Table S2). A substantial positive correlation existed between these prognostic BM genes indicating that they were highly interconnected (Figure 2A). As for the genomic variation that happened in these genes, 25.6% of the ccRCC samples possessed prognostic BM gene alterations and more of these alterations were missense mutations (Figure 2B). The top 3 mutated genes were HMCN1 (5%, risk gene), COL6A3 (3%, protective gene), and COL4A5 (2%, protective gene). In addition, the CNV of these genes was analyzed and a relatively low frequency of CNV was discovered, except for SPARC (21.0%, gain of function), and TGFBI (19.8%, gain of function) (Figures 2C, D).

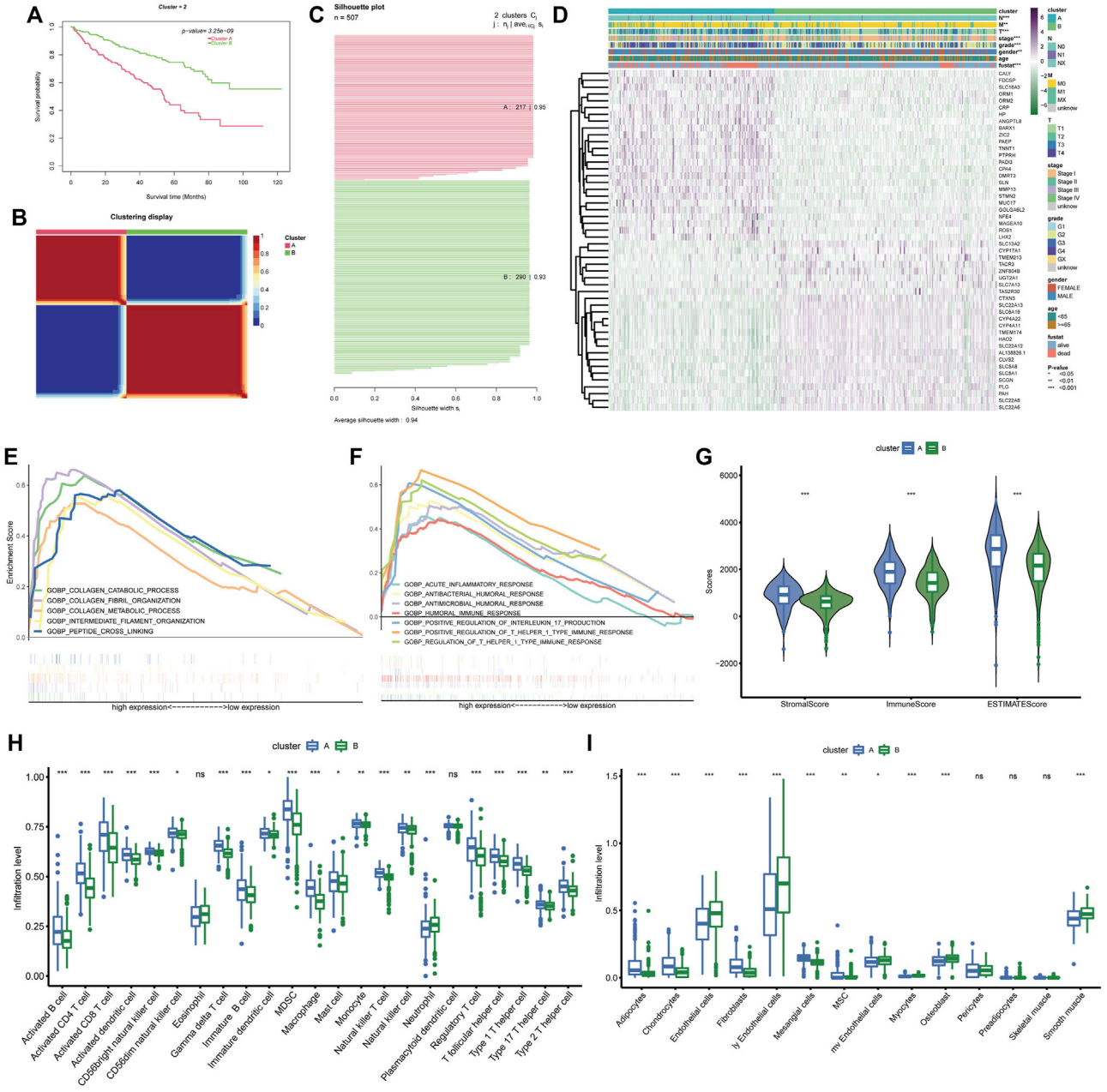


FIGURE 3
BM gene-based classification of ccRCC patients. **(A)** Survival analysis of the two ccRCC clusters classified based on BM genes. **(B)** The sample similarity matrix plot of the two identified clusters. **(C)** The Silhouette width plot of the two BM-related subtypes. **(D)** The heatmap depicting the distinct clinical features and gene expression between cluster **(A)** and cluster **(B)**. **(E, F)**. The biological processes that enriched in cluster **(A)**. **(G)** Comparison of the stromal score, immune score and ESTIMATE score between cluster **(A)** and cluster **(B)**. **(H, I)** The different infiltration levels of immune cells and stromal cells between the two BM-related subtypes.

3.2 Recognition of BM-based clusters of ccRCC

The 49 differentially expressed and prognostic BM genes were further used for the classification of ccRCC patients through CNMF algorithm. As the results show, the samples could be classified into two separate clusters A and B, in which cluster A had significantly lower OS than cluster B and the average silhouette width was 0.94 (Figures 3A–C). For the reason that the average silhouette widths were relatively low, classifications of 3, 4, and 5 clusters were not under consideration (Supplementary Figure S1). By comparing the clinical features in both clusters, we discovered that the ccRCC samples in cluster A had more advanced features including Fuhrmann grade, AJCC stage, T stage, N stage, and M stage. There were more males in cluster A than in cluster B while the age distributions in both clusters were similar (Figure 3D). The differentially expressed analysis also indicated that some metastasis-related genes, such as MMP13 and ROS1 were upregulated in cluster A. The following GSEA analysis demonstrated that basement membrane-related functions like collagen fibril organization and collagen catabolic process were

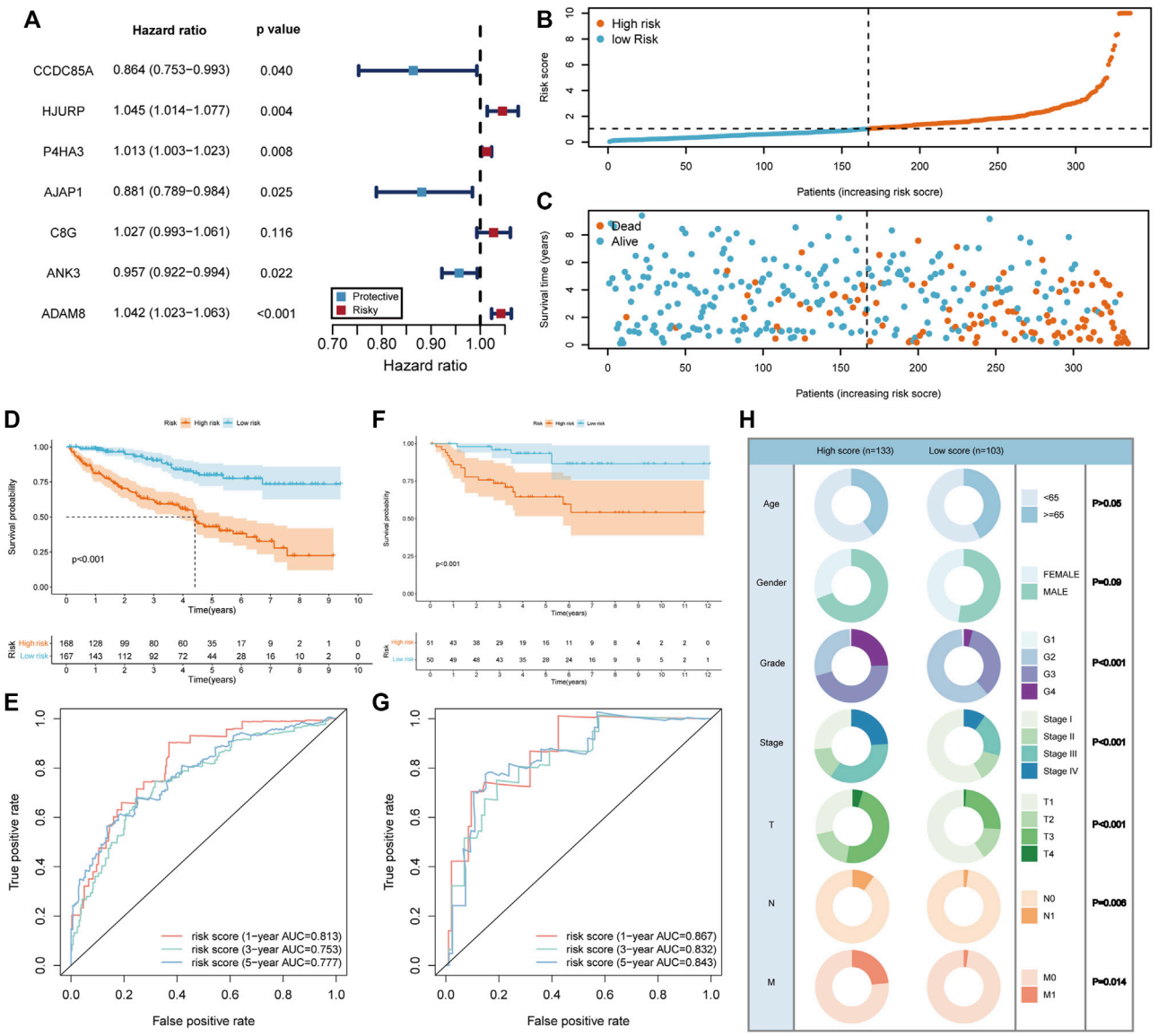


FIGURE 4
Construction and evaluation of the risk scoring system BMRS. **(A)** The forest plot illustrating the 7 genes resulted from multivariate cox regression analysis. **(B, C)** Distribution of the risk scores and survival status of ccRCC patients in the TCGA training cohort. **(D, F)** The Kaplan Meier survival curve comparing the overall survival between high and low risk groups in the TCGA training cohort and E-MTAB-1980 cohort respectively. **(E, G)** The tROC curve of the risk score for 1-, 3-, and 5-year survival in the TCGA training cohort and E-MTAB-1980 cohort. **(H)** Comparison of the clinical features in high and low risk groups.

more prominent in cluster A. Meanwhile, some immune-related biological processes were notably enriched in cluster A, such as positive regulation of T helper 1 type immune response and positive regulation of interleukin 17 production (Figures 3E, F). Considering these enriched immune functions, we further investigated the microenvironment components in both clusters through ESTIMATE and ssGSEA. It was demonstrated that the stromal score, immune score and ESTIMATE score in cluster A were higher than those in cluster B, indicating the rich stromal, immune components and low tumor purity in cluster A (Figure 3G). At the same time, cluster A had a higher level of nearly every type of infiltrating immune cell than cluster B except for eosinophils, neutrophils, and plasmacytoid dendritic cell (Figure 3H). As for stromal cells, more adipocytes, chondrocytes, fibroblasts, and mesangial cells were found in cluster A while cluster B possessed more endothelial cells, lymphatic endothelial cells and microvascular endothelial cells (Figure 3I). Therefore, the prognostic BM genes could divide ccRCC samples into two clusters with distinct clinical and TME features.

3.3 Construction of BMRS risk scoring system for prognosis prediction

Since the BM gene-based ccRCC subtypes could discriminate prognostic and clinical features, we would like to derive a more applicable risk scoring system from these subtypes. In the TCGA training cohort, the differentially expressed genes between clusters A

and B were identified (Supplementary Figure S2A). Based on these genes, univariate cox regression analysis was conducted and induced 414 prognostic genes (Supplementary Table S3). Subsequently, LASSO algorithm was applied to obtain 12 genes that were significantly associated with the prognosis of ccRCC (Supplementary Figure S2B, C). Then, these 12 genes were incorporated into multivariate cox regression analysis (Figure 4A), inducing a risk scoring system, BMRS, based on 7 genes, CCDC85A, AJAP1, ANK3, P4HA3, C8G, ADAM8, HJURP.

3.4 Evaluation of the predictive ability of BMRS

For the evaluation of BMRS system, we calculated risk scores for each sample in TCGA training cohort and divided them into high and low risk groups according to the median risk scores (Figure 4B). It could be recognized that samples with high risk were more frequently dead than those with low risk (Figure 4C). At the same time, survival analysis demonstrated that the high risk group had longer OS than the low risk group (Figure 4D). The results of tROC curve proved that BMRS system possessed a great prognostic capacity for 1-year (AUC = 0.813), 3-year (AUC = 0.753), and 5-year (AUC = 0.777) survival (Figure 4E). The same analyses were applied in the TCGA testing cohort and TCGA cohort as an internal validation and delivered similar results (Supplementary Figure S2D–K). As an external validation, the survival analysis result and tROC curve derived from E-MTAB-1980 also supported that BMRS could discriminate the prognosis of ccRCC and had high accuracy (Figures 4F, G, 1-year AUC = 0.867, 3-year AUC = 0.832, and 5-year AUC = 0.843). Furthermore, survival analyses of high and low risk groups in the TCGA cohort under various clinical situations were conducted and revealed that BMRS could discriminate ccRCC survival in most situations (Supplementary Figure S3). Meanwhile, the Fuhrmann grade, AJCC stage, T stage, N stage, and M stage in the high risk group were notably higher than those in the low risk group, which was in line with the poor prognosis in the high risk group (Figure 4H). Moreover, we evaluated the gene expression of the genes used for BMRS and the results demonstrated that most of the genes were differentially expressed in ccRCC (Supplementary Table S4), consistent with the above analysis. In this way, a reliable risk scoring system BMRS was constructed with considerable predictive capacity.

3.5 Establishment of a clinical nomogram based on BMRS

For better application of the BMRS in clinic, we evaluated its predictive ability with the consideration of the clinical variables. Analyzing risk score and clinical variables cooperatively, univariate cox regression analysis revealed that only age and gender could not predict ccRCC prognosis (Figure 5A). The following multivariate cox regression analysis uncovered the independent prognostic predictive ability of BMRS, T stage and M stage (Figure 5B). Thereafter, the Fuhrmann grade, T stage, M stage and BMRS risk were incorporated to establish a nomogram (Figure 5C). The calibration curve depicted the 3-year and 5-year OS predicted by the nomogram had a satisfactory consistency with those observed in clinic (Figure 5D). While compared with the clinically popular TNM stage, Zhou's and Xi's gene signatures, the nomogram could deliver higher net benefit in 3-year (Figure 5E) and 5-year OS prediction (Figure 5F). The survival analysis also demonstrated that samples with high points had significantly lower OS than those with low points (Figure 5G). Thus, BMRS could not only independently predict ccRCC survival but also assist in establishing more competitive predictive methods.

3.6 The genomic variation in different risk groups

After discovering the clinical significance of BMRS, further analyses of the underlying genomic alterations were under concern. In both risk groups, the mutation of VHL (44%), PBRM1 (38%), TTN (13%), and SETD2 (11%) comprised the major part of all gene alterations (Figure 6A). The following comparison revealed that 10 mutants were occurring more frequently in the high risk group (Figure 6B). Among these mutants, SETD2 was the most frequent and significant one and most of its mutation happened in the non-coding area (Figure 6C). Besides, the discovery of the relationship between gene variations uncovered a notable co-occurance of SETD2 and PBRM1 mutation. At the same time, the mutation of MUC16 and BAP1 also exhibited a significant positive inter-relationship while MUC17 and VHL mutation were mutually exclusive (Figure 6D). Apart from these mutational differences, the high risk group was demonstrated to have higher scores of SNV antigens and HRD, indicating the high genomic heterogeneity in patients with high BMRS risk (Figures 6E, F). Meanwhile, the CTA and ITH scores in the high risk group were significantly higher than those in the low risk group (Figures 6G, H).

3.7 The association between BMRS and the immune landscape of ccRCC

Considering that BMRS was derived from the BM-related clusters with distinct functional and immunal features, we also investigated the functions associated with BMRS. Similar to the above results, the high risk group was enriched with BM-related functions, such as extracellular structure organization and extracellular matrix organization. Meanwhile, there were some immune involvements recognized in the high risk group including humoral immune response and acute inflammatory response (Figure 7A). The comparison of immune scores between the two risk groups also indicated that the high risk group had a higher immune level than the low risk group (Figure 7B). For a better understanding of this immune involvement, we further evaluated the infiltrating immune cells in each group. As the results showed, most of the immune cells were more frequently infiltrated in the high risk group than in the low risk group except for eosinophil and neutrophil whose infiltrating levels were lower in the high risk group (Figure 7C). Survival analyses were conducted for each immune cell and it was revealed that some innate immune cells like neutrophil, and mast cell were protective cells. Most of the immune cells enriched in the high risk group were associated with poor survival, including suppressive immune cells (e.g., MDSC) and effector immune cells (e.g., activated CD8 T cell). For an exploration of this connection between abundant immune infiltration and poor survival, the functionality of the immune environment was assessed, delivering a result that ccRCC samples with high risk scored higher on immune suppression and T cell

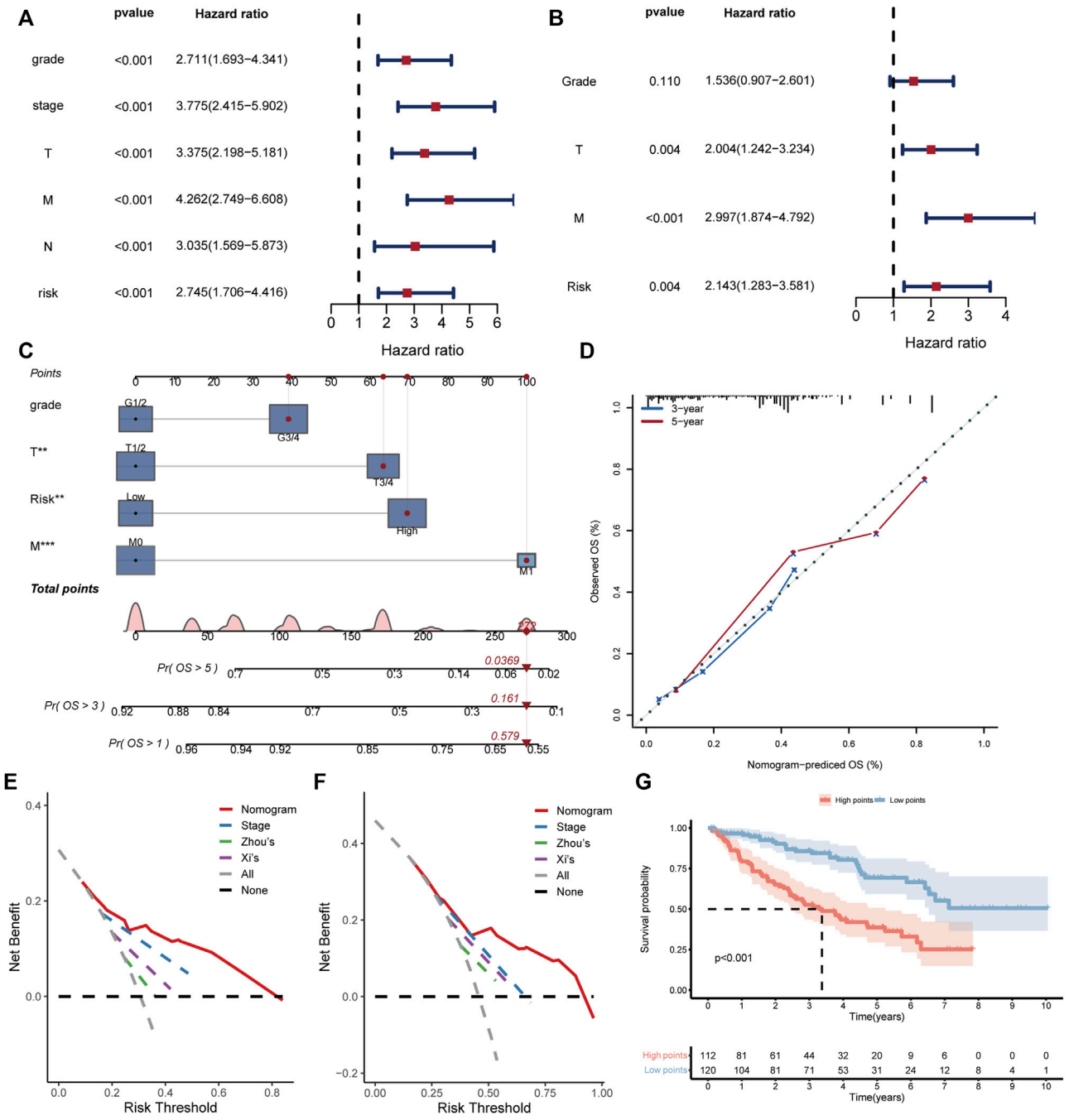


FIGURE 5
Establishment and evaluation of the clinical predictive nomogram. **(A, B)** The forest plots depicting the results of univariate cox regression analysis and multivariate cox regression analysis respectively. **(C)** The nomogram based on grade, T stage, M stage, and risk score. The red dot lines represented an example of the total point calculation. **(D)** The calibration curve testing predictability of the nomogram for 3- and 5-year survival. **(E, F)** The DCA curves comparing the net benefit of nomogram, TNM stage, Zhou's and Xi's gene signatures for prediction of 3- and 5-year survival respectively. **(G)** The Kaplan Meier survival curve comparing the overall survival of ccRCC patients with high and low points.

dysfunction than those with low risk (Figures 7D, E). Moreover, although high-risk samples had a higher level of most immune functions like APC co-stimulation and T cell co-stimulation, their scores of some negative immune functions such as APC co-inhibition and T cell co-inhibition were also greater than low-risk samples (Figure 7F). Therefore, BMRS may have a relationship with the suppressive immune microenvironment of ccRCC.

3.8 The therapeutic predictive potential of BMRS

After the investigation of BMRS and the immune microenvironment, the potential relationship between BMRS and immunotherapies was under consideration. Differential analysis of the gene expression of inhibitory immune checkpoints showed that

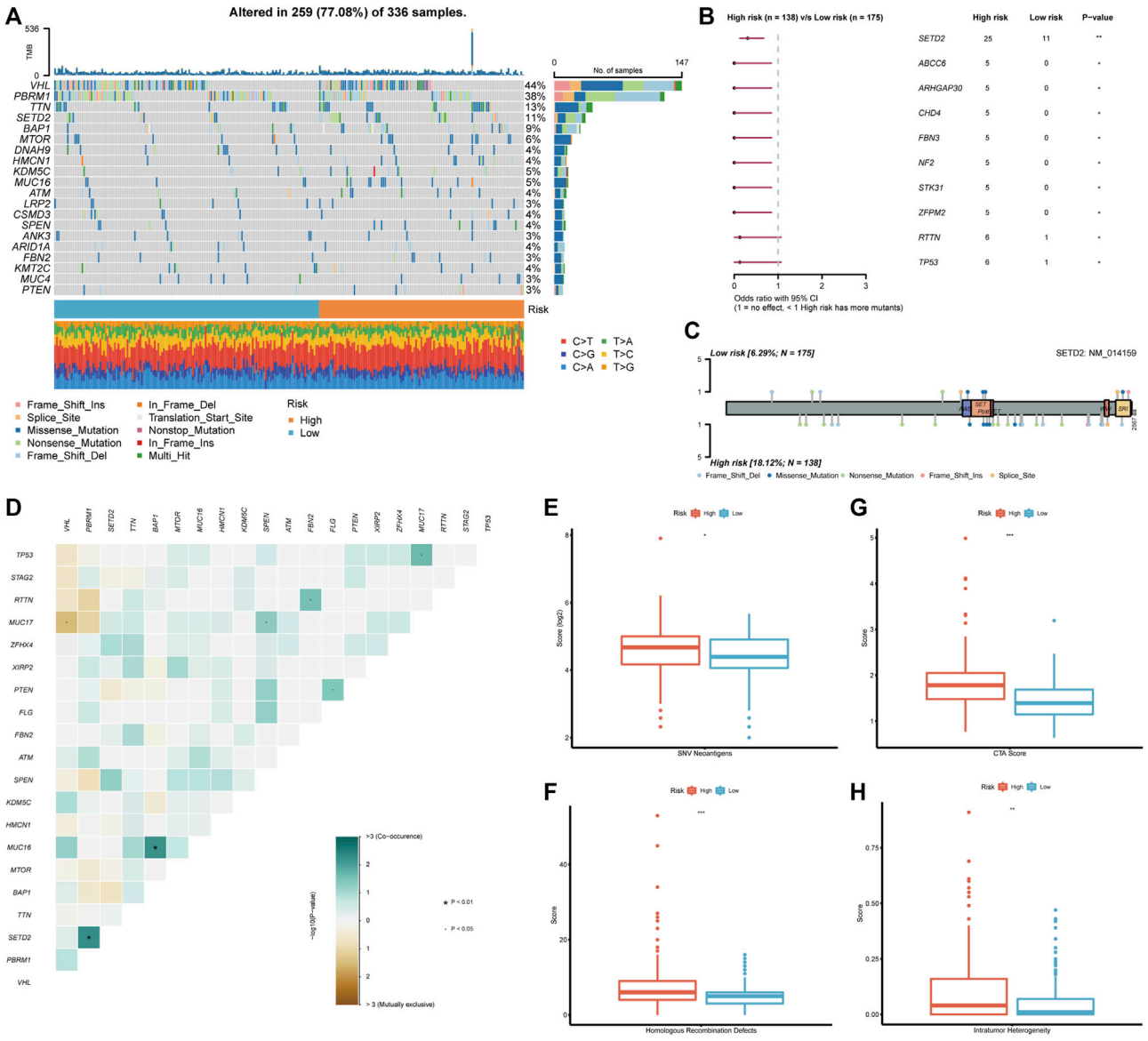


FIGURE 6
Genomic analysis in accordance with the BMRS risk groups. **(A)** The waterfall plot representing the top 20 mutated genes in both risk groups. **(B)** The mutants that significantly differentiated between high and low risk groups. **(C)** Exhibition of the mutation sites and types of SETD2 in both risk groups. **(D)** The correlation of the mutants in the high risk group. **(E–H)** Comparison of SNV neoantigens, HRD, CTA score, and ITH between high and low risk groups.

PD1, CTLA4, LAG3, and TIGIT were significantly upregulated in the high risk group (Figure 8A). Data on immunotherapeutic response predicted by TCIA demonstrated that high-risk ccRCC patients possessed a higher level of IPS to CTLA4 inhibitor and combination of PD1 and CTLA4 inhibitors than low-risk patients, indicating high-risk patients may respond better toward these two strategies of immunotherapies (Figure 8B). Apart from immunotherapies, we also correlated BMRS with some therapeutic drugs used in clinic. The half maximal inhibitory concentration (IC50) of sorafenib, sunitinib, and mitomycin C. were notably lower in the high risk group than in low risk group, suggesting that ccRCC patients with high risk may have better outcomes receiving these drugs (Figure 8C).

To discover some additional therapeutic targets, we analyzed the correlation between BMRS and some immune-related molecules (Figure 8D). There were multiple immune checkpoints (e.g., CXCL13, and CCL25), immune receptors (e.g., CXCR5, and CCR6) and MHC molecules (e.g., HLA-E) significantly related to BMRS, providing potential targets for therapeutic investigation. Moreover, as predicted by CMap (Figure 8E), molecular compounds like BRD-K23875128, which is Rho associated kinase inhibitor, could serve as therapeutic drugs for high-risk ccRCC patients. Some other compounds were also estimated to be effective in low-risk ccRCC patients, for example, PF-04217903, which is a kind of arginase inhibitor.

4 Discussion

With the progression of ccRCC research, more treatment options were available to improve the prognosis of ccRCC patients (Choueiri

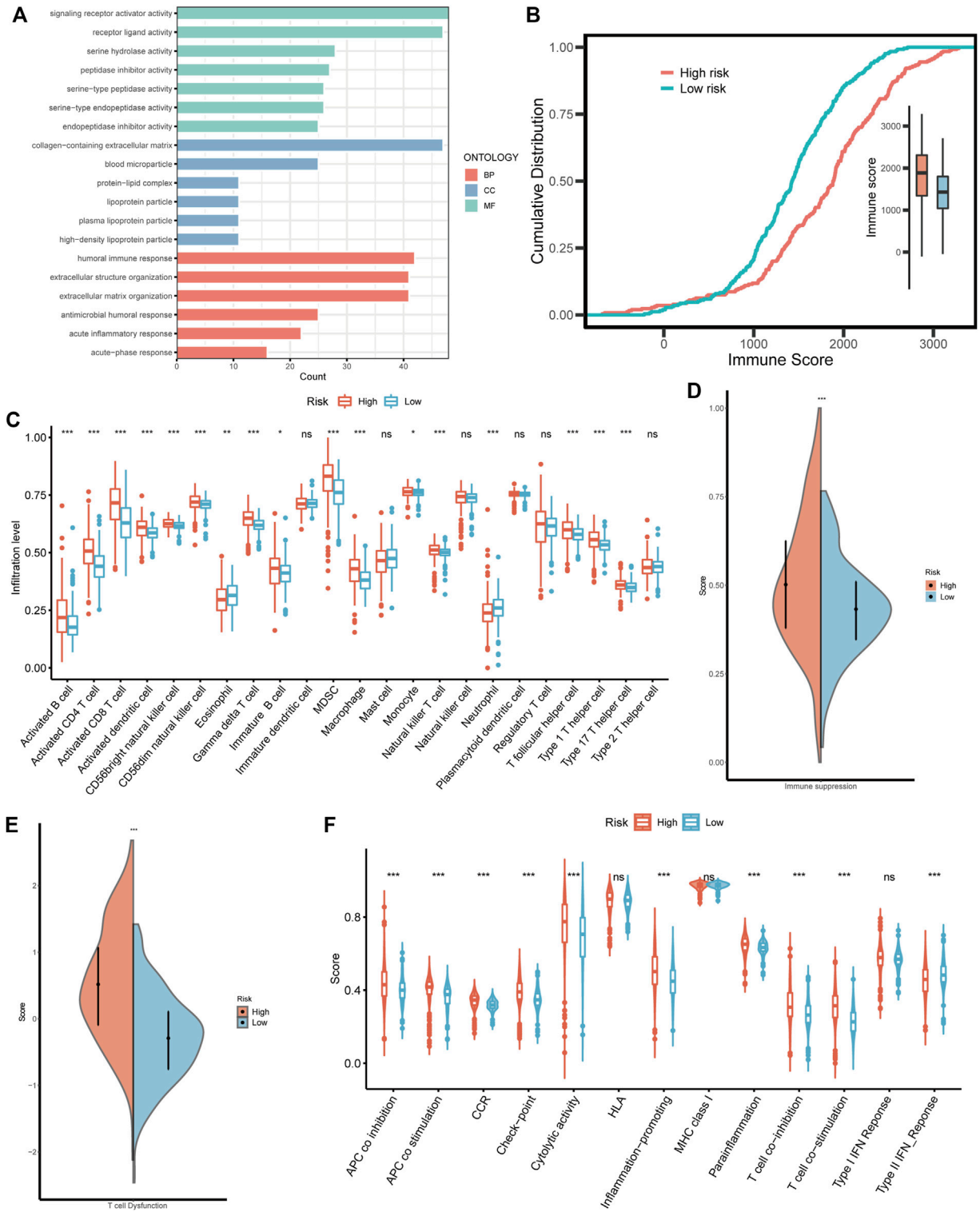


FIGURE 7
The immune landscape in different BMRS risk groups. **(A)** The enriched GO functions in the high risk group. **(B)** The different cumulative distributions of the immune score in different risk groups. **(C)** The distinct infiltration level of the immune cells in high and low risk groups. **(D, E)** Comparison of the scores of immune suppression and T cell dysfunction between the two BMRS groups. **(F)** The difference in immune functions between high and low risk groups.

and Motzer, 2017). However, the existing predictive markers for the prognosis and therapeutic response of ccRCC had limited functions due to the heterogeneity of this type of patient (Greef and Eisen, 2016). As a component of ECM, BM serves as not only a physical barrier against tumor invasion and metastasis but also a mediator of signals between microenvironment and cells (Bissell and Hines, 2011), indicating its potential as a target for the investigation of novel biomarkers. Through analyses of the expression of BM-related

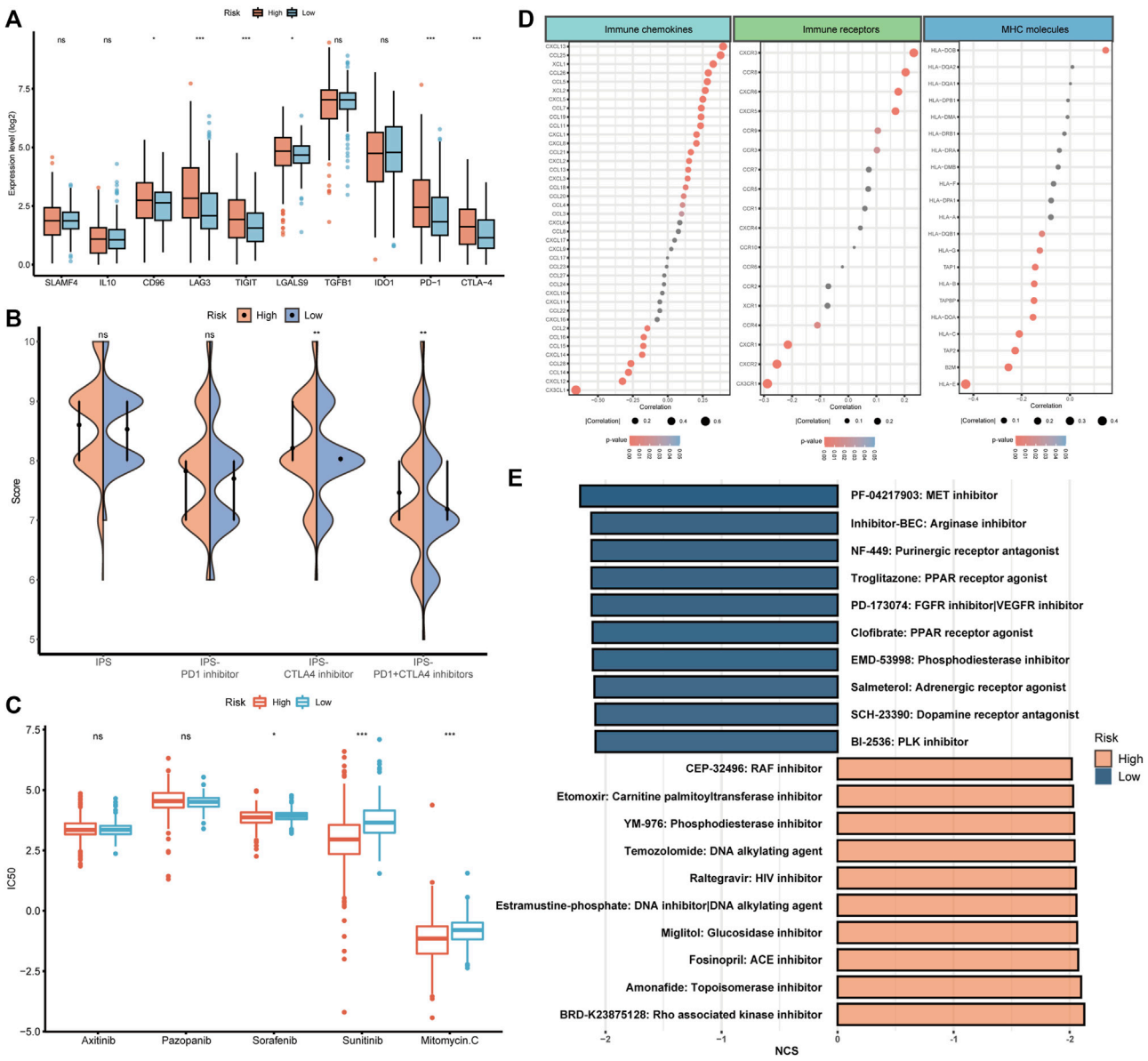


FIGURE 8
The potential of BMRS in guiding ccRCC treatment. **(A)** Comparison of the inhibitory immune checkpoint expression in different risk groups. **(B)** The IPS difference of4 types of ICI therapies in high and low BMRS groups. **(C)** Differential analysis of the IC50 of 5 therapeutic agents between BMRS risk groups. **(D)** Correlation between BMRS risk scores and the expression level of immune checkpoints, immune receptors and MHC molecules. **(E)** Potential therapeutic compounds predicted for high and low risk groups respectively.

genes, we discovered two subtypes with distinct clinical and microenvironmental features and established a 7-gene risk scoring system for the prediction of prognosis and treatment response of ccRCC patients.

According to statistics, the prognosis of ccRCC was largely influenced by the occurrence of distant metastasis, which could decrease the 5-year survival rate to less than 20% (Siegel et al., 2022). In line with this finding, our results demonstrated that BM-related genes could divide ccRCC patients into two clusters with distinct prognosis features. This could be explained by the defensive role of BM in preventing the tumor cells from invading the stroma (Valastyan and Weinberg, 2011). Matrix metalloproteinases (MMPs), a series of zinc-dependent proteinases, played an important role in the degradation of BM to weaken its barrier capacity. Without MMPs, the tumor cells could hardly squeeze through the nanosized pores in the BM (Wolf et al., 2013; Eatemadi et al., 2017). It was demonstrated that the increased expression of MMPs was correlated with the invasion and poor prognosis of carcinomas (Winer et al., 2018), and in our study, the upregulation of MMP13 was identified in cluster A. Moreover, the high infiltration of fibroblasts in cluster A could also be associated with the expression of MMPs. It was reported that tumor cells induced the secretion of MMPs from cancer-associated fibroblasts (CAFs) to modulate BM (Kessenbrock et al., 2010). Conversely, MMPs such as MMP1 secreted from tumor cells promote the transdifferentiation of fibroblasts into CAFs (Heneberg, 2016). In addition to the chemical changes, BM

could also be modulated into an invasion-favored status. A recent study revealed that in those BM with high plasticity, cells could mechanically enlarge the nanosized pores and migrate through BM without the help of MMPs (Wisdom et al., 2018). Meanwhile, the stiffness of BM could influence tumor invasion through a protein called netrin-4 (Net4). Net4 in BM mechanistically bound to laminin and diluted laminin ternary node complex, softening the BM and making it more resistant to tumor cell invasion (Reuten et al., 2021). Therefore, the different statuses of BM could influence the prognosis of ccRCC and may facilitate the survival prediction.

Based on the BM-related clusters, we constructed a risk scoring system called BMRS. Genes included in BMRS were found to be correlated with the progression and metastasis of cancers including both risk and protective genes. Holliday junction recognition protein (HJURP) is a kind of centromeric protein, being essential for the stimulation of chromosome division and cell mitosis (Zhang et al., 2021). Its upregulation was correlated with an increased invasion and migration capacity of cancer (Chen et al., 2019). Prolyl 4-hydroxylase alpha subunit 3 (P4HA3) is also a risk gene that could strengthen the motility and invasiveness of tumors (Song et al., 2018). As an enzymic subunit of prolyl 4-hydroxylase, P4HA3 was critical for the stability of collagen and its dysregulation would directly activate the invasiveness potential of tumor cells (Nakasuka et al., 2021). Meanwhile, a disintegrin and metalloproteinase 8 (ADAM8) was another gene related to enzyme production and could promote tumor metastasis probably through the degradation of ECM components (Conrad et al., 2019). Apart from risk genes, BMRS also contained protective genes including adherens junctional associated protein-1 (AJAP1) and ankyrin G (ANK3). AJAP1 belonged to multi-protein complexes named adherens junction, which is critical for cell adhesion and growth inhibition (Xu et al., 2019; Zhou et al., 2022b). ANK3, a family member of ankyrins, assisted in maintaining cell stability through anchoring cytoskeleton to the cell membrane (Wang et al., 2016). The downregulation of both genes could lead to the proliferation and invasion of cancer cells. However, little was known about coiled-coil domain-containing protein 85A (CCDC85A) and complement component 8 gamma (C8G) in tumorigenesis and further investigations were required.

Further analysis of the TME in ccRCC revealed that BM-based clusters and risk groups had distinct immune landscape. Higher immune scores and infiltration levels of immune cells were discovered in the cluster or group that possessed more enriched BM-related functions. It was reported that BM was one of the barriers through which the extravasation of lymphocytes into tumor sites should be overcome (Marchand et al., 2019). During this process, laminin, the major component of BM, played an important role in mediating the functions of lymphocyte trafficking (Pozzi et al., 2017). Evidence suggested that laminins containing LAMA4 favored the transmigration of T cells through providing some permissive signals while those with LAMA5 tended to oppose this process (Sixt et al., 2001). Besides, T cells might adhere less strongly to LAMA4-bearing laminins but migrate faster across these lamins than those with LAMA5 (Zhang et al., 2020). BM distributed with a higher level of LAMA4-containing laminins would have a higher potential for lymphocyte extravasation (Wu et al., 2009). However, we discovered that cluster with high immune cell infiltration was related to poor prognosis and most of the infiltrated lymphocytes in ccRCC including CD8 T lymphocyte and CD4 T lymphocyte were associated with low survival time. This could be due to the dysfunction of lymphocytes by the suppressive lymphocytes and tumor cells. Myeloid-derived suppressor cells (MDSCs) were a cluster of immune suppressive myeloid cells frequently found in cancers (Gabrilovich et al., 2007). As their name suggested, MDSCs exerted suppressive functions on various cells, especially T lymphocytes (Gabrilovich et al., 2012). They could induce tolerance of antigen-specific T lymphocytes mainly through the production of reactive oxygen species (ROS) to nitrate receptors on T lymphocytes and reduce their responsiveness, inhibiting the anti-tumor function (Nagaraj et al., 2007). Apart from suppressive immune cells, tumor cells themselves could transduce inhibitory signals to lymphocytes for immune evasion (Muenst et al., 2016). Inhibitory immune checkpoints such as PD1 were targeted by tumor cells to transform T lymphocytes into suppressive status (Daassi et al., 2020). Thus, BM remodeling may assist in the development of the rich but suppressive immune microenvironment in ccRCC.

With an increasing understanding of the mechanisms underlying tumorigenesis, therapeutic strategies developed against them were shown to be effective in cancer therapy. ICIs, functioning through interrupting the suppressive signals transduced by CTLA4 or PD1 to reactivate the anti-tumor immunity and prevent tumor immune evasion, was approved to be successful in the treatment of multiple cancers (Sharma and Allison, 2015). It was believed that normalization of the suppressive environment and restoration of the anti-tumor immunity would be more effective than directly enhancing the immune function (Sanmamed and Chen, 2018). In line with this statement, our finding suggested that clusters with high immunogenicity potential but suppressed responded better to ICIs. The higher level of SNV, HRD, and CTA in the high risk group indicated it possessed a higher potential to generate tumor-associated neoantigens, which were critical for the immune system to exert anti-tumor function (Meng et al., 2021; van Wilpe et al., 2021; Wang et al., 2021). Meanwhile, once the suppressive environment was removed, the relatively high infiltration level of immune cells served as an immune reservoir that supported a powerful immune reaction (Waldman et al., 2020). Additionally, the increased expression of immune checkpoints in the high risk group was thought to be predictive for immunotherapy (Thommen et al., 2018). Apart from ICI monotherapy, recent studies had domonstrated the benefits of treatment combined ICI and TKI, which could have the better therapeutic capacity (Quhal et al., 2021). Our results also demonstrated that the high risk group was more sensitive to TKIs including sorafenib and sunitinib, as well as chemotherapy like mitomycin C. The mutation analysis helped discover a co-occurance mutant pair, SETD2 and PBRM1, in which SETD2 mutation was correlated with a favorable outcome of ICI-treated patients (Lu et al., 2021) and PBRM1 mutation was associated with high angiogenesis and TKIs therapeutic outcomes (Motzer et al., 2020b), indicating the potential of the high risk group for combination therapy. It was demonstrated that TKIs therapy was associated with improved vessel extravasation and enhanced drug delivery to tumors (Zhou et al., 2008; Zhou and Gallo, 2009). Moreover, after TKI treatment, the amount of suppressive immune cells including Treg and MDSCs was found to be decreased in TME (Finke et al., 2008; Ko et al., 2009). Therefore, BMRS constructed in the current study could not only help individualized administration of immunotherapy but also assist in the combination therapy of ICIs and TKIs.

In addition to the existing therapeutic agents, the investigation was also focus on novel therapeutic targets. In patients with high BMRS, we identified the upregulation of chemokine (C-X-C motif) ligand 13 (CXCL13) and its receptor CXCR5. Their interaction could impede the tumor-specific cytotoxic function of CD8 T lymphocytes and be related to the recruitment of suppressive immune cells including MDSCs and Treg (Ammirante et al., 2014). Inhibitor targeting the CXCL13/CXCR5 axis was demonstrated to have an encouraging effect on cancer treatment (Hussain et al., 2019). A Rho associated kinase inhibitor, BRD-K23875128, may be a potential therapeutic agent for high risk patients. Rho kinase pathway took part in multiple cell functions and was implicated in tumor metastasis as well as ECM remodeling. It may be a candidate for combination treatment due to its ability to increase sensitivity to other therapeutic drugs (Kim et al., 2021). As for patients with low BMRS, there may also be some curative agents. Histocompatibility leucocyte antigen E (HLA-E) and its cognate inhibitory receptor NKG2A could serve as a novel immune checkpoint to be targeted for inducing anti-tumor immunity (Borst et al., 2020). As predicted by CMap, arginase inhibitor, could also potentially treat patients with high BMRS by promoting the T cells activation and proliferation to exert an anti-tumor immune response (Borek et al., 2020).

Overall, the current study identified BM-related subtypes of ccRCC and constructed a risk scoring system BMRS for prognosis and therapeutic prediction based on both public and clinical data. The mutation, TME and treatment analysis also provided potential novel therapeutic agents for further investigations. Comparing with the existing similar gene signatures, BMRS possessed higher net benefit in predicting the prognosis of ccRCC patients. Besides, we had validated the scoring system and its gene expression using external dataset and clinical samples. However, the lack of accessibility to clinical treatment-related data restricted our analysis of BMRS in predicting therapeutic response. More in-depth study could be conducted to validate the proposed potential therapeutic targets for ccRCC.

5 Conclusion

In summary, we recognized BM-related subtypes of ccRCC with distinct survival and TME features. A risk scoring system BMRS was established for prognosis prediction and individualized treatment instruction. Mechanistic investigations based on BM-related clusters and risk groups helped identified some therapeutic candidates. With more and more studies focusing on combination therapy, our results may provide certain practical instructions for clinical application and future research.

Data availability statement

The data and information demonstrated and analyzed throughout the present study were obtained from the Genomics Data Commons Data Portal (https://portal.gdc.cancer.gov/), ArrayExpress (https://www.ebi.ac.uk/arrayexpress/), Gene Set Enrichment Analysis (http://www.gsea-msigdb.org/gsea/index.jsp), xCell (https://xcell.ucsf.edu/), The Cancer Immunome Database (https://tcia.at/home), Genomics of Drug Sensitivity in Cancer (https://www.cancerrxgene.org), ConnectivityMap (CMap, https://clue.io). The patient information of the clinical transcriptome data could be accessed from the supplementary material (Supplementary Table S1).

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Guangdong Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YT: Conception and design, data analysis and interpretation, collection and assembly of data; CY: Conception and design, data analysis and interpretation; JZ and PZ: Data analysis and interpretation; SC, WZ, and BY: Collection and assembly of data; YY and YL: Administrative support. All authors wrote and approved the final version of the manuscript.

Funding

This work was supported by Guangdong Basic and Applied Basic Research Foundation (Grant/Award Number: 2020A1515011473) and Science and Technology Program of Guangzhou (Grant/Award Number: 202002030060).

Acknowledgments

The information utilized in this study was retrieved from The Cancer Genome Atlas (TCGA), ArrayExpress, xCell, cBioPortal for Cancer Genomics, The Cancer Immunome Database, Genomics of Drug Sensitivity in Cancer and Gene Set Enrichment Analysis. We sincerely appreciate their sharing of the data and information.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1038924/full#supplementary-material

References

Amin, A., and Hammers, H. (2018). The evolving landscape of immunotherapy-based combinations for frontline treatment of advanced renal cell carcinoma. *Front. Immunol.* 9, 3120. doi:10.3389/fimmu.2018.03120

Ammirante, M., Shalapour, S., Kang, Y., Jamieson, C. A. M., and Karin, M. (2014). Tissue injury and hypoxia promote malignant progression of prostate cancer by inducing CXCL13 expression in tumor myofibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* 111 (41), 14776–14781. doi:10.1073/pnas.1416498111

Barbour, J. D., Cox, N., and Huang, R. S. (2014). pRRophetic: An R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS ONE* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Bissell, M. J., and Hines, W. C. (2011). Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. Med.* 17 (3), 320–329. doi:10.1038/nm.2328

Blankenstein, T., Coulie, P. G., Gilboa, E., and Jaffee, E. M. (2012). The determinants of tumour immunogenicity. *Nat. Rev. Cancer* 12 (4), 307–313. doi:10.1038/nrc3246

Borek, B., Gajda, T., Golebiowski, A., and Blaszczyk, R. (2020). Boronic acid-based arginase inhibitors in cancer immunotherapy. *Bioorg Med. Chem.* 28 (18), 115658. doi:10.1016/j.bmc.2020.115658

Borst, L., van der Burg, S. H., and van Hall, T. (2020). The nkg2a-HLA-E Axis as a novel checkpoint in the tumor microenvironment. *Clin. Cancer Res.* 26 (21), 5549–5556. doi:10.1158/1078-0432.CCR-19-2095

Chang, J., and Chaudhuri, O. (2019). Beyond proteases: Basement membrane mechanics and cancer invasion. *J. Cell Biol.* 218 (8), 2456–2469. doi:10.1083/jcb.201903066

Chen, T., Zhou, L., Zhou, Y., Zhou, W., Huang, H., Yin, S., et al. (2019). HJURP promotes epithelial-to-mesenchymal transition via upregulating SPHK1 in hepatocellular carcinoma. *Int. J. Biol. Sci.* 15 (6), 1139–1147. doi:10.7150/ijbs.30904

Choueiri, T. K., and Motzer, R. J. (2017). Systemic therapy for metastatic renal-cell carcinoma. *N. Engl. J. Med.* 376 (4), 354–366. doi:10.1056/NEJMra1601333

Conrad, C., Benzel, J., Dorzweiler, K., Cook, L., Schlomann, U., Zarbock, A., et al. (2019). ADAM8 in invasive cancers: Links to tumor progression, metastasis, and chemoresistance. *Clin. Sci. (Lond)* 133 (1), 83–99. doi:10.1042/CS20180906

Daassi, D., Mahoney, K. M., and Freeman, G. J. (2020). The importance of exosomal PDL1 in tumour immune evasion. *Nat. Rev. Immunol.* 20 (4), 209–215. doi:10.1038/s41577-019-0264-y

de Velasco, G., Culhane, A. C., Fay, A. P., Hakimi, A. A., Voss, M. H., Tannir, N. M., et al. (2017). Molecular subtypes improve prognostic value of international metastatic renal cell carcinoma database consortium prognostic model. *Oncologist* 22 (3), 286–292. doi:10.1634/theoncologist.2016-0078

Díaz-Montero, C. M., Rini, B. I., and Finke, J. H. (2020). The immunology of renal cell carcinoma. *Nat. Rev. Nephrol.* 16, 721–735. doi:10.1038/s41581-020-0316-3

Eatemadi, A., Aiyelabegan, H. T., Negahdari, B., Mazlomi, M. A., Daraee, H., Daraee, N., et al. (2017). Role of protease and protease inhibitors in cancer pathogenesis and treatment. *Biomed. Pharmacother.* 86, 221–231. doi:10.1016/j.biopha.2016.12.021

Finke, J. H., Rini, B., Ireland, J., Rayman, P., Richmond, A., Golshayan, A., et al. (2008). Sunitinib reverses type-1 immune suppression and decreases T-regulatory cells in renal cell carcinoma patients. *Clin. Cancer Res.* 14 (20), 6674–6682. doi:10.1158/1078-0432.CCR-07-5212

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Gabrilovich, D. I., Bronte, V., Chen, S. H., Colombo, M. P., Ochoa, A., Ostrand-Rosenberg, S., et al. (2007). The terminology issue for myeloid-derived suppressor cells. *Cancer Res.* 67 (1), 425. doi:10.1158/0008-5472.CAN-06-3037

Gabrilovich, D. I., Ostrand-Rosenberg, S., and Bronte, V. (2012). Coordinated regulation of myeloid cells by tumours. *Nat. Rev. Immunol.* 12 (4), 253–268. doi:10.1038/nri3175

Greef, B., and Eisen, T. (2016). Medical treatment of renal cancer: New horizons. *Br. J. Cancer* 115 (5), 505–516. doi:10.1038/bjc.2016.230

Heneberg, P. (2016). Paracrine tumor signaling induces transdifferentiation of surrounding fibroblasts. *Crit. Rev. Oncol. Hematol.* 97, 303–311. doi:10.1016/j.critrevonc.2015.09.008

Ho, T. H., Serie, D. J., ParasraMka, M., Cheville, J. C., Bot, B. M., Tan, W., et al. (2017). Differential gene expression profiling of matched primary renal cell carcinoma and metastases reveals upregulation of extracellular matrix genes. *Ann. Oncol.* 28 (3), 604–610. doi:10.1093/annonc/mdw652

Hussain, M., Adah, D., Tariq, M., Lu, Y., Zhang, J., and Liu, J. (2019). CXCL13/CXCR5 signaling axis in cancer. *Life Sci.* 227, 175–186. doi:10.1016/j.lfs.2019.04.053

Jayadev, R., Morais, M. R. P. T., Ellingford, J. M., Srinivasan, S., Naylor, R. W., Lawless, C., et al. (2022). A basement membrane discovery pipeline uncovers network complexity, regulators, and human disease associations. *Sci. Adv.* 8 (20), eabn2265. doi:10.1126/sciadv.abn2265

Jhaveri, K. D., and Perazella, M. A. (2018). Adverse events associated with immune checkpoint blockade. *N. Engl. J. Med.* 378 (12), 1163. doi:10.1056/NEJMc1801663

Jonasch, E., Gao, J., and Rathmell, W. K. (2014). Renal cell carcinoma. *BMJ* 349, g4797. doi:10.1136/bmj.g4797

Kessenbrock, K., Plaks, V., and Werb, Z. (2010). Matrix metalloproteinases: Regulators of the tumor microenvironment. *Cell* 141 (1), 52–67. doi:10.1016/j.cell.2010.03.015

Kim, S., Han, J., and Kim, I. S. (2021). Rho-kinase as a target for cancer therapy and its immunotherapeutic potential. *Int. J. Mol. Sci.* 22 (23), 12916. doi:10.3390/ijms222312916

Ko, J. S., Zea, A. H., Rini, B. I., Ireland, J. L., Elson, P., Cohen, P., et al. (2009). Sunitinib mediates reversal of myeloid-derived suppressor cell accumulation in renal cell carcinoma patients. *Clin. Cancer Res.* 15 (6), 2148–2157. doi:10.1158/1078-0432.CCR-08-1332

Li, P., Ren, H., Zhang, Y., and Zhou, Z. (2018). Fifteen-gene expression based model predicts the survival of clear cell renal cell carcinoma. *Med. Baltim.* 97 (33), e11839. doi:10.1097/MD.0000000000011839

Liu, X., Qiao, Y., Chen, J., and Ge, G. (2022). Basement membrane promotes tumor development by attenuating T cell activation. *J. Mol. Cell Biol.* 14 (2), mjac006. doi:10.1093/jmcb/mjac006

Lu, M., Zhao, B., Liu, M., Wu, L., Li, Y., Zhai, Y., et al. (2021). Pan-cancer analysis of SETD2 mutation and its association with the efficacy of immunotherapy. *NPJ Precis. Oncol.* 5 (1), 51. doi:10.1038/s41698-021-00193-0

Marchand, M., Monnot, C., Muller, L., and Germain, S. (2019). Extracellular matrix scaffolding in angiogenesis and capillary homeostasis. *Semin. Cell Dev. Biol.* 89, 147–156. doi:10.1016/j.semcdb.2018.08.007

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Meng, X., Sun, X., Liu, Z., and He, Y. (2021). A novel era of cancer/testis antigen in cancer immunotherapy. *Int. Immunopharmacol.* 98, 107889. doi:10.1016/j.intimp.2021.107889

Motzer, R. J., Banchereau, R., Hamidi, H., Powles, T., McDermott, D., Atkins, M. B., et al. (2020). Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. *Cancer Cell* 38 (6), 803–817.e4. doi:10.1016/j.ccell.2020.10.011

Motzer, R. J., Escudier, B., McDermott, D. F., George, S., Hammers, H. J., Srinivas, S., et al. (2015). Nivolumab versus everolimus in advanced renal-cell carcinoma. *N. Engl. J. Med.* 373 (19), 1803–1813. doi:10.1056/NEJMoa1510665

Motzer, R. J., Jonasch, E., Boyle, S., Carlo, M. I., Manley, B., Agarwal, N., et al. (2020). NCCN guidelines insights: Kidney cancer, version 1.2021. *J. Natl. Compr. Canc Netw.* 18 (9), 1160–1170. doi:10.6004/jnccn.2020.0043

Muenst, S., Laubli, H., Soysal, S. D., Zippelius, A., TzAnkov, A., and Hoeller, S. (2016). The immune system and cancer evasion strategies: Therapeutic concepts. *J. Intern Med.* 279 (6), 541–562. doi:10.1111/joim.12470

Nagaraj, S., Gupta, K., Pisarev, V., Kinarsky, L., Sherman, S., Kang, L., et al. (2007). Altered recognition of antigen is a mechanism of CD8+ T cell tolerance in cancer. *Nat. Med.* 13 (7), 828–835. doi:10.1038/nm1609

Nakasuka, F., Tabata, S., Sakamoto, T., Hirayama, A., Ebi, H., Yamada, T., et al. (2021). TGF-beta-dependent reprogramming of amino acid metabolism induces epithelial-mesenchymal transition in non-small cell lung cancers. *Commun. Biol.* 4 (1), 782. doi:10.1038/s42003-021-02323-7

Pozzi, A., Yurchenco, P. D., and Iozzo, R. V. (2017). The nature and biology of basement membranes. *Matrix Biol.* 57-58, 1–11. doi:10.1016/j.matbio.2016.12.009

Quhal, F., Mori, K., Bruchbacher, A., Resch, I., Mostafaei, H., Pradere, B., et al. (2021). First-line immunotherapy-based combinations for metastatic renal cell carcinoma: A systematic review and network meta-analysis. *Eur. Urol. Oncol.* 4 (5), 755–765. doi:10.1016/j.euo.2021.03.001

Reuten, R., Zendehroud, S., Nicolau, M., Fleischhauer, L., Laitala, A., Kiderlen, S., et al. (2021). Basement membrane stiffness determines metastases formation. *Nat. Mater* 20 (6), 892–903. doi:10.1038/s41563-020-00894-0

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

Sanmamed, M. F., and Chen, L. (2018). A paradigm shift in cancer immunotherapy: From enhancement to normalization. *Cell* 175 (2), 313–326. doi:10.1016/j.cell.2018.09.035

Sharma, P., and Allison, J. P. (2015). The future of immune checkpoint therapy. *Science* 348 (6230), 56–61. doi:10.1126/science.aaa8172

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics. *CA Cancer J. Clin.* 72 (1), 7–33. doi:10.3322/caac.21708

Sixt, M., Engelhardt, B., Pausch, F., Hallmann, R., Wendler, O., and Sorokin, L. M. (2001). Endothelial cell laminin isoforms, laminins 8 and 10, play decisive roles in T cell recruitment across the blood–brain barrier in experimental autoimmune encephalomyelitis. *J. Cell Biol.* 153 (5), 933–946. doi:10.1083/jcb.153.5.933

Song, H., Liu, L., Song, Z., Ren, Y., Li, C., and Huo, J. (2018). P4HA3 is epigenetically activated by slug in gastric cancer and its deregulation is associated with enhanced metastasis and poor survival. *Technol. Cancer Res. Treat.* 17, 1533033818796485. doi:10.1177/1533033818796485

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality

worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Thommen, D. S., Koelzer, V. H., Herzig, P., Roller, A., Trefny, M., Dimeloe, S., et al. (2018). A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat. Med.* 24 (7), 994–1004. doi:10.1038/s41591-018-0057-z

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48 (4), 812–830 e14. doi:10.1016/j.immuni.2018.03.023

Valastyan, S., and Weinberg, R. A. (2011). Tumor metastasis: Molecular insights and evolving paradigms. *Cell* 147 (2), 275–292. doi:10.1016/j.cell.2011.09.024

van Wilpe, S., Tolmeijer, S. H., Koornstra, R. H. T., de Vries, I. J. M., Gerritsen, W. R., Ligtenberg, M., et al. (2021). Homologous recombination repair deficiency and implications for tumor immunogenicity. *Cancers (Basel)* 13 (9), 2249. doi:10.3390/cancers13092249

Waldman, A. D., Fritz, J. M., and Lenardo, M. J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* 20 (11), 651–668. doi:10.1038/s41577-020-0306-5

Wang, T., Abou-Ouf, H., Hegazy, S. A., Alshalalfa, M., Stoletov, K., Lewis, J., et al. (2016). Ankyrin G expression is associated with androgen receptor stability, invasiveness, and lethal outcome in prostate cancer patients. *J. Mol. Med. Berl.* 94 (12), 1411–1422. doi:10.1007/s00109-016-1458-4

Wang, Y., Shi, T., Song, X., Liu, B., and Wei, J. (2021). Gene fusion neoantigens: Emerging targets for cancer immunotherapy. *Cancer Lett.* 506, 45–54. doi:10.1016/j.canlet.2021.02.023

Winer, A., Adams, S., and Mignatti, P. (2018). Matrix metalloproteinase inhibitors in cancer therapy: Turning past failures into future successes. *Mol. Cancer Ther.* 17 (6), 1147–1155. doi:10.1158/1535-7163.MCT-17-0646

Wisdom, K. M., Adebowale, K., Chang, J., Lee, J. Y., Nam, S., Desai, R., et al. (2018). Matrix mechanical plasticity regulates cancer cell migration through confining microenvironments. *Nat. Commun.* 9 (1), 4144. doi:10.1038/s41467-018-06641-z

Wolf, K., Te Lindert, M., Krause, M., Alexander, S., Te Riet, J., Willis, A. L., et al. (2013). Physical limits of cell migration: Control by ECM space and nuclear deformation and tuning by proteolysis and traction force. *J. Cell Biol.* 201 (7), 1069–1084. doi:10.1083/jcb.201210152

Wragg, J. W., Finnity, J. P., Anderson, J. A., Ferguson, H. J. M., Porfiri, E., Bhatt, R. I., et al. (2016). MCAM and LAMA4 are highly enriched in tumor blood vessels of renal cell carcinoma and predict patient outcome. *Cancer Res.* 76 (8), 2314–2326. doi:10.1158/0008-5472.CAN-15-1364

Wu, C., Ivars, F., Anderson, P., Hallmann, R., Vestweber, D., Nilsson, P., et al. (2009). Endothelial basement membrane laminin alpha5 selectively inhibits T lymphocyte extravasation into the brain. *Nat. Med.* 15 (5), 519–527. doi:10.1038/nm.1957

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. (N Y)* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Xiong, X., Chen, C., Yang, J., Ma, L., Wang, X., Zhang, W., et al. (2022). Characterization of the basement membrane in kidney renal clear cell carcinoma to guide clinical therapy. *Front. Oncol.* 12, 1024956. doi:10.3389/fonc.2022.1024956

Xu, C., Liu, F., Xiang, G., Cao, L., Wang, S., Liu, J., et al. (2019). β-Catenin nuclear localization positively feeds back on EGF/EGFR-attenuated AJAP1 expression in breast cancer. *J. Exp. Clin. Cancer Res.* 38 (1), 238. doi:10.1186/s13046-019-1252-6

Xu, T., Le, T. D., Liu, L., Su, N., Wang, R., Sun, B., et al. (2017). CancerSubtypes: An R/bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics* 33 (19), 3131–3133. doi:10.1093/bioinformatics/btx378

Yi, M., Nissley, D. V., McCormick, F., and Stephens, R. M. (2020). ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Sci. Rep.* 10 (1), 10258. doi:10.1038/s41598-020-66986-8

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

Yurchenco, P. D. (2011). Basement membranes: Cell scaffoldings and signaling platforms. *Cold Spring Harb. Perspect. Biol.* 3 (2), a004911. doi:10.1101/cshperspect.a004911

Zhang, F., Yuan, D., Song, J., Chen, W., Wang, W., Zhu, G., et al. (2021). HJURP is a prognostic biomarker for clear cell renal cell carcinoma and is linked to immune infiltration. *Int. Immunopharmacol.* 99, 107899. doi:10.1016/j.intimp.2021.107899

Zhang, X., Wang, Y., Song, J., Gerwien, H., Chuquisana, O., Chashchina, A., et al. (2020). The endothelial basement membrane acts as a checkpoint for entry of pathogenic T cells into the brain. *J. Exp. Med.* 217 (7), e20191339. doi:10.1084/jem.20191339

Zhou, H., Huang, S., Shao, C., Zou, J., Zhou, A., Yu, J., et al. (2022). miR-1266-3p suppresses epithelial-mesenchymal transition in colon cancer by targeting P4HA3. *Anal. Cell Pathol. (Amst)* 2022, 1542117. doi:10.1155/2022/1542117

Zhou, Q., and Gallo, J. M. (2009). Differential effect of sunitinib on the distribution of temozolomide in an orthotopic glioma model. *Neuro Oncol.* 11 (3), 301–310. doi:10.1215/15228517-2008-088

Zhou, Q., Guo, P., and Gallo, J. M. (2008). Impact of angiogenesis inhibition by sunitinib on tumor distribution of temozolomide. *Clin. Cancer Res.* 14 (5), 1540–1549. doi:10.1158/1078-0432.CCR-07-4544

Zhou, T., Chen, W., Wu, Z., Cai, J., and Zhou, C. (2022). A newly defined basement membrane-related gene signature for the prognosis of clear-cell renal cell carcinoma. *Front. Genet.* 13, 994208. doi:10.3389/fgene.2022.994208

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 03 February 2023
DOI 10.3389/fgene.2023.1067666

OPEN ACCESS

EDITED BY
Shun Lu,
University of Electronic Science and Technology of China, China

REVIEWED BY
Lucia Guadalupe Taja Chayeb,
National Institute of Cancerology (INCAN), Mexico
Wenjie Shi,
Otto von Guericke University Magdeburg, Germany
Huiwu Ouyang,
Cold Spring Harbor Laboratory, United States
Saumya Pandey,
Indira IVF Hospital, India

*CORRESPONDENCE
Keke Huang,
148348824@qq.com
Xiaomei Liang,
drsherryleung@hotmail.com
Yifeng Wang,
wangyifeng@smu.edu.cn

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

RECEIVED 12 October 2022
ACCEPTED 20 January 2023
PUBLISHED 03 February 2023

CITATION
Lai W, Liao J, Li X, Liang P, He L, Huang K, Liang X and Wang Y (2023), Characterization of the microenvironment in different immune-metabolism subtypes of cervical cancer with prognostic significance.
Front. Genet. 14:1067666.
doi: 10.3389/fgene.2023.1067666

Characterization of the microenvironment in different immune-metabolism subtypes of cervical cancer with prognostic significance

Wujiang Lai[1], Jinrong Liao[1], Xiaoxuan Li[1], Peili Liang[2,3], Liqing He[1], Keke Huang[4]*, Xiaomei Liang[1]* and Yifeng Wang[1]*

[1]Obstetrics and Gynecology Center, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [2]Department of Obstetrics and Gynecology, Guangdong Provincial Key Laboratory of Major Obstetric Diseases, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [3]Center for Reproductive Medicine/Department of Fetal Medicine and Prenatal Diagnosis/BioResource Research Center, Guangdong Provincial Key Laboratory of Major Obstetric Diseases, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [4]Department of Obstetrics, Shunde Hospital, The First People's Hospital of Shunde, Southern Medical University, Foshan, Guangdong, China

Introduction: Immune cell infiltration and metabolic reprogramming may have great impact on the tumorigenesis and progression of malignancies. The interaction between these two factors in cervical cancer remains to be clarified. Here we constructed a gene set containing immune and metabolism related genes and we applied this gene set to molecular subtyping of cervical cancer.

Methods: Bulk sequencing and single-cell sequencing data were downloaded from the Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database respectively. Immune and metabolism related genes were collected from Immport and Kyoto encyclopedia of genes and genomes (KEGG) database respectively. Unsupervised consensus clustering was performed to identify the molecular subtypes. Cibersort was applied to evaluate the immune cells infiltration status. Differential expression analysis and Gene set enrichment analysis (GSEA) were performed to characterize the molecular pattern of different subtypes. Multivariate Cox regression analysis was used for prognosis prediction model construction and receiver operating characteristic (ROC) curve was used for performance evaluation. The hub genes in the model were verified in single-cell sequencing dataset and clinical specimens. *In vitro* experiments were performed to validate the findings in our research.

Results: Three subtypes were identified with prognostic implications. C1 subgroup was in an immunosuppressive state with activation of mitochondrial cytochrome P450 metabolism, C2 had poor immune cells infiltration and was characterized by tRNA anabolism, and the C3 subgroup was in an inflammatory state with activation of aromatic amino acid synthesis. The area under the ROC curve of the constructed model was 0.8, which showed better performance than clinical features. IMPDH1 was found to be significantly upregulated in tumor tissue and it was demonstrated that IMPDH1 could be a novel therapeutic target *in vitro*.

Discussion: In summary, our findings suggested novel molecular subtypes of cervical cancer with distinct immunometabolic profiles and uncovered a novel therapeutic target.

KEYWORDS

cervical cancer, molecular subclassification, immune microenvironment, metabolic reprogramming, prognosis

Introduction

Cervical cancer (CC), a malignancy associated with high-risk human papillomavirus (HPV), remains the second leading cause of death in women aged 20–39 despite widespread early screening and prophylactic vaccination. It was demonstrated that a total of 4,152 women died of CC in 2019 (Siegel et al., 2022). It has been observed that there are significant individual differences in the prognosis of cancer patients despite controlling for factors such as age and clinical stage, which is likely due to tumor heterogeneity (Punt et al., 2017). For example, polymorphisms of TLRs that play a key role in innate immunity were found to be correlated with susceptibility to cervical cancer (Pandey et al., 2011). However, the tumor heterogeneity has not been fully clarified and the prognostic effect of the International Federation of Gynecology and Obstetrics (FIGO) staging system does not meet the clinical needs (Kupets and Covens, 2001). Therefore, it is necessary to explore the molecular heterogeneity and establish a more complete prognostic evaluation system, which may help improve the precision treatment for cervical cancer.

Cancer cells tend to be in the spotlight of many studies on cancer biology. However, it has been confirmed that the tumor microenvironment (TME) is a non-negligible factor in tumorigenesis, in which cancer cells may interact with extracellular matrix (ECM) and stromal cells. The TME is composed of a variety of cells, including fibroblasts, endothelial cells, mesenchymal cells and immune cells (Hanahan and Coussens, 2012). During interactions with tumor cells, the phenotype of both stromal cells and immune cells can be shaped to support tumor cell growth (Coussens et al., 2013). In the context of choric inflammation caused by cancer cells, myeloid cell precursors may be induced to proliferate and differentiate into the myeloid derived suppressor cells (MDSCs) with the binding of soluble tumor necrosis factor to the corresponding receptor (Sobo-Vujanovic et al., 2016). Once differentiated, MDSCs may be home to the TME and the subsequent vicious cycle of chronic inflammation, immunosuppression, tumor growth and differentiation cannot be stopped, which may result in poor prognosis (Ugel et al., 2015). Tumor associated macrophages (TAMs) were abundant in solid tumors and their appearance promoted tumor cell invasion and metastasis (Qian et al., 2009; Qian et al., 2011). It was demonstrated that the differentiation states of macrophages can be influenced by cancer cells (Mantovani and Sica, 2010). For example, Yang Cheng et al. found that PKN2 derived from colon cancer cells can inhibit M2 phenotype polarization, which may help promote anti-tumor immune response and improve prognosis (Cheng et al., 2018). $CD8^+$ T cells play an important role in tumor suppression but they can be exhausted during the progression of cancer. Yongshuai Jiang et al. found that PRMT5 derived from cancer cells could suppress the function of tumor infiltrating T cells and promote the development of cervical cancer (Jiang et al., 2021). In addition, dendritic cells (DCs) expressing PD-1 were found to be correlated with advanced stages, elevated preoperative squamous cell carcinoma antigen levels and lymph-vascular space invasion, which suggests its role in immune surveillance dysfunction in cervical cancer (Wang et al., 2022). Overall, tumor-infiltrating immune cells were one of the important factors for the survival and prognosis of cancer patients, but the heterogeneity of TME in cervical cancer has not been fully elucidated and deserves further exploration.

It is well known that metabolic reprogramming is another characteristic of malignancies and plays an important role in tumor progression. Otto Warburg found that tumor cells generate energy through glycolysis under aerobic conditions and produce a lot of lactic acid, which shapes a hypoxic and acidic tumor microenvironment (Koppenol et al., 2011). Cytotoxic T cells are the most important cells in the anti-tumor immune response and their function was restricted by the glucose metabolism of tumor cells, resulting in tumor progression (Chang et al., 2015). It was shown that tumor glycolysis impacted T cell infiltration in the TME and impaired the efficacy of adoptive T cell therapy (Cascone et al., 2018). Aerobic glycolysis in tumor cells has been shown to promote the infiltration of MDSCs, thereby suppressing anti-tumor immune responses in triple-negative breast cancer (Li et al., 2018). On the other hand, immunomodulatory cells like M2 macrophages and regulatory T cells may impact the function of T cells by depleting the arginine in the TME (Speiser et al., 2016). Glutamine metabolism in tumor cells was enhanced in the TME where aerobic glycolysis produced large amounts of lactate, which caused glutamine deprivation to infiltrated immune cells and affected their proliferation (Carr et al., 2010). Tryptophan metabolism was found to play a role in immunosuppression state in various tumors (Platten et al., 2019). IDO1, one of tryptophan metabolizing enzymes, was demonstrated to correlate with low tumor infiltration of T cells in colorectal cancer (Brandacher et al., 2006), ovarian cancer (Inaba et al., 2009), and endometrial cancer (Ino et al., 2008). High levels of IDO and TDO have also been shown to contribute to impaired anti-tumor immune responses (Munn et al., 2005). Luc Pilotte et al. found that activation of tryptophan metabolism-related enzyme TDO2 in tumor cells can significantly inhibit the activation of T cells (Pilotte et al., 2012). Mutations of the isocitrate dehydrogenase genes IDH1 and IDH2 were found in the lower-grade glioma (LGG) and the mutated forms could convert α-ketoglutarate (α-KG) to the oncometabolite R-2-hydroxyglutarate (2HG) (Ichimura, 2012). Kohanbash G et al. found that IDH-MUT glioma reduced the production of T cells attracting chemokines and the accumulation of T cells was suppressed (Kohanbash et al., 2017). Therefore, tumor metabolic status can have an impact on infiltrating immune cells and may be a target for improving the efficacy of tumor immunotherapy.

In order to explore the heterogeneity of the TME in cervical cancer, an immune-metabolism related gene set was constructed and was applied to the identification of subgroups with prognostic significance in cervical cancer from The Cancer Genome Atlas Project (TCGA). Immune cell infiltration and metabolism status were evaluated respectively and the interaction was explored. Besides, we constructed a prognostic model in cervical cancer and validated the expression of key genes on a single-cell dataset from Gene Expression Omnibus (GEO). Our findings suggested the existence of immuno-metabolic subgroups of cervical cancer and uncovered novel therapeutic targets for cervical cancer.

Methods and materials

Human cervical cancer cohorts

The gene expression data of cervical cancer and the corresponding clinical information in the Cancer Genome Atlas (TCGA) were

downloaded from the UCSC data portal (https://xenabrowser. net), which consists of 306 cervical tumor samples and 3 normal samples. We used the corresponding annotation file from the same database to convert the transcriptome raw count value in TCGA cohort to transcripts per kilobase million (TPM) values.

Construction of the immunometabolism gene set and identification of the subtypes

851 immune related genes were downloaded from the Immport database (https://www.immport.org/home) and 1,401 metabolism related genes were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg). We applied the univariate Cox proportional hazards model to evaluate the association of these genes with overall survival. We included the genes with HR < 0.8 or HR > 1.2 and p values < 0.05 in the subsequent sample clustering. ConsensusClusterPlus, an R package designed for unsupervised consensus clustering, was used to identify the subclusters of the TCGA-CESC cohort.

Evaluation of differences in gene expression and metabolism pathways among the subtypes

R (4.2.1) was used to identify the differentially expressed genes (DEGs) among the cervical cancer subtypes with the EdgeR package. Genes with an absolute value greater than 2 and an FDR value less than 0.05 were considered as DEGs. To clarify the pathway status among subtypes, Gene set variation analysis (GSVA) was performed with GSVA package to calculate the enrichment score of different pathways in every single CESC sample, including "hallmark gene set," "KEGG gene set," "GO biological processes," "GO cellular components" and "GO molecular functions" downloaded from the Molecular Signatures Database (MsigDB, https://www.gsea-msigdb.org/gsea/msigdb).

Evaluation of the immuno-metabolic microenvironment

Cibersort, a deconvolution algorithm for dissecting the cell component in bulk sequence data, was used to characterize the abundance of various immune cells infiltrated and the corresponding cell states (Newman et al., 2015). As for the evaluation of difference in metabolic status, 85 metabolic pathways and the corresponding gene sets were acquired from the KEGG database and the single sample gene set enrichment analysis (ssGSEA) algorithm was used to calculate the enrichment scores. Subsequently, Limma package was used to perform the differential analysis of the metabolic pathways.

Generation of the prognostic gene signature

The entire cohort was randomly divided into a training dataset (70%) and a validation dataset (30%). In the training cohort, the Lasso-Cox regression analysis was used to select the genes with prognostic value, Then, a multivariate Cox hazard ratio model was constructed with 10 genes selected. The risk score was calculated based on the expression data in the validation cohort and the corresponding coefficients in the model. According to the coefficients in the model, the formula for calculating the risk score is: risk score = FLT3LG*(−0.147) + IMPDH1*(-0.011) + OPRD1*0.836 + MOCS1*0.163 + IL1B*0.331 + GALNT10*0.489 + TNFRSF11B*(−0.048) + LDHC*(−1.075) + ISG20*(−0.444) + TRAV12_3*(−0.134). Finally, We used the survivalROC package to evaluate the prognostic value. With the median value of risk score, we divided the entire cohort into high and low risk groups, for which survival analysis was performed.

Gene expression analysis at single cell resolution

To elucidate the cell specificity of gene expression and validate the difference between tumor and normal samples. GSE168652, a single cell sequencing data of cervical cancer derived from Hua's research (Li et al., 2021) was downloaded from Gene Expression Omnibus (GEO) datasets. Seurat (version 4.1.1) was used to perform quality control, data filtration, data scale, dimension reduction, clustering and cell type annotation, in which the criteria were set the same as the original research. For pseudo-time trajectory analysis, Monocle (version 2.20.0) was used to analyze the cell state transition of cancer cells and visualize the gene expression patterns along the trajectory.

Immunohistochemistry

Human cervical cancer tissue sections were retrospectively obtained from surgical resections that were fixed in buffered formalin, embedded in paraffin, and stored at the Zhujiang Hospital, Southern Medical University, Guangzhou, China. The corresponding clinical data were obtained from medical records and identified. The ethics committee of Zhujiang Hospital approved the use of the clinical specimens. For IHC staining, antigen retrieval was performed by heat treatment in a microwave oven for 21 min in Tris-ethylene diamine tetraacetic acid (EDTA) buffer solution (0.05 mol/L Tris, 0.001 mol/L EDTA; pH 8.5). Endogenous peroxidase activity was inactivated using 0.3% H_2O_2 for 10 min followed by washing with PBS (Gibco, C14190500BT). After blocking by 5% BSA for 20 min, the slides were incubated overnight at 4°C with the following primary antibodies used (proteintech, 22092-1-AP-50UL). After washing with PBS, the sections were incubated with HRP conjugated goat anti-rabbit IgG secondary antibodies (Cell Signalling, 7074) for 50 min. Finally, immunoreactivity was detected using 3,3-diaminobenzidine (Servicebio, G1211), followed by re-staining with hematoxylin. Images were obtained by using 3D HISTECH (Pannoramic MIDI II).

Cell culture

Human cervical cancer cell lines HeLa, SiHa, Caski and c33a were purchased from the Procell Life Science&Technology Co.,Ltd. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) with supplement of 10% fetal bovine serum (FBS) (Gibco, 10099-141C) and

100 units/mL penicillin and streptomycin (Sigma, St. Louis, MO, United States) at 37°C with a humidified atmosphere of 5% CO_2 maintenance.

Western blotting

Cells were homogenized in RIPA lysis buffer (sc-24948; Santa Cruz Biotechnology, Inc.), and protein contents were measured using Bicin-choninic Acid (BCA) protein assay kit (CWBIO, CW0014S). Incubate cell protein with SDS-PAGE loading buffer (CWBIO, CW0052S) at 100°C for 10 min to denature the protein. After electrophoresis with SDS-PAGE, the separated proteins from the gel were transferred onto polyvinylidene fluoride (PVDF) membrane and then were subjected to western blotting with specific primary antibodies followed by detection with horseradish peroxidase-conjugated secondary antibody (solarbio, SE134-1mL, SE131-1 mL) and enhanced chemiluminescence (Merck millipore, WBKLS0100). The antibodies used in this study include the following: anti-IMPDH1 antibody (proteintech, 22092-1-AP-50UL, 1:2000); GAPDH (MC4) Mouse Monoclonal Antibody (Beijing Ray Antibody Biotech, RM 2002, 1:50000). ImageJ software was used to quantify the protein bands. Target protein expression was normalized to GAPDH to correct for loading.

Quantitative real-time PCR

Total RNA from cells was isolated by TRIzol extraction according to the manufacturer's instructions (Thermo Fisher, 15596026), and cDNA was synthesized with a reverse-transcription kit (Vazyme, R323-01). The quantitative real-time PCR (qRT-PCR) experiment was conducted using SYBR Green Real-Time PCR Master Mix Kit (Vazyme, Q711-02) with the Light Cycler LC480 (Roche). Primer pairs for quantitative real-time PCR were synthesized from Tsingke Biotechnology Co., Ltd. IMPDH1(F: 5′-CAGCAGGTGTGACGT TGAAAG-3′, R: 5′-AGCTCATCGCAATCATTGACG-3′); ACTB(F: 5′-AGAGCTACGAGCTGCCTGAC-3′, R: 5′-AGCACT GTGTTGGCGTACAG-3′) Values were calculated by the change in threshold method (ΔΔCT).

CCK8

Cells were seeded at 5,000 cells per well in 96-well plates according to the manufacturer's guidelines. Cells were allowed to adhere overnight, and treated with different interventions (n = 3 wells/group) for the indicated time. A total of 10 μL Cell Counting Kit-8 (CCK8) reagents (APExBIO, K1018-5) were added and incubated for 1 h. Then, the absorbance was read at 450 nm. Statistical analysis (mean ± SD) with triplicates is shown.

Flow cytometry

After induction of apoptosis, cells from each treatment condition were washed once in PBS. The apoptotic rate was evaluated according to the protocols provided by the Annexin V-FITC Apoptosis Detection Kit (Beyotime, China). Generally, 1 × 10^5 cells were diluted within buffer, and stained with FITC-conjugated Annexin V and PI according to the manufacturer's instructions. The cell mixture was cultured at room temperature for 20 min and then analyzed by the CytoFLEX instrument (Beckman).

Transfection

For knockdown assays, short interfering RNAs (siRNAs) targeting IMPDH1 were synthesized by RiboBio Co., Ltd. (Guangdong, China). The sequence are listed: genOFFTM st-h-IMPDH1_001: 5′-GGTGAT GACGCCAAGGATT-3′; genOFFTM st-h-IMPDH1_002: 5′-GCA CCGACCTGAAGAAGAA-3′; genOFFTM st-h-IMPDH1_003: 5′-GTACAAGGTGGCTGAGTAT-3′; All cells were transfected using Lipofectamine 3000 Reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instructions.

Statistical analysis

Cox regression model was used to evaluate the hazard ratio and prognostic significance of genes in the OS. KM and Cox regression analysis were applied to calculate the significance of difference in OS, PFS and DSS. Log-rank test was used to evaluate the statistical difference of the KM curves. For evaluation of the predictive power of immuno-metabolism risk score to OS, the time-dependent area under the receiver operating characteristic curve (AUC) and C-index (also termed concordance index) were calculated. Higher value of these two indicators represented better accuracy. In terms of correlation analysis, the Spearman method was used to calculate the correlation coefficient and the p-value. Kruskal–Wallis test was used when the statistical difference of distribution in three or more groups was examined and Wilcoxon test was used when comparisons contain only two groups. If not specified, p values were two-sided and $p < 0.05$ was defined as statistically significant.

Results

Identification of the immune-metabolism subtypes of cervical cancer

The immune related genes and the metabolism related genes were downloaded from the Immport database and KEGG database respectively. We constructed an immune-metabolism gene set and determined the candidate genes with prognostic values using Univariate Cox proportional hazards regression model analysis. As a result, the top three immune related genes with unfavorable prognosis were SHC4(HR = 5.3, p = 0.044, 95%CI, 1–26), LEPR (HR = 3.4, p = 0.0064, 95%CI, 1.4–8.1) and OPRD1(HR = 3.3, p = 0.00082, 95%CI, 1.6–6.6) while TRGC1(HR = 0.016, p = 0.0051, 95% CI, 0.00092–0.29), TRGC2(HR = 0.048, p = 0.0029, 95%CI, 0.0065–0.35) and ANGPTL6(HR = 0.079, p = 0.00017, 95%CI, 0.021–0.3) were the top three significant protective factors (Figure 1A). It was reported that SHC4 was involved in the progression of hepatocellular cancer (Urabe et al., 2020) and prostate cancer (Zhang et al., 2022). LEPR was found to be overexpressed in epithelial ovarian cancer indicating poor

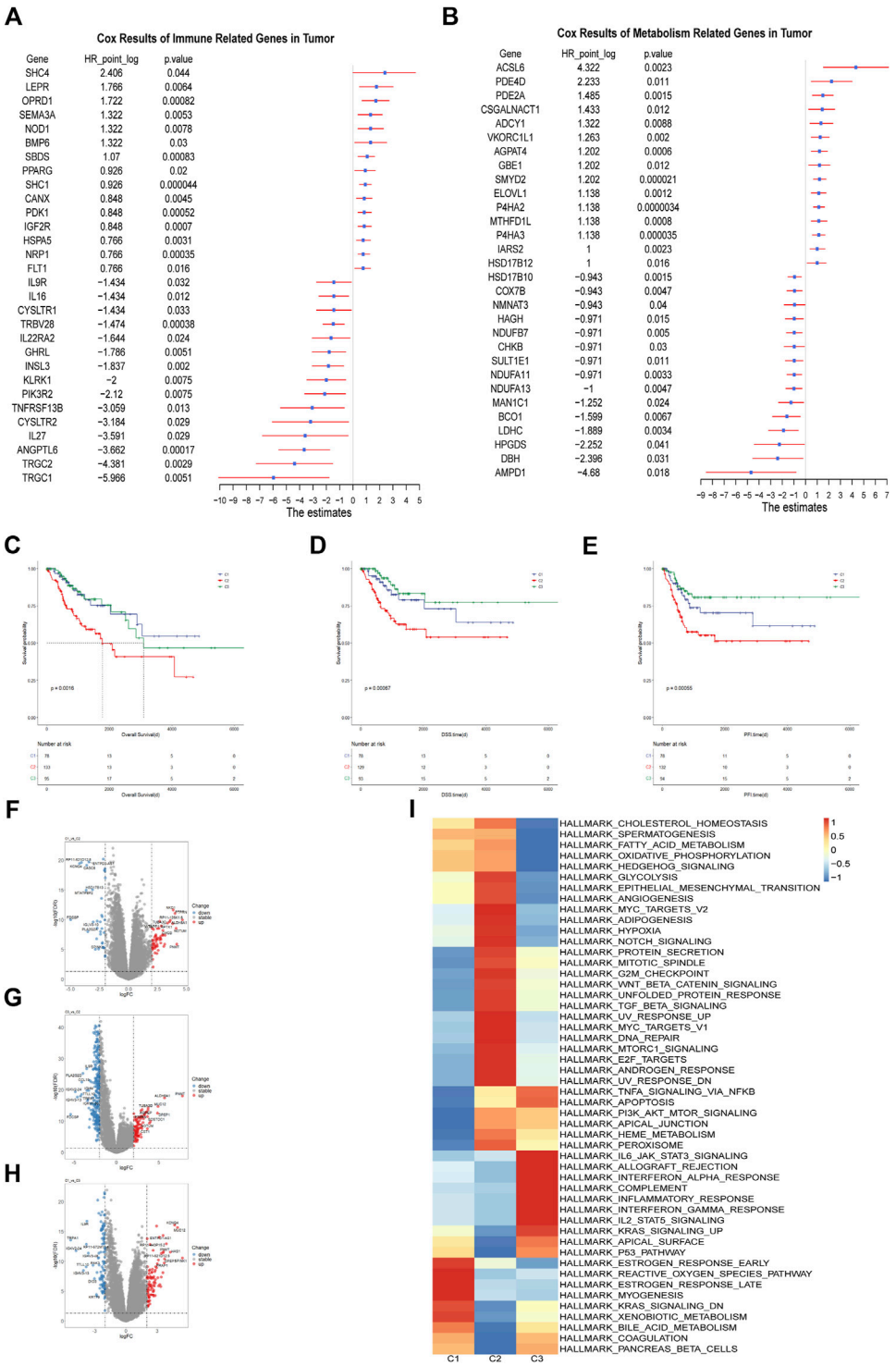


FIGURE 1
Identification of the Immune-metabolism Subtypes of Cervical Cancer **(A)** Hazard ratio of top 30 immune related genes that meet the requirements of HR < 0.8 or HR > 1.2 and $p < 0.05$ associated with overall survival. **(B)** Hazard ratio of top 30 metabolism related genes that meet the requirements of HR < 0.8 or HR > 1.2 and $p < 0.05$ associated with overall survival. **(C–E)** Kaplan-Meier curves of overall survival, progression free interval and disease specific survival among the subtypes in TCGA cohort. **(F–H)** The volcano plots of differentially expressed genes among the three subtypes in TCGA cohort. **(I)** Heatmap of differentially enriched hallmark pathways from GSEA database among three subtypes in TCGA cohort.

progression-free survival (Uddin et al., 2009) and its somatic mutation was found to increase the susceptibility to hepatocarcinogenesis (Ikeda et al., 2014).

In terms of metabolism, we identified ACSL6 (HR = 20, $p = 0.0023$, 95%CI, 2.9–140), PDE4D (HR = 4.7, $p = 0.011$, 95%CI, 1.4–16) and PDE2A (HR = 2.8, $p = 0.0015$, 95%CI, 1.5–5.2) as the top three risk

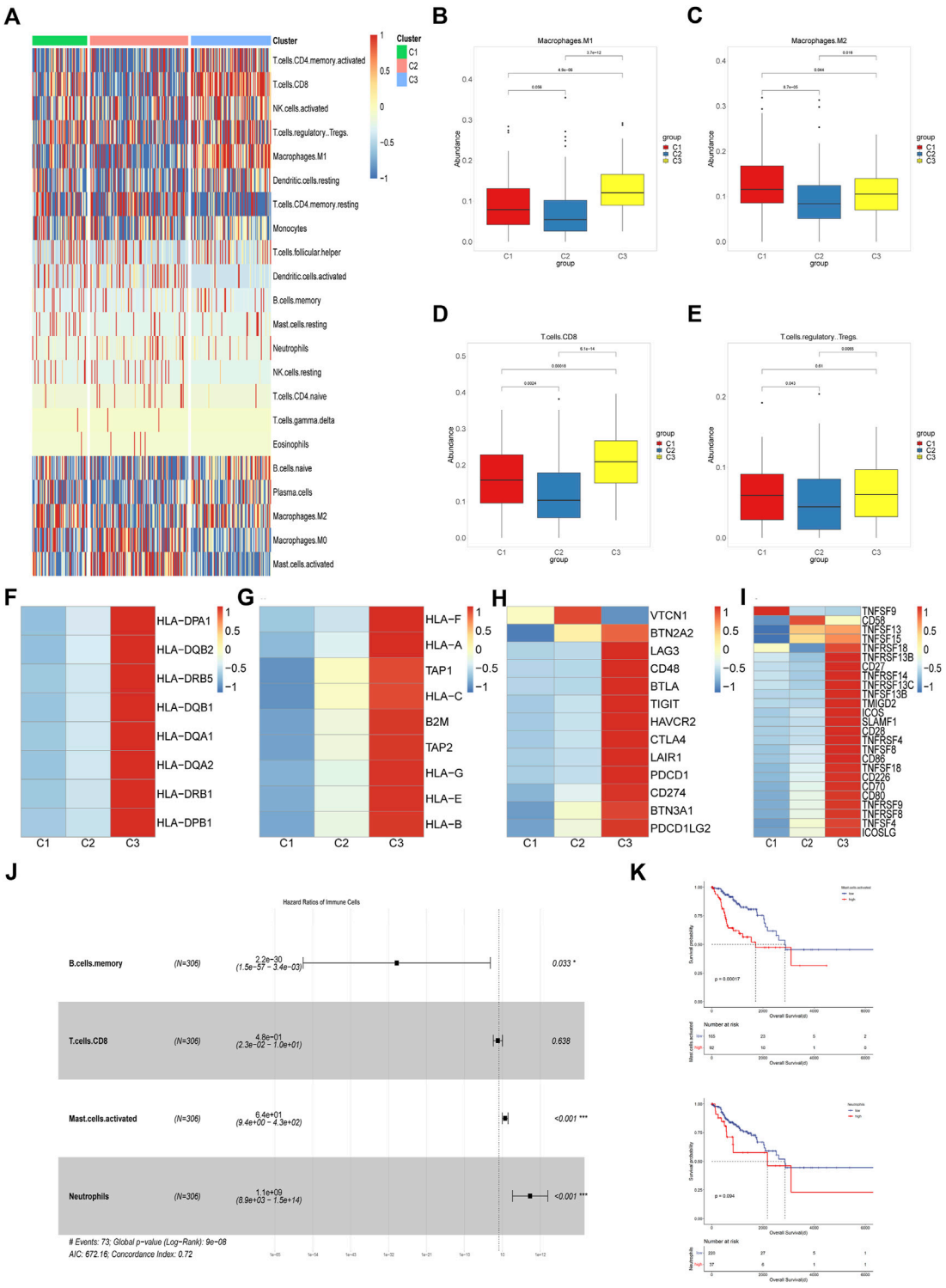


FIGURE 2
Characterization of the Immune Microenvironment Among Different Subpopulations **(A)** Heatmap of the abundance of 22 immune cells among three subgroups in TCGA cohort. **(B–E)** Boxplots of abundance of M1 macrophages, M2 macrophages, $CD8^+$ T cells and Tregs among three different subtypes in TCGA cohort. The differences were compared using the Kruskal–Wallis test. **(F,G)** Heatmap of differential expression of MHC molecules among three subtypes in TCGA cohort. **(H,I)** Difference of co-stimulatory and co-inhibitory molecules expression level in three subtypes. **(J)** Multivariate cox regression model constructed based on the immune cells with prognostic value in univariate cox analysis. **(K)** Kaplan-Meier curves of overall survival between cohorts in TCGA with high and low abundance of activated mast cells and neutrophils infiltrated.

factors. AMPD1 (HR = 0.039, p = 0.018, 95%CI, 0.0027–0.57), DBH (HR = 0.19, p = 0.031, 95%CI, 0.043–0.86) and HPGDS (HR = 0.21, p = 0.041, 95%CI, 0.047–0.94) were identified as the top three protective factors (Figure 1B). PDE4D was demonstrated to be a tumor-promoting factor in prostate cancer (Rahrmann et al., 2009) and BRAF-mutated melanoma (Delyon et al., 2017). Inhibition of PDE4D helped to overcome tamoxifen resistance in ER-positive breast cancer (Mishra et al., 2018).

Those genes with HR > 1.2 or HR < 0.8 and p < 0.05 in the Cox regression model were selected for the subsequent clustering analysis, which contained 154 immune-related genes and 195 metabolism related genes. We applied the unsupervised cluster analysis for the dataset and it was indicated that the entire cohort could be divided into three subgroups (Supplementary Figures 1A–1F). Survival analysis showed that overall survival (OS), progression free survival (PFS) and disease specific survival (DSS) differed significantly among these three subgroups. The C2 subgroup showed worst prognosis (Figures 1C–E). There were distinct gene expression patterns among the three subpopulations so we analyzed the differentially expressed genes (DEGs) among them (Figures 1F–H). Finally, we identified 97 DEGs between C1 and C2, 197 DEGs between C1 and C3 and 337 DEGs between C2 and C3. To explore the difference in signal pathway status, we performed GSVA analysis in the three cervical cancer subgroups. It was indicated that C2 was characterized by some cancer-related pathways such as angiogenesis, hypoxia, and epithelial-mesenchymal transition while C3 was shown to have inflammatory signature, with inflammatory response, interferon gamma response and reactive oxygen species pathway significantly upregulated. As for C1 subgroup, it was characterized by estrogen response, KRAS signaling and xenobiotic metabolism activation (Figure 1I). It is well known that the rapid progression of malignancy may create a hypoxic microenvironment and hypoxia can stimulate the expression of some angiogenesis related factors, which may exacerbate tumor immunosuppression and adversely affect patient outcomes (Rahma and Hodi, 2019). Taken together, We identified three subgroups of cervical cancer with unique molecular features and prognostic significance.

Characterization of the immune microenvironment among different subpopulations

Emerging evidence showed that solid tumor harbor rather complex components, which included immune cells, fibroblasts, endothelial cells and mesenchymal cells and tumor cells. Tumor microenvironment plays an important role in cancer development, immune escape and metastasis (Quail and Joyce, 2013). Different immune microenvironment components were closely related to the responsiveness of chemotherapy, immunotherapy and patient prognosis (Liu et al., 2022). With Cibersort, we evaluated the abundance of 22 immune cells infiltrated in the three subgroups. It was shown that most immune cells infiltrated poorly in the C2 subgroup while abundantly in the C3 subgroup, moderately in the C1 subgroup, which indicated that C3 subgroup had better anti-tumor immune response than others (Figure 2A).Pandey et al. (2009) found that TLRs, the pathogen recognition receptors mainly expressed on immune cells, were correlated with susceptibility to cervical cancer. Therefore we analyzed the expression of TLR2, TLR3 and TLR4 among three subtypes and found that C3 had the highest level, which may account for better innate immunity and better prognosis in the C3 subgroup (Supplementary Figures 2A–2C). Anti-tumor immune response includes multiple steps and various kinds of immune cells were involved in it such as dendritic cells, B cells, macrophages, nature killer cells, and T cells (Motz and Coukos, 2013). Macrophages can be influenced by tumor cells to differentiate into M1 or M2 subtypes, wherein M1 is a pro-inflammatory and anti-tumor subtype while M2 is an anti-inflammatory and tumor-promoting subtype. We compared the abundance of macrophages among the three subgroups and found that C2 subpopulation had the lowest abundance of both subtypes of macrophages while C1 and C3 were dominated by M2 and M1 macrophages respectively (Figures 2B, C). $CD8^+$ T cells are an important part of anti-tumor immunity but are prone to depletion phenotype transformation in the tumor microenvironment, which is one of the reasons for the low response rate of tumor immunotherapy. We found that $CD8^+$ T cell abundance was highest in the C3 subpopulation, lowest in the C2 subpopulation, and intermediate in the C1 subpopulation, which may indicate that C2 has the worst anti-tumor immune status (Figure 2D). As for regulatory T cells, both C1 and C3 were significantly more abundant than C2 while there was no significant difference between C1 and C3, suggesting that C1 has immunosuppressive characteristics (Figure 2E). Antigen presentation is an important step in immune response, which is dependent on the expression of major histocompatibility complex (MHC). It was reported that cancer cells can evade attack by immune cells through downregulating the expression of MHC molecules (Marincola et al., 2000; van der Burg et al., 2016). So we evaluated the expression of several MHC I/II molecules and found that C1 and C2 subpopulations showed lower expression level than C3 subpopulation (Figures 2F, G). Activation and expansion of T cells require the co-stimulatory molecules and we found that C1 and C2 had significantly lower levels of co-stimulatory molecules compared to C3 (Figure 2H). In addition, high levels of co-inhibitory molecules were detected in the C3 subpopulation, which may indicate that C3 subpopulation benefits from immune checkpoint blockade (ICB) therapy (Figure 2I). It is well known that cytotoxic T cells depend on interferon gamma (IFN-γ) and granzyme B to attack cancer cells (Jenkins and Griffiths, 2010). In our study significantly lower expression level of these two genes and lower enrichment score of IFN-γ signaling were found in C1 and C2, suggesting impaired T cells function in these two subpopulations (Supplementary Figures 2D–2E). Finally, we assessed the prognostic significance of various immune cells in the cohort. Multivariate cox regression analysis showed that the abundance of memory B cells was an independent protective factor and that activated mast cells and neutrophils were independent risk factors (Figure 2J). Survival analysis showed that the population with higher mast cell abundance had a significantly worse prognosis than the population with lower mast cell abundance while neutrophil abundance did not significantly distinguish the cohort for survival differences (Figure 2K).

Overall, we identified three subgroups of the cervical cancer cohort based on the immune-metabolism gene set and analyzed the immune infiltration status of the three subgroups. We found that C3 had the highest abundance of immune cell infiltration and was characterized by inflammation, C2 had the least abundance of immune cells and had the worst prognosis, and C1 had some immune cell infiltration but mainly had an immune-exhausted phenotype.

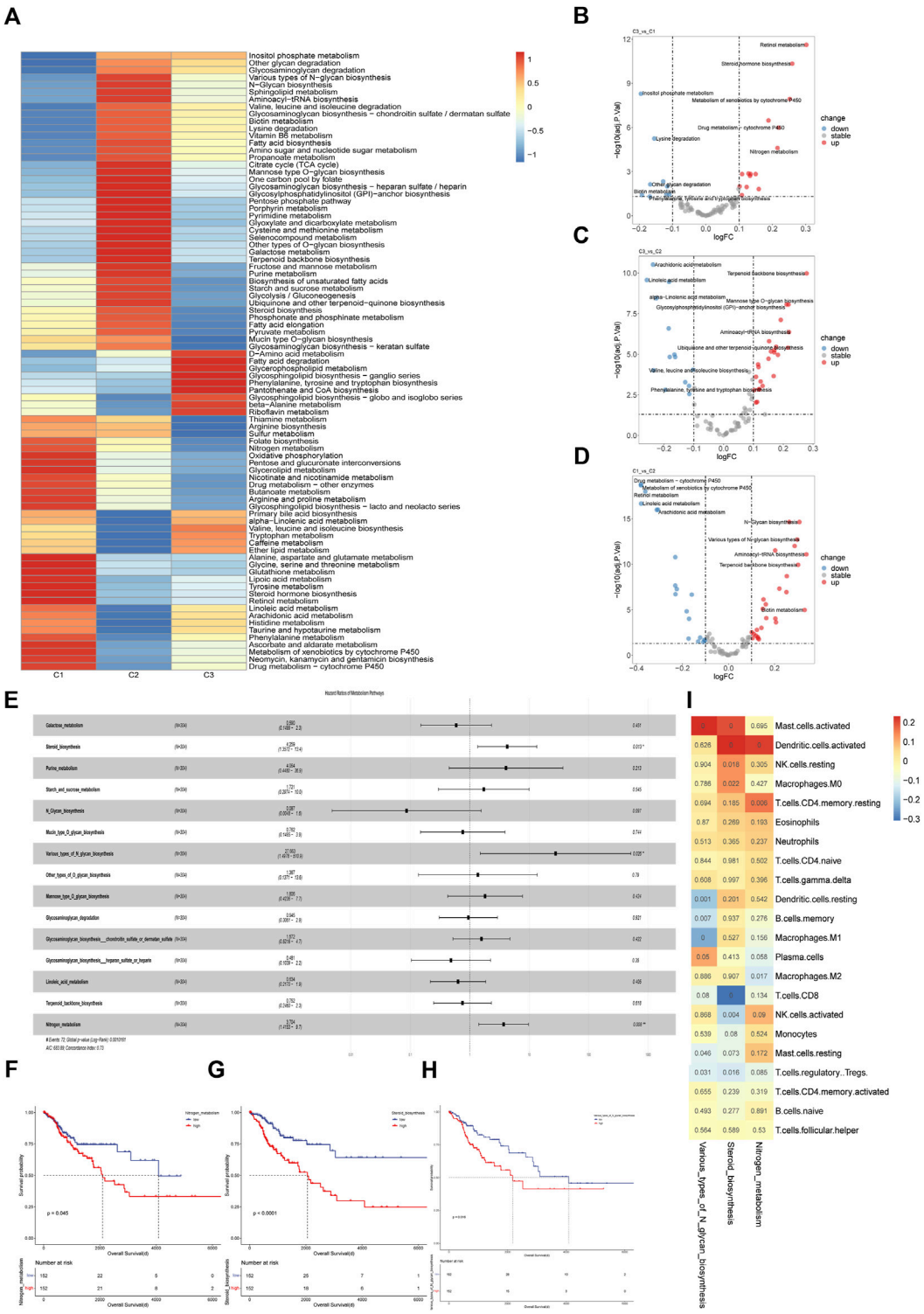


FIGURE 3
Metabolic characteristics of cervical cancer subgroups **(A)** Heatmap of differentially enriched metabolic pathways in three subtypes in TCGA cohort. The enrichment analysis was performed with GSVA algorithm. **(B–D)** Volcano plot of differentially enriched metabolic pathways among three subtypes in TCGA cohort. The differential analysis was performed based on the GSVA analysis. **(E)** Multivariate cox regression model constructed based on the metabolic pathways with prognostic value in univariate cox analysis. **(F–H)** Kaplan-Meier curves of overall survival between cohorts in TCGA with high and low enrichment score of nitrogen metabolism, steroid biosynthesis and various types of N glycan biosynthesis. The enrichment analysis was performed with GSVA algorithm. **(I)** Correlation matrix of the specific metabolic pathways and the 22 immune cells infiltrated in the whole cohort. Correlation coefficients are represented in the form of heatmap using colored scale ranging from blue (minimum correlation) to red (maximum correlation) and the *p*-value were presented.

Metabolic characteristics of cervical cancer subgroups

It was known that malignancies undergo metabolic reprogramming in order to adapt to the needs of rapid proliferation (Martínez-Reyes and Chandel, 2021), resulting in abnormal accumulation of metabolites in the solid tumor microenvironment, thereby affecting various cellular components in the tumor microenvironment (Li et al., 2019).

Using the GSVA algorithm, we performed an enrichment analysis of 85 metabolic pathways in the KEGG database for the cervical cancer cohort. The enrichment score heatmap showed a distinct pattern in metabolic status among the three subgroups (Figure 3A). Next, we performed differential analysis of metabolic pathways among the three subgroups and the results showed that a total of 107 differential metabolic pathways were identified. After ranking by logFC value, the top 5 pathways were included in subsequent analysis, and only those upregulated in comparison with other subgroups were considered significantly enriched pathways. It was illustrated that the drug metabolism-cytochrome P450 pathway, metabolism of xenobiotics by cytochrome P450 pathway and retinol metabolism pathway were significantly upregulated in the C1 subpopulation (Figure 3B). The metabolic pathways that were significantly upregulated in C2 were aminoacyl-tRNA biosynthesis pathway and terpenoid backbone biosynthesis pathway (Figure 3C). As for C3, phenylalanine, tyrosine and tryptophan biosynthesis were most significantly upregulated (Figure 3D). To explore the prognostic significance of multiple metabolic pathways in the whole cohort, we performed a univariate cox analysis for each metabolic pathway, and only those metabolic pathways with statistical significance were included in the subsequent multivariate cox regression analysis. The results showed that steroid biosynthesis, various type of N-glycan biosynthesis and nitrogen pathway were independent risk factors in multivariate cox regression analysis (Figure 3E). Based on the median of these three metabolic pathways enrichment scores, we divided the cohort into groups with high and low levels of the corresponding pathways, and then performed survival analysis. We found that higher activity of the three metabolic pathways was associated with worse prognosis (Figures 3F–H). Next, we performed a correlation analysis between immune cell abundance and the enrichment scores of metabolic pathways with independent prognostic value (Figure 3I). The data showed that both nitrogen metabolism ($r = 0.22$, $p < 0.01$) and steroid biosynthesis ($r = 0.23$, $p < 0.01$) pathway were significantly positively correlated with activated dendritic cell. M1 macrophages were significantly negatively correlated with the various type of N-glycan biosynthesis pathway ($r = -0.24$, $p < 0.01$) while M2 macrophages were significantly negatively correlated with the nitrogen metabolism pathway ($r = -0.14$, $p = 0.017$). The abundance of $CD8^+$ T cells ($r = -0.31$, $p < 0.01$) and activated NK cells ($r = -0.16$, $p = 0.004$) showed significantly negative correlation with steroid biosynthesis pathway. Regulatory T cells (Tregs) are significantly negatively correlated with the various type of N-glycan biosynthesis ($r = -0.12$, $p = 0.031$) and steroid biosynthesis pathway ($r = 0.14$, $p = 0.016$). According to the published research, it was demonstrated that metabolic reprogramming in tumor cells could impair the anti-tumor immunity (Hung et al., 2021; Kao et al., 2022). Therefore, it was indicated that correlation and interaction existed between tumor cells and infiltrated immune cells in cervical cancer. The phenotype and function of immune cells may be affected by tumor metabolic reprogramming, which may promote immune escape in CC.

Construction of an immuno-metabolic prognostic model for cervical cancer

Based on the immune-metabolism gene set, we identified patient subgroups with significantly different prognosis. Therefore, the immunometabolism gene set was applied to construct a cervical cancer prognosis prediction model. Firstly, we divided 70% of the cervical cancer cohort into the training cohort and the remaining 30% into the validation cohort. Then, with the lasso-cox regression method, we screened the 10 genes with the prognostic value from the immune metabolism gene set (Figures 4A, B), which included FLT3LG, IMPDH1, OPRD1, MCOS1, IL1B, GALNT10, TNFRSF11B, LDHC, ISG20 and TRAV12-3. Next, a multivariate cox regression model based on the 10 prognostic-related genes was constructed in the training cohort and the risk scores were calculated based on the corresponding gene coefficients in the model (Figure 4C). With the risk scores, we performed a prognostic prediction in the validation cohort and the area under the receiver operating curve of the prediction model was 0.8, which was higher than that of the clinical stage prediction model of 0.69 (Figure 4D). Based on the median risk score, the whole cohort was divided into a high-risk group and a low-risk group, and there was a significant difference in survival between the two groups ($p < 0.0001$) (Figure 4E), indicating that immuno-metabolic factors were important for the prognosis of cervical cancer patients. Additionally, a multivariate cox regression analysis incorporating immuno-metabolic risk model scores with FIGO stage, TMN stage, and age was performed, and risk score was found to be an independent adverse prognostic factor (Figure 4F). Finally, a nomogram for prognosis prediction was constructed for cervical cancer (Figure 4G).

IMPDH1 was a significant prognostic risk gene in cervical cancer

Bulk sequencing failed to distinguish the effects of cellular components in the TME on tumorigenesis while single-cell sequencing can make up for this limitation. To verify the cellular origin of genes in prognostic models, a single-cell transcriptome dataset of cervical cancer was downloaded from Gene Expression Omnibus (GEO) database and used for analysis. Data filtration, integration, and dimensionality reduction clustering were performed according to the parameter in original research. A total of 13 cell clusters were identified and cancer cells, endometrial stromal cells, endothelial cells, fibroblasts, lymphocytes, macrophages, and smooth muscle cells were annotated respectively according to corresponding cell surface markers (Figures 5A, B). IMPDH1 and ISG20 were mainly expressed in tumor cells, while the remaining genes were less specific to tumor cell origin (Figure 5C). Then we performed a pseudo-time analysis of tumor cells and compared the expression differences of these genes between normal cells and tumor cells (Figure 5D). The results showed that the expression levels of FLT3LG, GALNT10, IL1B, IMPDH1 and ISG20 were higher in tumor cells than in normal cells, while the remaining genes could not be identified because of the low expression levels in tumor cells (Figure 5E). In addition, we found that with the evolution of tumor cell status, the expression levels of FLT3LG, GALNT10, IL1B, and IMPDH1 remained stable while ISG20 gradually increased (Figure 5F). It was shown that IMPDH1 was one of the isoforms of inosine-5′-monophosphate dehydrogenase (IMPDH) which contributed to the formation of

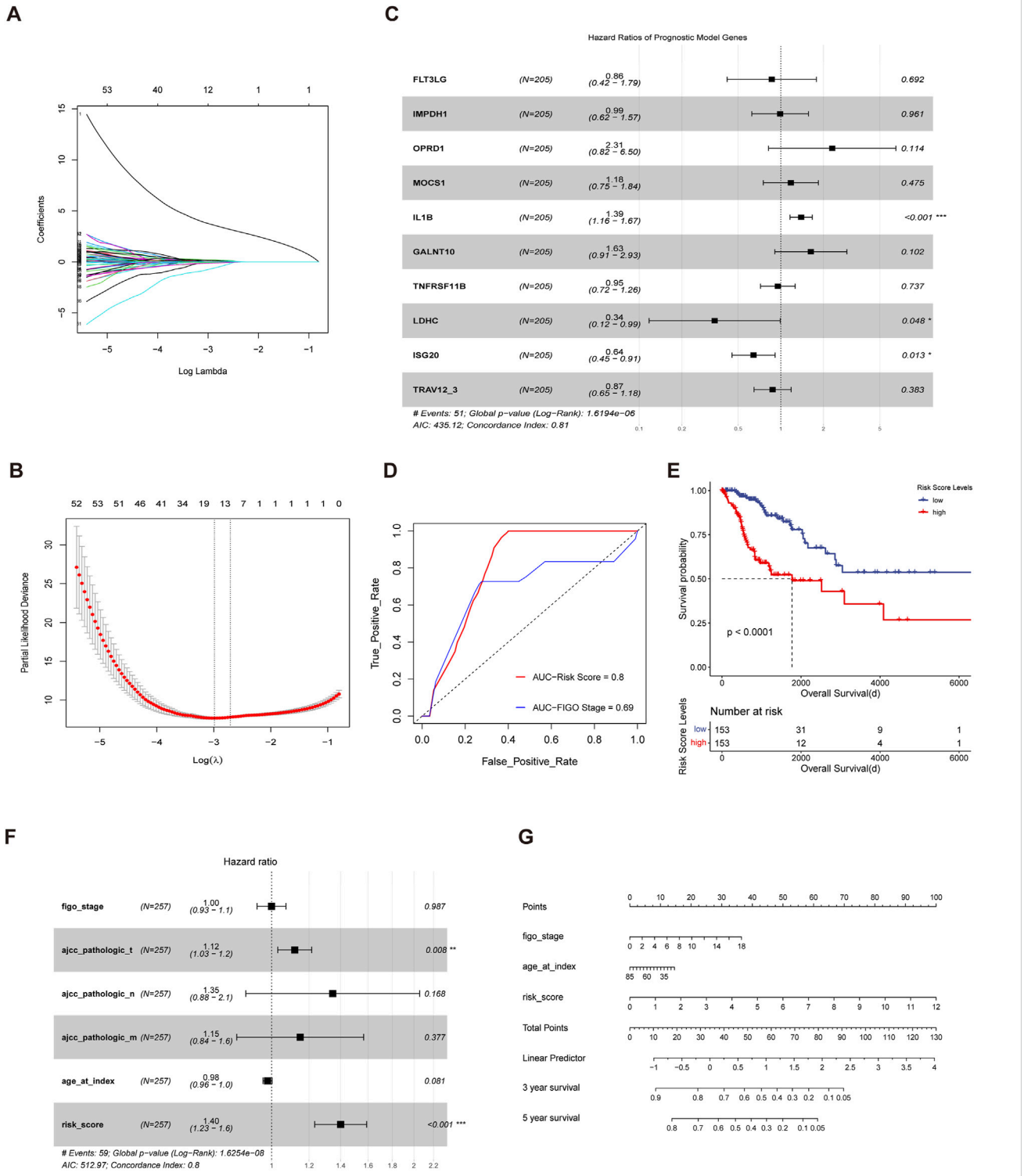


FIGURE 4
Construction of an immuno-metabolic prognostic model for cervical cancer **(A,B)** Fit and cvfit plots of LASSO screen. **(C)** Forest plot of Multivariate cox regression model constructed based on the genes screened out from LASSO in training cohort. **(D)** ROC curves measuring the predictive value of risk score and clinical stage. The area under the ROC curve was 0.82 and 0.69 for the risk score and clinical stage, respectively. **(E)** Kaplan-Meier curves of overall survival between the cohorts of TCGA with low and high risk score. **(F)** Forest plot of the multivariate cox regression model constructed with clinical stage, age and risk score. **(G)** Nomogram for predicting probability of survival at 3 and 5 years in cervical cancer.

cytoophidia and tumor progression (Ruan et al., 2020a). For further analysis of IMPDH1, IMPDH2 and ISG20, we divided the TCGA cohort into high and low level groups based on the mean expression levels of corresponding molecules. The results showed that there were significant prognostic differences between groups with different expression levels of IMPDH1 but not IMPDH2 or ISG20 (Figure 5G).

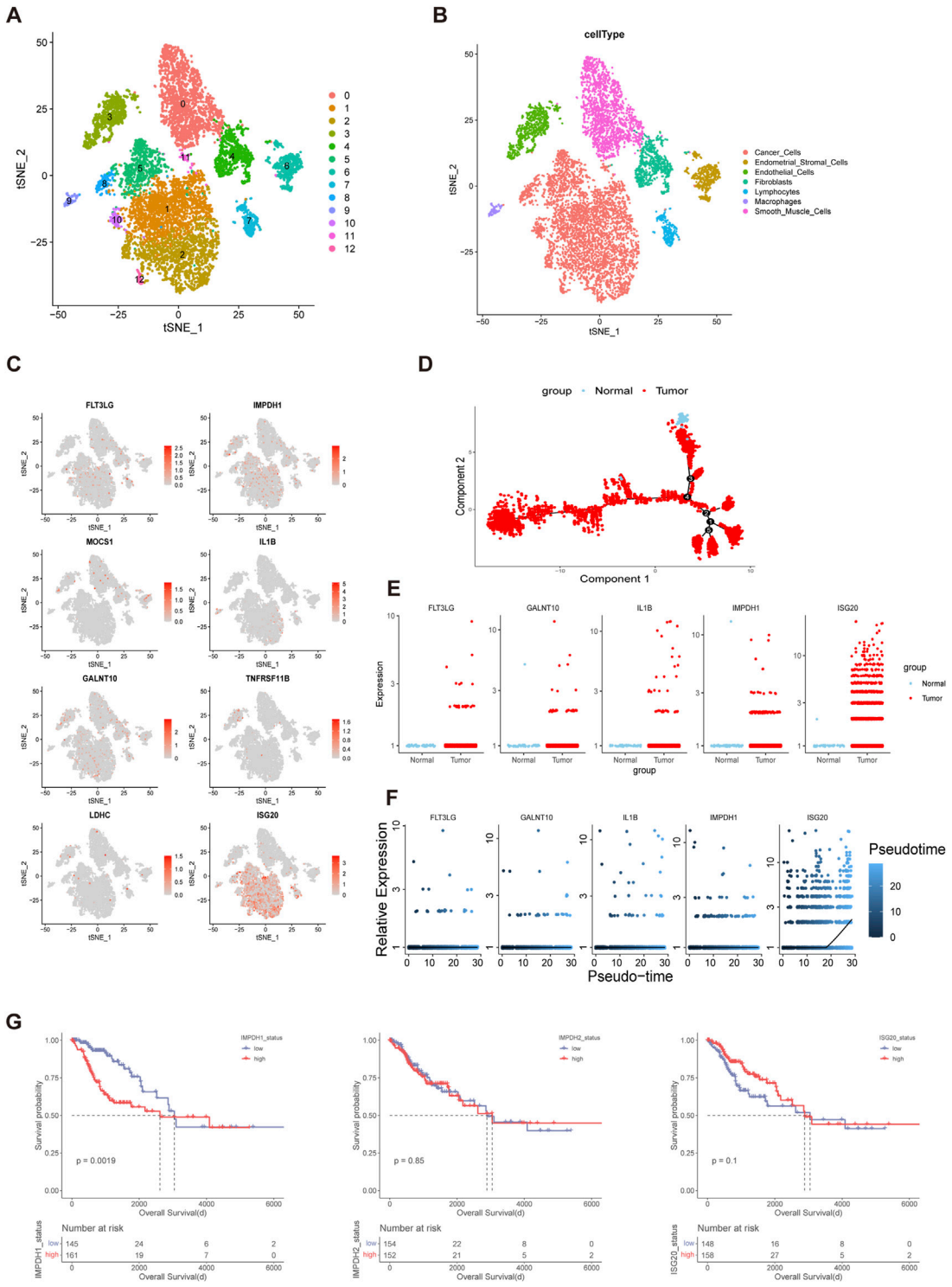


FIGURE 5
IMPDH1 was A Significant Prognostic Risk Gene in Cervical Cancer **(A,B)** The t-distributed stochastic neighbor embedding (t-SNE) plot demonstrating the main cell clusters in cervical cancer and identification of the main cell types. **(C)** Heatmap shows the expression of genes in the risk models. **(D)** Development trajectory plot of cervical cancer cells in pseudo-time analysis. **(E)** Differential expression of the genes involved in the risk model between the normal cervix cells and cervical cancer cells. **(F)** Pseudo-time analysis showing the expression patterns of the genes in the risk model along the tumor progression. **(G)** Kaplan-Meier curves of overall survival of the cohorts with low and high expression level of IMPDH1, IMPDH2 and ISG20 in TCGA.

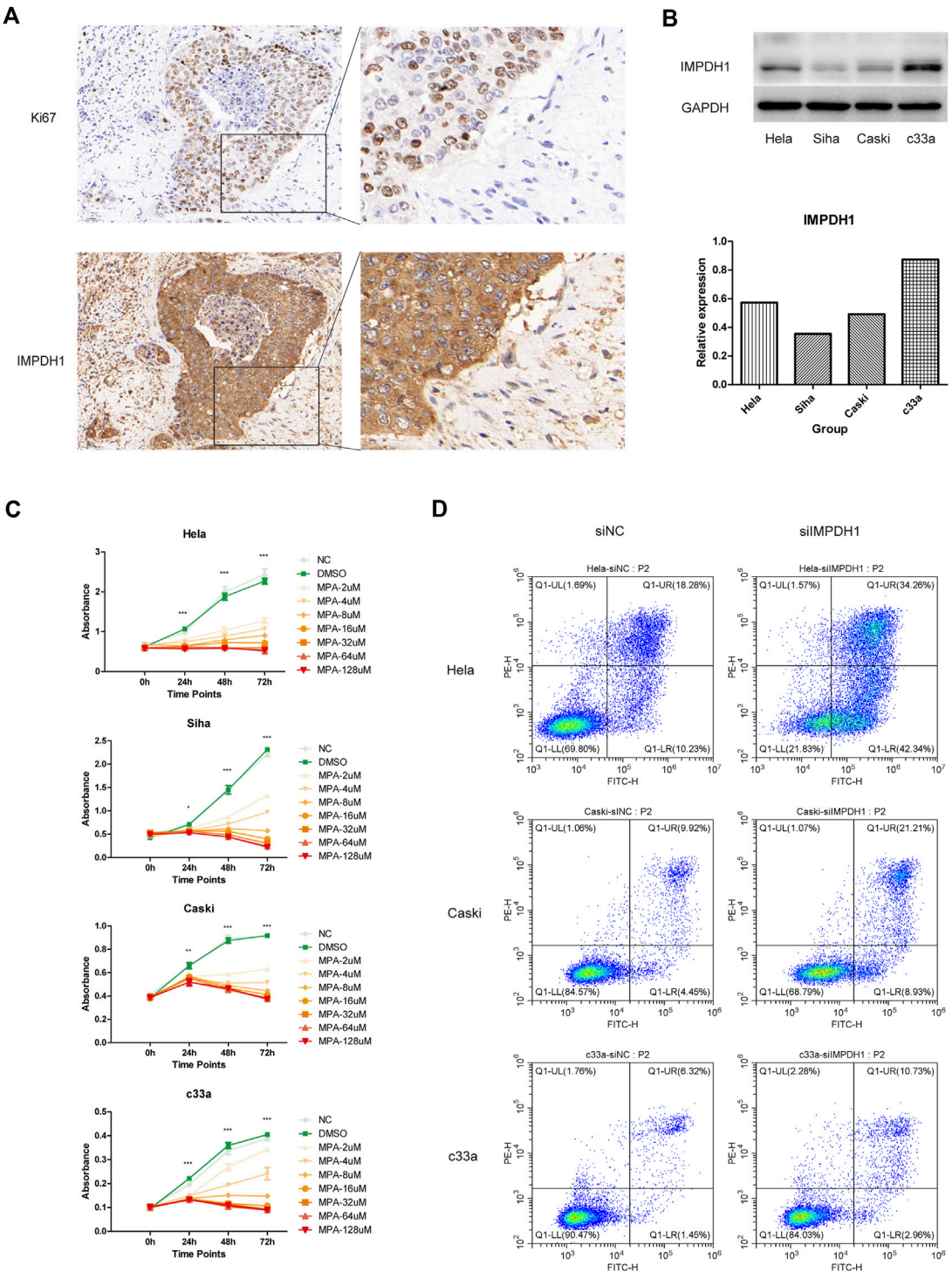


FIGURE 6
In vitro validation **(A)** IMPDH1 protein expression in normal cervix tissue and cervical cancer determined using immunohistochemistry. **(B)** Different expression level of IMPDH1 in Hela, Siha, Caski and c33a cell lines. **(C)** CCK8 assay of cell viability under intervention at different concentrations of MPA. **(D)** Flow cytometry analysis of apoptosis rate of cells transfected with siRNA. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

In vitro validation

It was found at single cell resolution that IMPDH1 was mainly expressed in cervical cancer cells. Therefore we validated our finding on clinical specimens with immunohistochemistry. Compared with normal cervix tissue, the expression of IMPDH1 in cervical cancer was significantly increased, suggesting IMPDH1 may contribute to the progression of cervical cancer (Figure 6A). Then we examined the expression level of IMPDH1 in four cervical cancer cell lines Hela, Caski, c33a and Siha with western blotting. It was shown that 4 cell lines have

different expression level (Figure 6B). Mycophenolic acid (MPA) was a pan-inhibitor of IMPDH and it was demonstrated that targeting IMPDH with MPA can significantly inhibit growth of the ASCL1^{low} small cell lung cancer cell (Huang et al., 2018). In our research, we found that the cell viability of cervical cancer cell lines could also be significantly inhibited with MPA, which indicated IMPDH could be targeted in cervical cancer (Figure 6C). In addition, siRNA was used to knock down the expression of IMPDH1 in Hela, Caski and c33a cell lines. Western blotting and qRT-PCR confirmed the knockdown efficiency (Supplementary Figures 3A–3C). With flow cytometry, we found that the apoptosis rate was higher in cells transfected with siRNA (Figure 6D). Taken together, IMPDH1 may contribute to the growth of cervical cancer and it can be a novel therapeutic target in cervical cancer.

Discussion

In the present study, we identified three molecular subgroups with distinct prognosis in cervical cancer from immuno-metabolic perspective. The differences in tumor infiltrating lymphocytes and metabolic characteristics among the three subgroups were evaluated with Cibersort algorithm and GSVA algorithm respectively. In addition, a risk model based on immune and metabolism related genes was constructed for prognosis prediction in cervical cancer, which showed higher accuracy than current FIGO stage in the validation cohort. With the single cell sequencing data from Hua's research (Li et al., 2021), we explored the hub genes' cellular localization and their expression patterns during cancer progression. Finally, we found that IMPDH1 may be a key gene in tumorigenesis, whose expression may help shape the TME and promote tumor progression.

Tumor heterogeneity is an important factor affecting the survival and prognosis of patients with cancer. Molecular stratification has been applied to multiple malignancies in order to help inform appropriate clinical decisions, including prostate cancer (Tang et al., 2022), breast cancer (Wolf et al., 2022), hepatocellular carcinoma (Molina et al., 2022), intrahepatic cholangiocarcinoma (Martin-Serrano et al., 2022) and small cell lung cancer (Rudin et al., 2019). It was demonstrated that multiple molecular subtypes of cervical cancer may be uncovered from different aspects (Meijer and Steenbergen, 2017). Integration analysis from The Cancer Genome Atlas Research Network revealed different molecular features of cervical cancer, which may help personalize clinical management (Cancer Genome Atlas Research Network et al., 2017). Maud Kamal et al. discovered the different integration signatures of HPV genome in cervical cancer that may imply prognostic significance (Kamal et al., 2021). With the 50 genes having the largest expression variation, Xiaojun Zhu et al. demonstrated two molecular subgroups in cervical cancer and explored the heterogeneity, which provided novel targets for diagnosis and treatment (Zhu et al., 2022). Here we provided a novel classification approach to dissect the heterogeneity of cervical cancer from the immunological and metabolic perspectives. Three immuno-metabolic subtypes were identified in cervical cancer with significantly different prognosis. The immune infiltration status was poorer in C2 subgroup than the other two subgroups and C2 had the worst prognosis. The expression level of immune checkpoint molecules was higher in the C3 subgroup, which suggested better response to immune checkpoint blockade therapy. Due to the limited number of samples, there was no significant prognostic difference between C1 and C3 subgroups but they showed distinct metabolism status, indicating novel therapeutic targets for tumor metabolism. Therefore, this stratification strategy may contribute to individualized treatment for cervical cancer.

Immune microenvironment was one of the most important facets of tumorigenesis (Schreiber et al., 2011). It was demonstrated that the abundance of tumor infiltrated lymphocytes (TILs) was associated with favorable patient prognosis (Jérôme et al., 2006). Consistent with the existing evidence, C2 subgroup showed the least immune cells abundance and therefore the prognosis was worst among the three subgroups. The phenotype of TILs can be shaped towards immunosuppressed by tumor cells during cancer progression. For example, macrophages infiltrating the tumor can be induced to differentiate into the M2 phenotype that tends to suppress immune response (Luca and Pollard Jeffrey, 2018). In our study, macrophages in C1 subgroup were characterized with anti-inflammatory phenotype while macrophages in C3 subgroup were pro-inflammatory. In terms of PFI and DSS, C3 subgroup showed better prognosis than C1 subgroup though the difference was not statistically significant. Therefore, we speculated that the status of immune cells may determine the prognosis of cervical cancer. It was shown that the lactate derived from tumor cells can influence the phenotype of macrophages in lung cancer and melanoma (Colegio et al., 2014). Our research showed that the glycolysis level is rather higher in C1 subgroup, which may explain the difference of macrophage differentiation between C1 and C3 subgroups. Besides, activated mast cells were found abundant in C2 subgroup and they indicated poor prognosis in the whole cohort, which was consistent with the existing evidence (Huang et al., 2008). Different expression level of TLRs was detected among the three subgroups, which indicated that distinct status of innate immunity against HPV may exist in cervical cancer patients and correlate with prognosis. Due to the data limitation, we cannot distinguish the expression level or polymorphism of TLRs in cancer cells in TCGA cohort and therefore the molecular mechanism by which TLRs promote cervical cancer progression needs further experimental research.

Distinct metabolic patterns were uncovered among the three immuno-metabolic subgroups. Compared with the other two subgroups, C2, the subgroup with the worst prognosis, was characterized with terpenoid backbone biosynthesis and aminoacyl-tRNA biosynthesis. 3-Hydroxy-3-methylglutaryl-CoA synthase 1 (HMGCS1), a metabolic enzyme that participated in terpenoid backbone biosynthesis, was demonstrated to involve in the progression of cervical cancer (Zhang et al., 2020a). Besides, Li et al. (2022) showed that isoprenylcysteine carboxyl methyltransferase (ICMT) may mediate the malignant development of cervical carcinoma. Other genes in terpenoid backbone biosynthesis pathway have been demonstrated to play a role in tumorigenesis of breast cancer (Yu et al., 2021), prostate cancer (Seshacharyulu et al., 2019) and renal cell carcinoma (Huang et al., 2021). During protein synthesis, aminoacyl-tRNA biosynthesis was an Indispensable pathway catalyzed by 20 essential enzymes that ligate the amino acids to their corresponding tRNAs and there was evidence that the synthetases may cause diseases when they were mutated or expressed abnormally (Kwon et al., 2019). For example, the expression of glycyl-tRNA synthetase (GRS) was found to be an indicator of unfavorable outcomes in renal, urothelial, liver, breast and endometrial cancers (Thul and Lindskog, 2018). In our research, we found a correlation between the metabolism pathway mentioned above and the rarity of TILs. Therefore we speculated that metabolism related genes may impact the abundance and status of TILs in addition to their involvement in tumor metabolic reprogramming, which deserves further exploration in cervical cancer.

Ten genes were used for the construction of the prognostic model, which included FLT3LG, IMPDH1, OPRD1, MOCS1, IL1B, GALNT10, TNFRSF11B, LDHC, ISG20 and TRAV12-3. This model showed better predictive power than FIGO stage system, suggesting its potential for clinical application. FLT3LG, the formative cytokine for cDC1, was shown to be capable of controlling the levels of type I conventional dendritic cells in TME and increasing the responsiveness of patients to anti-PD-1 immunotherapy (Barry et al., 2018). In contrast, we validated the expression levels of FLT3LG at single cell resolution and found that it was upregulated mainly in cancer cells, which indicated its distinct function in cervical cancer. IMPDH1 is a rate-limiting enzyme of guanosine triphosphate (GTP) *de novo* synthesis and it can also form a filamentous structure called cytoophidia (Keppeke et al., 2020). According to Ruan et al. (2020b)'s research, cytoophidia formed by IMPDH1 may contribute to the metastasis of clear cell renal cell carcinoma. Consistently, IMPDH1, but not IMPDH2, was found to be an indicator of poor prognosis in our cohort, suggesting that cytoophidia may be a therapeutic target in cervical cancer. Dysregulated Inflammation cytokine can exacerbate tumor development. MOCS1 is a gene involved in the molybdenum cofactor biosynthesis pathway (Reiss and Hahnewald, 2011) but little was known about its function in malignancies. IL1B was one of the IL-1 family proteins and it was demonstrated to be a therapeutic target in cancer (de Mooij et al., 2017). Altered glycosylation was found to occur in malignancy (Oliveira-Ferrer et al., 2017) and GALNT10 is one of the glycosyltransferases whose expression was associated with poor prognosis in high grade ovarian serous cancer (Zhang et al., 2020b). TNFRSF11B, also termed osteoprotegerin, was demonstrated to involve in the progression of gastric cancer (Luan et al., 2020), melanomas (Oliver et al., 2013) and colon cancer (Zhang et al., 2021) but little was known about its function in cervical cancer. LDHC is one of the isozymes in lactate dehydrogenase family which catalyzes the interconversion of pyruvate and l-lactate (Markert et al., 1975). Remy Thomas et al. demonstrated that LDHC could be a targetable cancer antigen for cancer immunotherapy (Thomas et al., 2020). ISG20 is a 20 kDa protein that was capable of inhibiting multiple viruses and its expression was shown to contribute to poor survival in glioma (Gao et al., 2019). Surprisingly, some genes in the model were not detected at single cell resolution such as OPRD1, a gene encoding the delta-opioid receptor and TRAV12-3, a gene encoding the T cell receptor alpha variable, which may be due to the sample heterogeneity of single cell sequencing. Besides, TRAV12-3 served as a protective factor in the model and it suggested a role for robust immunity in preventing HPV-related tumors.

There were some limitations existing in our research. Firstly, the identification of tumor subtypes was based on the analysis of the public data without further exploration in experiment or clinical investigation. Second, with limited data resources, the construction and verification of our prognostic model were performed in different parts of one cohort, which needs further validation in external cohort. Finally, bioinformatics methods were used to evaluate the prognostic prediction power of 10 key genes in cervical cancer but their underlying molecular mechanisms in the tumorigenesis and progression deserve further research. Though the clinical specimens and experimental data showed that IMPDH1 can be a therapeutic target in cervical cancer, more detailed mechanism needs further research.

Taken together, our research provided a novel perspective for molecular stratification in cervical cancer. Distinct metabolic patterns were found in three subgroups and the correlation between TILs and metabolic pathways were explored. Then a risk model for prognostic prediction was constructed and it showed better performance than clinical stage. Immune-metabolism risk score exhibited unfavorable prognostic significance. Finally, we verified the genes in model at single cell resolution to figure out the cellular localization and found that five genes were significantly upregulated in cervical cancer cells. These findings will help adjust the management strategy for cervical cancer according to their heterogeneity and uncover novel targets for cancer 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 Medical Ethics Committee of Zhujiang Hospital, Southern Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

WL and JL conceived and designed the study. XXL and LH were responsible for data collection and analysis. Manuscript was written by WL. KH and XL contributed to the design of experiments and data analysis. PL contributed to implementation of experiments. KH, XL, and YW contributed to the revision of the manuscript. WL, JL, XXL, PL, and LH contributed equally to this article. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by grants from the Guangdong Basic and Applied Basic Research Foundation (2021A1515011638).

Acknowledgments

We thank 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/fgene.2023.1067666/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Results of unsupervised consensus clustering **(A)** Consensus matrix legend plot indicating the cluster probability ranging from 0 to 1 with different colors. **(B–D)** Consensus matrix heatmap of different k value (*k*=2, *k*=3, and *k*=4 respectively). **(E)** Cumulative distribution function (CDF) plot of consensus clustering. **(F)** Delta area plot of consensus clustering.

SUPPLEMENTARY FIGURE S2
Expression of TLRs and cytotoxicity related molecules and signaling. **(A-C)** Boxplot of the expression of TLR2, TLR3 and TLR4 among three subtypes. The differences were compared using the Kruskal-Wallis test. **(D–E)** Boxplot of the expression of grazyme B and IFN-γ among three subtypes. The differences were compared using the Kruskal-Wallis test. **(F)** Boxplot of the enrichment score of IFN-γ signaling among three subtypes. The differences were compared using the Kruskal-Wallis test.

SUPPLEMENTARY FIGURE S3
Validation of the knockdown efficiency of siRNA **(A-C)** Results of qRT-PCR and western blotting confirmed the knockdown efficiency of siRNA targeting IMPDH1.

References

Barry, K. C., Hsu, J., Broz, M. L., Cueto, F. J., Binnewies, M., Combes, A. J., et al. (2018). A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments. *Nat. Med.* 24 (8), 1178–1191. doi:10.1038/s41591-018-0085-8

Brandacher, G., Perathoner, A., Ladurner, R., Schneeberger, S., Obrist, P., Winkler, C., et al. (2006). Prognostic value of indoleamine 2,3-dioxygenase expression in colorectal cancer: Effect on tumor-infiltrating T cells. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 12 (4), 1144–1151. doi:10.1158/1078-0432.ccr-05-1966

Cancer Genome Atlas Research Network; Albert Einstein College of Medicine; Analytical Biological Services; Barretos Cancer Hospital; Baylor College of Medicine; Beckman Research Institute of City of Hope, et al. (2017). Integrated genomic and molecular characterization of cervical cancer. *Nature* 543 (7645), 378–384. doi:10.1038/nature21386

Carr, E. L., Kelman, A., Wu, G. S., Gopaul, R., Senkevitch, E., Aghvanyan, A., et al. (2010). Glutamine uptake and metabolism are coordinately regulated by ERK/MAPK during T lymphocyte activation. *J. Immunol.* 185 (2), 1037–1044. (Baltimore, Md : 1950). doi:10.4049/jimmunol.0903586

Cascone, T., McKenzie, J. A., Mbofung, R. M., Punt, S., Wang, Z., Xu, C., et al. (2018). Increased tumor glycolysis characterizes immune resistance to adoptive T cell therapy. *Cell. metab.* 27 (5), 977–987. doi:10.1016/j.cmet.2018.02.024

Chang, C. H., Qiu, J., O'Sullivan, D., Buck, M. D., Noguchi, T., Curtis, J. D., et al. (2015). Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell.* 162 (6), 1229–1241. doi:10.1016/j.cell.2015.08.016

Cheng, Y., Zhu, Y., Xu, J., Yang, M., Chen, P., Xu, W., et al. (2018). PKN2 in colon cancer cells inhibits M2 phenotype polarization of tumor-associated macrophages via regulating DUSP6-Erk1/2 pathway. *Mol. cancer* 17 (1), 13. doi:10.1186/s12943-017-0747-z

Colegio, O. R., Chu, N. Q., Szabo, A. L., Chu, T., Rhebergen, A. M., Jairam, V., et al. (2014). Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. *Nature* 513 (7519), 559–563. doi:10.1038/nature13490

Coussens, L. M., Zitvogel, L., and Palucka, A. K. (2013). Neutralizing tumor-promoting chronic inflammation: A magic bullet? *Sci. (New York, NY)* 339 (6117), 286–291. doi:10.1126/science.1232227

de Mooij, C. E. M., Netea, M. G., van der Velden, W. J. F. M., and Blijlevens, N. M. A. (2017). Targeting the interleukin-1 pathway in patients with hematological disorders. *Blood* 129 (24), 3155–3164. doi:10.1182/blood-2016-12-754994

Delyon, J., Servy, A., Laugier, F., André, J., Ortonne, N., Battistella, M., et al. (2017). PDE4D promotes FAK-mediated cell invasion in BRAF-mutated melanoma. *Oncogene* 36 (23), 3252–3262. doi:10.1038/onc.2016.469

Gao, M., Lin, Y., Liu, X., Li, Y., Zhang, C., Wang, Z., et al. (2019). ISG20 promotes local tumor immunity and contributes to poor survival in human glioma. *Oncoimmunology* 8 (2), e1534038. doi:10.1080/2162402x.2018.1534038

Hanahan, D., and Coussens, L. M. (2012). Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell.* 21 (3), 309–322. doi:10.1016/j.ccr.2012.02.022

Huang, Bo, Zhang, Lei, Zhang, Gui-Mei, Dong, Li, Song, Chuanwang, Bo, Li, et al. (2008). SCF-mediated mast cell infiltration and activation exacerbate the inflammation and immunosuppression in tumor microenvironment. *Blood* 112 (4), 1269–1279. doi:10.1182/blood-2008-03-147033

Huang, F., Ni, M., Chalishazar, M. D., Huffman, K. E., Kim, J., Cai, L., et al. (2018). Inosine monophosphate dehydrogenase dependence in a subset of small cell lung cancers. *Cell. metab.* 28 (3), 369–382. doi:10.1016/j.cmet.2018.06.005

Huang, J., Zhao, X., Li, X., Peng, J., Yang, W., and Mi, S. (2021). HMGCR inhibition stabilizes the glycolytic enzyme PKM2 to support the growth of renal cell carcinoma. *PLoS Biol.* 19 (4), e3001197. doi:10.1371/journal.pbio.3001197

Hung, M. H., Lee, J. S., Ma, C., Diggs, L. P., Heinrich, S., Chang, C. W., et al. (2021). Tumor methionine metabolism drives T-cell exhaustion in hepatocellular carcinoma. *Nat. Commun.* 12 (1), 1455. doi:10.1038/s41467-021-21804-1

Ichimura, K. (2012). Molecular pathogenesis of IDH mutations in gliomas. *Brain tumor pathol.* 29 (3), 131–139. doi:10.1007/s10014-012-0090-4

Ikeda, A., Shimizu, T., Matsumoto, Y., Fujii, Y., Eso, Y., Inuzuka, T., et al. (2014). Leptin receptor somatic mutations are frequent in HCV-infected cirrhotic liver and associated with hepatocellular carcinoma. *Gastroenterology* 146 (1), 222–232. doi:10.1053/j.gastro.2013.09.025

Inaba, T., Ino, K., Kajiyama, H., Yamamoto, E., Shibata, K., Nawa, A., et al. (2009). Role of the immunosuppressive enzyme indoleamine 2,3-dioxygenase in the progression of ovarian carcinoma. *Gynecol. Oncol.* 115 (2), 185–192. doi:10.1016/j.ygyno.2009.07.015

Ino, K., Yamamoto, E., Shibata, K., Kajiyama, H., Yoshida, N., Terauchi, M., et al. (2008). Inverse correlation between tumoral indoleamine 2,3-dioxygenase expression and tumor-infiltrating lymphocytes in endometrial cancer: Its association with disease progression and survival. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 14 (8), 2310–2317. doi:10.1158/1078-0432.ccr-07-4144

Jenkins, M. R., and Griffiths, G. M. (2010). The synapse and cytolytic machinery of cytotoxic T cells. *Curr. Opin. Immunol.* 22 (3), 308–313. doi:10.1016/j.coi.2010.02.008

Jérôme, G., Anne, C., Fatima, S-C., Amos, K., Bernhard, M., Christine, L-P., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Sci. (New York, NY)* 313 (5795), 1960–1964. doi:10.1126/science.1129139

Jiang, Y., Yuan, Y., Chen, M., Li, S., Bai, J., Zhang, Y., et al. (2021). PRMT5 disruption drives antitumor immunity in cervical cancer by reprogramming T cell-mediated response and regulating PD-L1 expression. *Theranostics* 11 (18), 9162–9176. doi:10.7150/thno.59605

Kamal, M., Lameiras, S., Deloger, M., Morel, A., Vacher, S., Lecerf, C., et al. (2021). Human papilloma virus (HPV) integration signature in cervical cancer: Identification of MACROD2 gene as HPV hot spot integration site. *Br. J. cancer* 124 (4), 777–785. doi:10.1038/s41416-020-01153-4

Kao, K. C., Vilbois, S., Tsai, C. H., and Ho, P. C. (2022). Metabolic communication in the tumour-immune microenvironment. *Nat. Cell. Biol.* 24 (11), 1574–1583. doi:10.1038/s41556-022-01002-x

Keppeke, G. D., Andrade, L. E. C., Barcelos, D., Fernandes, M., and Landman, G. (2020). IMPDH-based cytoophidium structures as potential theranostics in cancer. *Mol. Ther. J. Am. Soc. Gene Ther.* 28 (7), 1557–1558. doi:10.1016/j.ymthe.2020.06.006

Kohanbash, G., Carrera, D. A., Shrivastav, S., Ahn, B. J., Jahan, N., Mazor, T., et al. (2017). Isocitrate dehydrogenase mutations suppress STAT1 and CD8+ T cell accumulation in gliomas. *J. Clin. investigation* 127 (4), 1425–1437. doi:10.1172/jci90644

Koppenol, W. H., Bounds, P. L., and Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* 11 (5), 325–337. doi:10.1038/nrc3038

Kupets, R., and Covens, A. (2001). Is the international federation of Gynecology and Obstetrics staging system for cervical carcinoma able to predict survival in patients with cervical carcinoma?: An assessment of clinimetric properties. *Cancer* 92 (4), 796–804. doi:10.1002/1097-0142(20010815)92:4<796::aid-cncr1385>3.0.co;2-3

Kwon, N. H., Fox, P. L., and Kim, S. (2019). Aminoacyl-tRNA synthetases as therapeutic targets. *Nat. Rev. Drug Discov.* 18 (8), 629–650. doi:10.1038/s41573-019-0026-3

Li, W., Tanikawa, T., Kryczek, I., Xia, H., Li, G., Wu, K., et al. (2018). Aerobic glycolysis controls myeloid-derived suppressor cells and tumor immunity via a specific CEBPB isoform in triple-negative breast cancer. *Cell. metab.* 28 (1), 87–103. doi:10.1016/j.cmet.2018.04.022

Li, X., Wenes, M., Romero, P., Huang, S. C., Fendt, S. M., and Ho, P. C. (2019). Navigating metabolic pathways to enhance antitumour immunity and immunotherapy. *Nat. Rev. Clin. Oncol.* 16 (7), 425–441. doi:10.1038/s41571-019-0203-7

Li, C., Guo, L., Li, S., and Hua, K. (2021). Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and transcriptional activities of ECs in CC. *Mol. Ther. Nucleic acids* 24, 682–694. doi:10.1016/j.omtn.2021.03.017

Li, Y., Gao, X., Yang, C., Yan, H., and Li, C. (2022). CircRNA hsa_circ_0018289 exerts an oncogenic role in cervical cancer progression through miR-1294/ICMT axis. *J. Clin. laboratory analysis* 36 (5), e24348. doi:10.1002/jcla.24348

Liu, K., Cui, J. J., Zhan, Y., Ouyang, Q. Y., Lu, Q. S., Yang, D. H., et al. (2022). Reprogramming the tumor microenvironment by genome editing for precision cancer therapy. *Mol. cancer* 21 (1), 98. doi:10.1186/s12943-022-01561-5

Luan, F., Li, X., Cheng, X., Huangfu, L., Han, J., Guo, T., et al. (2020). TNFRSF11B activates Wnt/β-catenin signaling and promotes gastric cancer progression. *Int. J. Biol. Sci.* 16 (11), 1956–1971. doi:10.7150/ijbs.43630

Luca, C., and Pollard Jeffrey, W. (2018). Targeting macrophages: Therapeutic approaches in cancer. *Nat. Rev. Drug Discov.* 17 (12), 887–904. doi:10.1038/nrd.2018.169

Mantovani, A., and Sica, A. (2010). Macrophages, innate immunity and cancer: Balance, tolerance, and diversity. *Curr. Opin. Immunol.* 22 (2), 231–237. doi:10.1016/j.coi.2010.01.009

Marincola, F. M., Jaffee, E. M., Hicklin, D. J., and Ferrone, S. (2000). Escape of human solid tumors from T-cell recognition: Molecular mechanisms and functional significance. *Adv. Immunol.* 74, 181–273. doi:10.1016/s0065-2776(08)60911-6

Markert, C. L., Shaklee, J. B., and Whitt, G. S. (1975). Evolution of a gene. Multiple genes for LDH isozymes provide a model of the evolution of gene structure, function and regulation. *Sci. (New York, NY)* 189 (4197), 102–114. doi:10.1126/science.1138367

Martin-Serrano, M. A., Kepecs, B., Torres-Martin, M., Bramel, E. R., Haber, P. K., Merritt, E., et al. (2022). Novel microenvironment-based classification of intrahepatic cholangiocarcinoma with therapeutic implications. *Gut*. doi:10.1136/gutjnl-2021-326514

Martínez-Reyes, I., and Chandel, N. S. (2021). Cancer metabolism: Looking forward. *Nat. Rev. Cancer* 21 (10), 669–680. doi:10.1038/s41568-021-00378-6

Meijer, C. J. L. M., and Steenbergen, R. D. M. (2017). Gynaecological cancer: Novel molecular subtypes of cervical cancer - potential clinical consequences. *Nat. Rev. Clin. Oncol.* 14 (7), 397–398. doi:10.1038/nrclinonc.2017.52

Mishra, R. R., Belder, N., Ansari, S. A., Kayhan, M., Bal, H., Raza, U., et al. (2018). Reactivation of cAMP pathway by PDE4D inhibition represents a novel druggable Axis for overcoming tamoxifen resistance in ER-positive breast cancer. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 24 (8), 1987–2001. doi:10.1158/1078-0432.ccr-17-2776

Molina, L., Zhu, J., Trépo, E., Bayard, Q., Amaddeo, G., Calderaro, J., et al. (2022). Biallelic hydroxymethylbilane synthase inactivation defines a homogenous clinico-molecular subtype of hepatocellular carcinoma. *J. hepatology* 77, 1038–1046. doi:10.1016/j.jhep.2022.05.018

Motz, G. T., and Coukos, G. (2013). Deciphering and reversing tumor immune suppression. *Immunity* 39 (1), 61–73. doi:10.1016/j.immuni.2013.07.005

Munn, D. H., Sharma, M. D., Baban, B., Harding, H. P., Zhang, Y., Ron, D., et al. (2005). GCN2 kinase in T cells mediates proliferative arrest and anergy induction in response to indoleamine 2,3-dioxygenase. *Immunity* 22 (5), 633–642. doi:10.1016/j.immuni.2005.03.013

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Oliveira-Ferrer, L., Legler, K., and Milde-Langosch, K. (2017). Role of protein glycosylation in cancer metastasis. *Seminars cancer Biol.* 44, 141–152. doi:10.1016/j.semcancer.2017.03.002

Oliver, J. L., Alexander, M. P., Norrod, A. G., Mullins, I. M., and Mullins, D. W. (2013). Differential expression and tumor necrosis factor-mediated regulation of TNFRSF11b/osteoprotegerin production by human melanomas. *Pigment Cell. & melanoma Res.* 26 (4), 571–579. doi:10.1111/pcmr.12091

Pandey, S., Rama Devi, M., Srivastava, M., Srivastava, K., Singh, S., Srivastava, S., et al. (2009). Impact of Toll-like receptors [TLR] 2 (-196 to -174 del) and TLR 4 (Asp299Gly, Thr399Ile) in cervical cancer susceptibility in North Indian women. *Gynecol. Oncol.* 114 (3), 501–505. doi:10.1016/j.ygyno.2009.05.032

Pandey, S., Balraj, M., Srivastava, M., Singh, S., Srivastava, K., Lal, P., et al. (2011). Evaluation of Toll-like receptors 3 (c.1377C/T) and 9 (G2848A) gene polymorphisms in cervical cancer susceptibility. *Mol. Biol. Rep.* 38 (7), 4715–4721. doi:10.1007/s11033-010-0607-z

Pilotte, L., Larrieu, P., Stroobant, V., Colau, D., Dolusic, E., Frédérick, R., et al. (2012). Reversal of tumoral immune resistance by inhibition of tryptophan 2,3-dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 109 (7), 2497–2502. doi:10.1073/pnas.1113873109

Platten, M., Nollen, E. A. A., Röhrig, U. F., Fallarino, F., and Opitz, C. A. (2019). Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat. Rev. Drug Discov.* 18 (5), 379–401. doi:10.1038/s41573-019-0016-5

Punt, C. J., Koopman, M., and Vermeulen, L. (2017). From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat. Rev. Clin. Oncol.* 14 (4), 235–246. doi:10.1038/nrclinonc.2016.171

Qian, B., Deng, Y., Im, J. H., Muschel, R. J., Zou, Y., Li, J., et al. (2009). A distinct macrophage population mediates metastatic breast cancer cell extravasation, establishment and growth. *PloS one* 4 (8), e6562. doi:10.1371/journal.pone.0006562

Qian, B. Z., Li, J., Zhang, H., Kitamura, T., Zhang, J., Campion, L. R., et al. (2011). CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* 475 (7355), 222–225. doi:10.1038/nature10138

Quail, D. F., and Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* 19 (11), 1423–1437. doi:10.1038/nm.3394

Rahma, O. E., and Hodi, F. S. (2019). The intersection between tumor angiogenesis and immune suppression. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 25 (18), 5449–5457. doi:10.1158/1078-0432.ccr-18-1543

Rahrmann, E. P., Collier, L. S., Knutson, T. P., Doyal, M. E., Kuslak, S. L., Green, L. E., et al. (2009). Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen. *Cancer Res.* 69 (10), 4388–4397. doi:10.1158/0008-5472.can-08-3901

Reiss, J., and Hahnewald, R. (2011). Molybdenum cofactor deficiency: Mutations in GPHN, MOCS1, and MOCS2. *Hum. Mutat.* 32 (1), 10–18. doi:10.1002/humu.21390

Ruan, H., Song, Z., Cao, Q., Ni, D., Xu, T., Wang, K., et al. (2020). IMPDH1/YB-1 positive feedback loop assembles cytoophidia and represents a therapeutic target in metastatic tumors. *Mol. Ther. J. Am. Soc. Gene Ther.* 28 (5), 1299–1313. doi:10.1016/j.ymthe.2020.03.001

Ruan, H., Song, Z., Qi, C., Ni, D., Xu, T., Wang, K., et al. (2020). IMPDH1/YB-1 positive feedback loop assembles cytoophidia and represents a therapeutic target in metastatic tumors. *Mol. Ther.* 28 (5), 1299–1313. doi:10.1016/j.ymthe.2020.03.001

Rudin, C. M., Poirier, J. T., Byers, L. A., Dive, C., Dowlati, A., George, J., et al. (2019). Molecular subtypes of small cell lung cancer: A synthesis of human and mouse model data. *Nat. Rev. Cancer* 19 (5), 289–297. doi:10.1038/s41568-019-0133-9

Schreiber, R. D., Old, L. J., and Smyth, M. J. (2011). Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. *Sci. (New York, NY)* 331 (6024), 1565–1570. doi:10.1126/science.1203486

Seshacharyulu, P., Rachagani, S., Muniyan, S., Siddiqui, J. A., Cruz, E., Sharma, S., et al. (2019). FDPS cooperates with PTEN loss to promote prostate cancer progression through modulation of small GTPases/AKT axis. *Oncogene* 38 (26), 5265–5280. doi:10.1038/s41388-019-0791-9

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics, 2022. *CA a cancer J. Clin.* 72 (1), 7–33. doi:10.3322/caac.21708

Sobo-Vujanovic, A., Vujanovic, L., DeLeo, A. B., Concha-Benavente, F., Ferris, R. L., Lin, Y., et al. (2016). Inhibition of soluble tumor necrosis factor prevents chemically induced carcinogenesis in mice. *Cancer Immunol. Res.* 4 (5), 441–451. doi:10.1158/2326-6066.cir-15-0104

Speiser, D. E., Ho, P. C., and Verdeil, G. (2016). Regulatory circuits of T cell function in cancer. *Nat. Rev. Immunol.* 16 (10), 599–611. doi:10.1038/nri.2016.80

Tang, F., Xu, D., Wang, S., Wong, C. K., Martinez-Fundichely, A., Lee, C. J., et al. (2022)., 376. New York, NY), eabe1505. eabe1505. doi:10.1126/science.abe1505Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets*Science*6596

Thomas, R., Shaath, H., Naik, A., Toor, S. M., Elkord, E., and Decock, J. (2020). Identification of two HLA-A*0201 immunogenic epitopes of lactate dehydrogenase C (LDHC): Potential novel targets for cancer immunotherapy. *Cancer Immunol. Immunother.* 69 (3), 449–463. doi:10.1007/s00262-020-02480-4

Thul, P. J., and Lindskog, C. (2018). The human protein atlas: A spatial map of the human proteome. *Protein Sci. a Publ. Protein Soc.* 27 (1), 233–244. doi:10.1002/pro.3307

Uddin, S., Bu, R., Ahmed, M., Abubaker, J., Al-Dayel, F., Bavi, P., et al. (2009). Overexpression of leptin receptor predicts an unfavorable outcome in Middle Eastern ovarian cancer. *Mol. cancer* 8, 74. doi:10.1186/1476-4598-8-74

Ugel, S., De Sanctis, F., Mandruzzato, S., and Bronte, V. (2015). Tumor-induced myeloid deviation: When myeloid-derived suppressor cells meet tumor-associated macrophages. *J. Clin. investigation* 125 (9), 3365–3376. doi:10.1172/jci80006

Urabe, F., Kosaka, N., Sawa, Y., Yamamoto, Y., Ito, K., Yamamoto, T., et al. (2020). miR-26a regulates extracellular vesicle secretion from prostate cancer cells via targeting SHC4, PFDN4, and CHORDC1. *Sci. Adv.* 6 (18), eaay3051. doi:10.1126/sciadv.aay3051

van der Burg, S. H., Arens, R., Ossendorp, F., van Hall, T., and Melief, C. J. (2016). Vaccines for established cancer: Overcoming the challenges posed by immune evasion. *Nat. Rev. Cancer* 16 (4), 219–233. doi:10.1038/nrc.2016.16

Wang, Y. M., Qiu, J. J., Qu, X. Y., Peng, J., Lu, C., Zhang, M., et al. (2022). Accumulation of dysfunctional tumor-infiltrating PD-1+ DCs links PD-1/PD-L1 blockade

immunotherapeutic response in cervical cancer. *Oncoimmunology* 11 (1), 2034257. doi:10.1080/2162402x.2022.2034257

Wolf, D. M., Yau, C., Wulfkuhle, J., Brown-Swigart, L., Gallagher, R. I., Lee, P. R. E., et al. (2022). Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell.* 40 (6), 609–623.e6. doi:10.1016/j.ccell.2022.05.005

Yu, T. J., Liu, Y. Y., Li, X. G., Lian, B., Lu, X. X., Jin, X., et al. (2021). PDSS1-Mediated activation of camk2a-STAT3 signaling promotes metastasis in triple-negative breast cancer. *Cancer Res.* 81 (21), 5491–5505. doi:10.1158/0008-5472.can-21-0747

Zhang, J., Jiang, M., Qian, L., Xiao, L., Song, W., Gao, Y., et al. (2020). The STAT3-miR-223-TGFBR3/HMGCS1 axis modulates the progression of cervical carcinoma. *Mol. Oncol.* 14 (9), 2313–2331. doi:10.1002/1878-0261.12737

Zhang, G., Lu, J., Yang, M., Wang, Y., Liu, H., and Xu, C. (2020). Elevated GALNT10 expression identifies immunosuppressive microenvironment and dismal prognosis of patients with high grade serous ovarian cancer. *CII* 69 (2), 175–187. doi:10.1007/s00262-019-02454-1

Zhang, J. R., Hou, P., Wang, X. J., Weng, Z. Q., Shang-Guan, X. C., Wang, H., et al. (2021). TNFRSF11B suppresses memory CD4+ T cell infiltration in the colon cancer microenvironment: A multiomics integrative analysis. *Front. Immunol.* 12, 742358. doi:10.3389/fimmu.2021.742358

Zhang, X., Zhang, H., Liao, Z., Zhang, J., Liang, H., Wang, W., et al. (2022). SHC4 promotes tumor proliferation and metastasis by activating STAT3 signaling in hepatocellular carcinoma. *Cancer Cell. Int.* 22 (1), 24. doi:10.1186/s12935-022-02446-9

Zhu, X., Li, S., Luo, J., Ying, X., Li, Z., Wang, Y., et al. (2022). Subtyping of human papillomavirus-positive cervical cancers based on the expression profiles of 50 genes. *Front. Immunol.* 13, 801639. doi:10.3389/fimmu.2022.801639

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fgene.2023.1093163

OPEN ACCESS

EDITED BY
Yi Yao,
Renmin Hospital of Wuhan University, China

REVIEWED BY
Muhammad Khan,
Guangzhou Medical University Cancer Hospital, China
Emil Bulatov,
Kazan Federal University, Russia

*CORRESPONDENCE
Dilimulati Ismtula,
mlt0306@sina.com
Chenming Guo,
gcm_xjmu@yeah.net

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

RECEIVED 08 November 2022
ACCEPTED 14 March 2023
PUBLISHED 23 March 2023

CITATION
Wang Y, Bi X, Luo Z, Wang H, Ismtula D and Guo C (2023), Gelsolin: A comprehensive pan-cancer analysis of potential prognosis, diagnostic, and immune biomarkers.
Front. Genet. 14:1093163.
doi: 10.3389/fgene.2023.1093163

Gelsolin: A comprehensive pan-cancer analysis of potential prognosis, diagnostic, and immune biomarkers

Yiyang Wang[1], Xiaojuan Bi[1,2], Zhiwen Luo[1], Haiyan Wang[1], Dilimulati Ismtula[1]* and Chenming Guo[1]*

[1]Department of Breast Surgery, Center of Digestive and Vascular, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China, [2]State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Medicine Institute, Urumqi, China

Introduction: Gelsolin (GSN), a calcium-regulated actin-binding protein, is out of balance in various cancers. It can mediate cytoskeletal remodeling and regulate epithelial-mesenchymal conversion (EMT), but the studies on GSN function in pan-cancer are limited.

Methods: We studied the transcription level, prognostic impact, diagnostic value, genetic, epigenetic modification, methylation level and immune significance of GSN in pan-cancer to fully comprehend the function of GSN in various malignancies based on multiple databases like The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO).

Results: Pan-cancer research showed that GSN was downregulated in most tumors and expressed differently in immunological and molecular subtypes of many cancers. GSN had varying impacts on the prognosis of various tumor types. However, all had moderate to high diagnostic efficiency, and serum GSN had good diagnostic value in breast cancer patients (AUC = 0.947). Moreover, GSN was a distinguishing prognosis factor for some specific cancer types. The GSN protein was hypophosphorylated, and its promoter was hypermethylated in most cancers. GSN was linked to the infiltration level of several immunity cells and was essential in anti-tumor immune cell infiltration. KEGG and GSEA analyses showed that GSN was vital in the functions and proteoglycans processes in cancer, chemokine signaling pathway and other immune-related pathways, DNA methylation and cell cycle.

Discussion: In conclusion, GSN possesses the ability to be a predictive, diagnostic, and immune indicator in pan-cancer.

KEYWORDS

pan-cancer, GSN, prognosis, serodiagnostics, tumor immunity, methylation

1 Introduction

The Gelsolin (GSN) gene found on chromosome 9q33.2 encodes a calcium-regulated actin regulatory protein of 782 amino acids. It comprises six gelsolin-like homologous domains (G1-G6). G1 and G4 bind two Ca^{2+} in two forms of shared Ca^{2+} with actin and completely wrapped Ca^{2+}, while G2, G3, G5, and G6 each bind one Ca^{2+} in the form of wholly wrapped Ca^{2+}, reorganizing the actin cytoskeleton, which affects cell motility, cell division and apoptosis (Choe et al., 2002; Nag et al., 2009). GSN is widely found in plasma and cytoplasm and acts as a transcriptional cofactor in

signal transduction, and epigenetic changes affect its expression and activity and are essential for various diseases, including cancer, infection and inflammation, and heart damage (Li et al., 2012).

GSN is intimately linked to different types of tumor development as a vital controller of cell activity, division and death. The low GSN expression in colon cancer tissues is a favorable factor that improves the prognosis of colon cancer patients (Kim et al., 2018; Chen et al., 2019a) due to the silencing of GSN impedes colorectal cancer cell migration and invasiveness and induces cell cycle stagnation (Huang et al., 2022). Similarly, GSN, strongly expressed in bladder cancer tissues, is a major gene for poor prognosis. However, upregulated transcription factor 3 (ATF3) inhibits bladder cancer metastasis through upregulated GSN-mediated actin cytoskeletal remodeling (Yuan et al., 2013; Yang et al., 2020). In liver cancer, patients have a poor prognosis with high GSN expression, possibly because GSN overexpression increases the aggressiveness of cancer cells *via* controlling epithelial-mesenchymal transition (EMT) (Zhang et al., 2020b). Similarly, GSN, upregulated expression in non-small cell lung cancer (NSCLC) and breast cancer tissues, can mediate EMT action to increase cancer cell aggressiveness (Chen et al., 2015; Liu et al., 2021). Furthermore, upregulated GSN may suppress cancer cell proliferation and metastasis for glioblastoma and myelodysplastic syndrome (MDS) (Zhang et al., 2020a; Deng et al., 2020). For kidney cancer, knocking down GSN can inhibit cancer cell proliferation and metastasis (Xu et al., 2017). Accordingly, GSN may be an excellent prognostic biomarker in the above cancers, but research on its prognostic value in other cancers needs to be more extensive and clear.

GSN is divided into secretory and cytosolic types, and secretory GSN may have good value in diagnosing cancer. Serum GSN can act as a diagnostic marker for colon and esophageal adenocarcinoma (Shah et al., 2018; Chen et al., 2019b). Also, plasma GSN can be a factor to distinguish whether people with diabetes have pancreatic ductal adenocarcinoma (Peng et al., 2020). Additionally, the combined area under the curve (AUC) can reach 0.85, which is twice as accurate as the tumor marker CA19-9 alone (Peng et al., 2020). GSN is also associated with chemotherapy resistance, and GSN expression levels in gynecological and head and neck cancers tissues are positively correlated with *in vitro* and *in vivo* chemical resistance (Abedini et al., 2014; Wang et al., 2014). Since current research is limited to a single type of cancer, prognosis, or mechanism, investigating GSN function in a pan-cancer is crucial.

This study examined GSN expression and its diagnostic and predictive significance in pan-cancer and used the serum of breast cancer patients to verify it. Moreover, we explored protein phosphorylation, methylation modification, epigenetic alteration and other aspects and discussed the correlation between GSN expression and immunological response, immune cell infiltration and immune-related gene expression. Finally, functional and pathway enrichment analyses were carried out, providing ideas for further functional experiments. This pan-cancer analysis demonstrated the essential function of GSN in cancer.

2 Material and methods

2.1 Data downloading and GSN expression difference analysis

RNA-sequencing (RNA-seq) data and related medical data of the pan-cancer cohort (n = 15,776) were gained from UCSC XENA (https://xenabrowser.net/datapages/), containing genotype tissue expression (GTEx) of 33 different cancers and normal tissues from The Cancer Genome Atlas (TCGA). Transcripts Per Million (TPM) formatted expression spectrum data was transformed by log2 and merged with subsequent analyses. Data were validated using expression data from 8 datasets from the Gene Expression Omnibus (GEO) database, containing GSE42568 (platform: GPL570), GSE9750 (platform: GPL96), GSE20916 (platform: GPL570), GSE31547 (platform: GPL96), GSE54129 (platform: GPL570), GSE225638 (platform: GPL570), GSE43176 (platform: GPL96) and GSE15471 (platform: GPL570).

2.2 Explore diagnostic and prognostic potential of GSN

The Cox regression model examined the connection between GSN expression and outcome in patients with each tumor. Information was obtained regarding patient survival comprises overall survival (OS), disease-specific survival (DSS), disease-free survival (DFS), and progression-free survival (PFS). For conducting COX regression and plotting Kaplan-Meier (KM) curves, the "survival" and "survminer" packages were used. Forest and venn plots were created utilizing the "ggplot2" package to show the finding. The PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) analyzed 12 datasets involving 8 tumors to examine the connection between GSN expression and patient survival prognoses.

The R package "pROC" was utilized to conduct receiver operating characteristic (ROC) curve analysis to explore GSN predicted values in TCGA tumor tissues and the values in the matching GTEx and TCGA normal tissues. AUC between 0.7 and 0.9 indicates that TUBA1B has a specific diagnostic ability. AUC > 0.9 indicated good diagnostic ability.

2.3 Serum sample collection and ELISA

A total of 37 breast cancer patients who had surgery in the First Affiliated Hospital of Xinjiang Medical University were selected. Simultaneously, 31 healthy women who had a physical examination were randomly chosen as normal controls. Inclusion criteria were: ① All patients diagnosed with medical treatment for the first time, and blood samples were taken before systemic chemotherapy, radiotherapy, endocrine therapy, targeted therapy and surgical treatment. ② All patients pathologically diagnosed with primary breast invasive carcinoma (BRCA). ③ All patients had complete medical records. Exclusion criteria were: patients with other malignant tumors, autoimmune diseases, liver disease, kidney disease, and infectious diseases. All subjects signed informed consent, and the study complied with the Declaration of Helsinki of the World Medical Association. Also, the First Affiliated Hospital Ethics Committee of Xinjiang Medical University accepted this study (20220309–167). Venous blood was collected from the fasting elbows of all subjects into vacuum collection vessels without any anticoagulant, and the supernatant was collected after centrifugation for examination. ELISA detected serum GSN content (ab270215, Abcam, United Kingdom).

2.4 Nomograms development and calibration

First, we assessed risk factors that affected patient prognosis using univariate and multivariate Cox regressions; therefore, variables with p-values < 0.1 was employed for subsequent multivariate Cox analyses. GSN expression was split into high- and low-expression groups utilizing the average to be the threshold and then included as an independent factor. The criteria selected in the multivariate COX regression analysis were incorporated into the nomogram, and the consistency index (C-index) determined the predictive validity of nomogram, where 1,000 was used as a replicates number. Plot calibration curves were conducted to contrast the predicted and the real operating systems.

2.5 Use of online databases

The connection between GSN expression and different human cancer subtypes was examined using the "Subtypes" module of TISIDB database (http://cis.hku.hk/TISIDB/index.php) obtaining the connection between methylation levels and the degree of immune cell infiltration (Ru et al., 2019).

The "TCGA" module of UALCAN database (http://ualcan.path.uab.edu/index.html) was employed to contrast the GSN promoter methylation levels of various malignancies between normal and TCGA samples (Chandrashekar et al., 2022). From the "CPTAC" module, the protein content and phosphorylation level of pan-cancer tissue and its corresponding normal tissue were analyzed.

To verify the differential expression of GSN at the protein level, immunohistochemical images of nine cancer tissues and the matching healthy tissues with various GSN expressions and protein content were gained from the Human Protein Atlas (HPA) database (https://www.proteinatlas.org/) (Thul and Lindskog, 2018).

The "Mutation" module in the Gene Set Cancer Analysis (GSCA) database (http://bioinfo.life.hust.edu.cn/GSCA/#/) was employed to assess the copy number variation% (CNV%) within every cancer, the connection between the expression of GSN and its methylation levels and CNV shifts, and the impacts of GSN methylation levels and CNV changes on pan-cancer prognosis (Liu et al., 2018).

The "OncoPrint" module of cBioPortal database (https://www.cbioportal.org) (Gao et al., 2013) was employed to examine the levels of GSN genetic changes in the "TCGA Pan-Cancer Atlas Studies" dataset (10.443 samples with mutation data in 32 studies). The "Cancer Types Summary" module assesses the recurrence of GSN changes, genetic mutation number, mutation type, and CNV in every cancer form. The GSN mutation site was evaluated by the "mutation" module and demonstrated in the 3D structure of its protein.

The GSN percentage within each CNV and Single Nucleotide Variation (SNV) type in pan-cancer, was obtained from The Catalogue of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic).

2.6 Association of GSN with tumor immunity

First, we examined the associations between GSN and tumor mutational load (TMB) and microsatellite instability (MSI) in several cancers using Sangerbox 3.0 (http://vip.sangerbox.com/) online database. The "GSVA" and "org.Hs.eg.db" tools were performed to compute StromalScore, ImmuneScore, and ESTIMATEScore in the ESTIMATE algorithm (Yoshihara et al., 2013). Eight genes were selected as immune checkpoint-associated transcripts, and the connection between their expression and GSN expression in pan-cancer was evaluated. Moreover, a list of genes for immune activator, immunosuppressive, chemokine, chemokine receptor, and major histocompatibility complex (MHC) molecules was gained from the Gene Set Enrichment Analysis (GSEA) database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). The association between GSN expression levels and immune-related gene expression was examined utilizing the Spearman correlation coefficient.

The single-sample GSEA (ssGSEA) algorithm was employed to assess the invasion level of 24 immunological cells in pan-cancer (Bindea et al., 2013). The EPIC, TIMER, CIBERSORT, MCPCOUNTER algorithm of the "Immune" module of TIMER2.0 database (http://timer.cistrome.org/) was utilized to study the link between GSN expression and levels of immune cell infiltration in pan-cancer, containing cancer-associated fibroblasts (CAFs), CD8+ T cells, CD4+ T cells, regulatory T cells (Tregs), B cells, neutrophils, monocytes, myeloid dendritic cells (mDCs), macrophages, and natural killer cells (NKs) (Li et al., 2020). The "Gene_Corr" module was utilized to examine the association between GSN expression and biomarker expression of CAFs and mDCs.

2.7 Function and pathway enrichment analysis

GSN-targeted binding proteins were studied utilizing the STRING database (https://string-db.org/). Experimentally detected GSN-binding proteins were created using setting STRING parameters, and protein-protein interaction (PPI) networks were formed. The "Similar Genes Detection" module in GEPIA2 (http://gepia2.cancer-pku.cn/#index) was utilized to create the first 100 genes co-expressed with GSN (Tang et al., 2019). The "clusterProfiler" and the "org.Hs.eg.db" tools' were utilized to perform enrichment analysis of the GSN functionality, and a bubble chart shows five of each item. Moreover, GSEA was used to elucidate the functional pathways of differential GSN in the two expression groups of varying cancer cohorts, with a gene set of "c2.cp.v7.2.symbols.gmt" from MSigDB, and all analyses were repeated 5,000 times. A ridge plot shows the highest 15 "Reactom pathways" for each cancer type.

2.8 Statistical analysis

R software (vs. 4.0.3, https://www. R-project.org/) for statistical analysis, the "ggplot2" package for a visualization, was used for

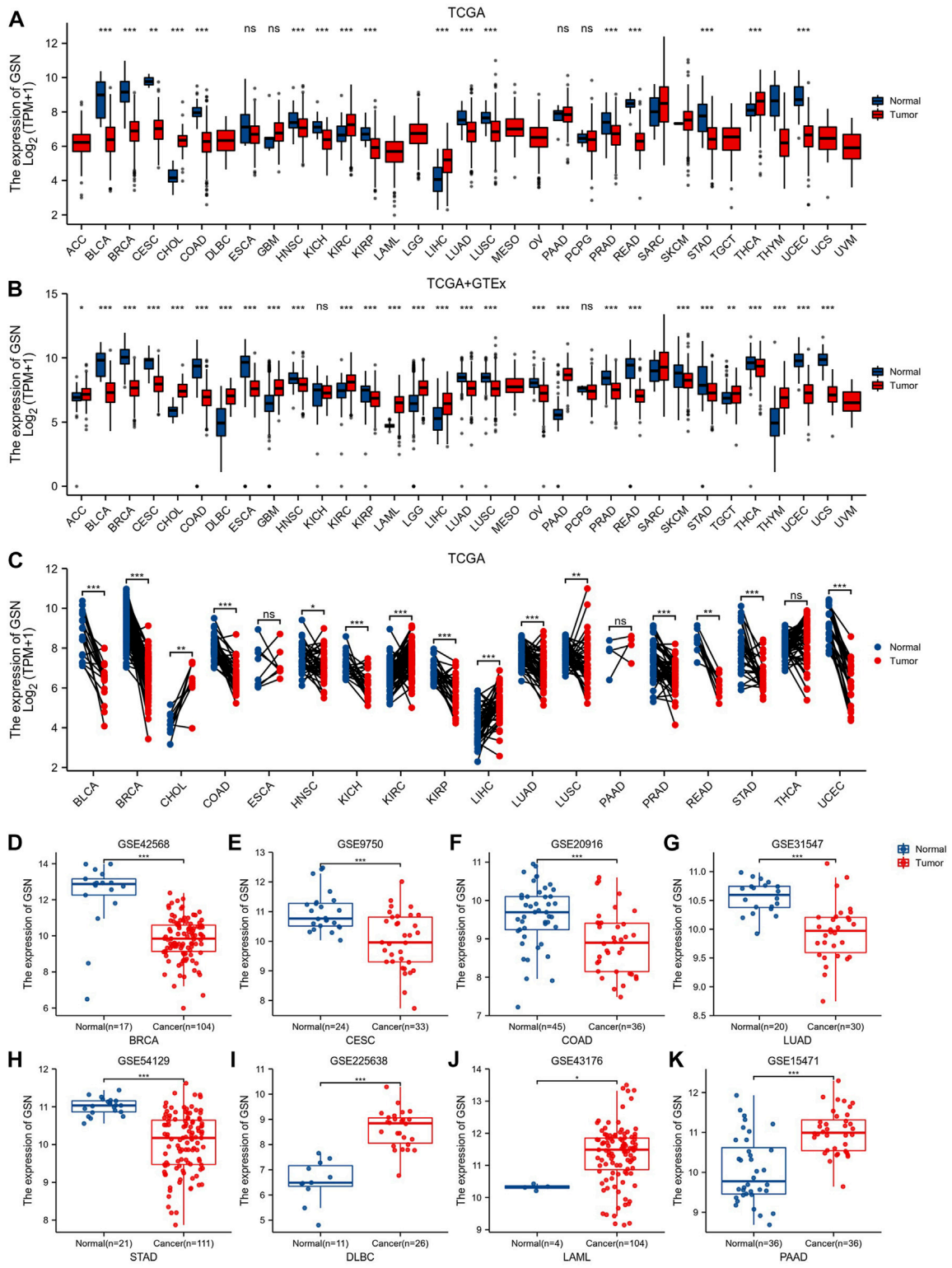


FIGURE 1
Differences in GSN expression in 33 cancers. **(A)** GSN mRNA expression difference between TCGA tumor and normal tissues. **(B)** GSN mRNA expression difference between tumor and normal tissues with data from the TCGA and GTEx. **(C)** GSN mRNA expression in TCGA tumor and paired normal tissues. The differential expression of GSN was analyzed using BRCA (GSE42568) **(D)**, CESC (GSE9750) **(E)**, COAD (GSE20916) **(F)**, LUAD (GSE31547) **(G)**, STAD (GSE54129) **(H)**, DLBC (GSE225638) **(I)**, LAML (GSE43176) **(J)**, and PAAD (GSE15471) **(K)** datasets in GEO databases. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

statistical analyses. The Mann-Whitney U test was utilized to examine variations in expression levels of GSN in unmatched samples, and Wilcoxon signed rank test was employed for paired samples. The Spearman correlation coefficient was utilized to study the association between GSN expression and m6A methylation regulators, TMB, MSI, immune score, and immune-related genes. $p < 0.05$ was regarded as statistically significant.

3 Results

3.1 Differences of GSN expression in pan-carcinoma and its subtypes

Figure 1A shows GSN expression levels in bladder urothelial carcinoma (BLCA), BRCA, cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), and uterine corpus endometrial carcinoma (UCEC) was reduced contrasted with that of healthy tissues. In contrast, cholangiocarcinoma (CHOL), kidney renal clear cell carcinoma (KIRC), liver hepatocellular carcinoma (LIHC) and thyroid carcinoma (THCA) GSN mRNA expression patterns were elevated compared with the matching healthy tissue levels.

GTEx normal tissue was matched to TCGA cancer tissue to create more persuasive outcomes. Furthermore, we found a significant elevation in GSN expression of 11 malignancies: adrenocortical carcinoma (ACC), CHOL, lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), glioblastoma multiforme (GBM), KIRC, acute myeloid leukemia (LAML), brain lower grade glioma (LGG), LIHC, pancreatic adenocarcinoma (PAAD), testicular germ cell tumors (TGCT), thymoma (THYM). Conversely, in 17 malignancies: BLCA, BRCA, CESC, COAD, esophageal carcinoma (ESCA), HNSC, KIRP, LUAD, LUSC, ovarian serous cystadenocarcinoma (OV), PRAD, READ, skin cutaneous melanoma (SKCM), STAD, THCA, UCEC and uterine carcinosarcoma (UCS), GSN expression was downregulated contrasted with in healthy tissue ($p < 0.05$; Figure 1B). In both data, we found different results at GSN expression levels in THCA. In paired samples from 18 malignancies, we discovered that GSN mRNA expression patterns were elevated significantly in malignancies such as CHOL, KIRC and LIHC compared to neighboring normal tissues and significantly downregulated in cancers: BLCA, BRCA, COAD, HNSC, KICH, KIRP, LUAD, LUSC, PRAD, READ, STAD and UCEC ($p < 0.05$; Figure 1C). By analyzing the GEO dataset, we discovered that GSN expression patterns were reduced significantly in BRCA ($p = 6.2e-07$), CESC ($p = 1.8e-04$), COAD ($p = 2.9e-04$), LUAD ($p = 9.2e-06$) and STAD ($p = 4.4e-09$) compared to the corresponding normal tissues (Figures 1D–H). Concurrently, DLBC ($p = 9.7e-09$), LAML ($p = 0.01$) and PAAD ($p = 3.5e-07$) were significantly elevated (Figures 1I–K).

Herein, the correlation between GSN expression in different tumor stages was found that in tumors with decreased GSN expression, including BLCA, THCA and SKCM, the decrease in GSN expression was more significant in early cancers (Figures 2A–C). In tumors with elevated GSN expression, including KIRC, the increase in GSN expression was more significant in early cancers (Figure 2D). This suggested that GSN has the potential to serve as an essential clinical indicator for the early diagnosis of malignancy in these cancers.

Then, we employed the TISIDB database to investigate the differential expression of GSN in various pan-cancer immunological as well as molecular subtypes. Figures 2E–S show that GSN expression varied in 15 cancer subtypes with distinct molecular subtypes. For tumor types with high GSN expression, the molecular subtype of CIMP-low in ACC exhibited the most significant GSN expression (Figure 2E), LGm6-GBM for GBM (Figure 2H), Mesenchymal-like for LGG (Figure 2K) and iCluster:2 for LIHC (Figure 2L). Meanwhile, for tumor types with low GSN expression, GSN expression was the lowest in the molecular subtype of LumB for BRCA (Figure 2F), HM-SNV for COAD (Figure 2G), Basal for HNSC (Figure 2I), C2b for KIRP (Figure 2J), primitive for LUSC (Figure 2M), Proliferative for OV (Figure 2N), Wnt-altered for PCPG (Figure 2O), 1-ERG for PRAD (Figure 2P), RAS_Hotspot_Mutants for SKCM (Figure 2Q), HM-indel for STAD (Figure 2R) as well as CN_HIGH for UCEC (Figure 2S).

Moreover, we discovered that GSN expression was significantly associated with various immunological subtypes of 18 malignancies: ACC, BLCA, BRCA, COAD, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, OV, PCPG, PRAD, STAD, THCA, UCEC and UVM. In many cancers, GSN expression was highest in the C3 (inflammatory) immune subtype and lowest in the C4 (lymphocyte-depleted) immune subtype (Figure 2T). In summary, immune and molecular subtypes exhibited various GSN expressions.

3.2 Prognostic and diagnostic value of GSN in pan-carcinoma

To comprehend if GSN expression influences the outcome of cancer patients, we carried out a survival analysis according to GSN expression in cancer patients using the PrognoScan database. Herein, 11 datasets were included: (GSE5287, GSE17536, GSE14333, GSE8970, GSE12417, GSE4412, GSE1456, GSE3494, GSE4922, GSE4475, and GSE13213) from bladder cancer, colorectal cancer, LAML, LGG, BRCA, DLBC, and LUAD. Higher GSN expression was related to worse outcomes in bladder cancer patients, colorectal cancer, LAML, and LGG (Cox $p < 0.05$; Figures 3A–G). Lower GSN expression was related to poorer prognoses in BRCA, DLBC, and LUAD patients (Cox $p < 0.05$; Figures 3H–M).

Next, we utilized TCGA RNA-seq data to examine the prognostic value of GSN, including OS, DSS, and PFS. For OS, we found that low-expression GSN was an adverse factor affecting OS in patients with CESC, KIRC, SARC, and low-expression GSN was a protective variable for BLCA, LAML and LGG patients ($p < 0.05$; Figure 4A and Supplementary

10.3389/fgene.2023.1093163

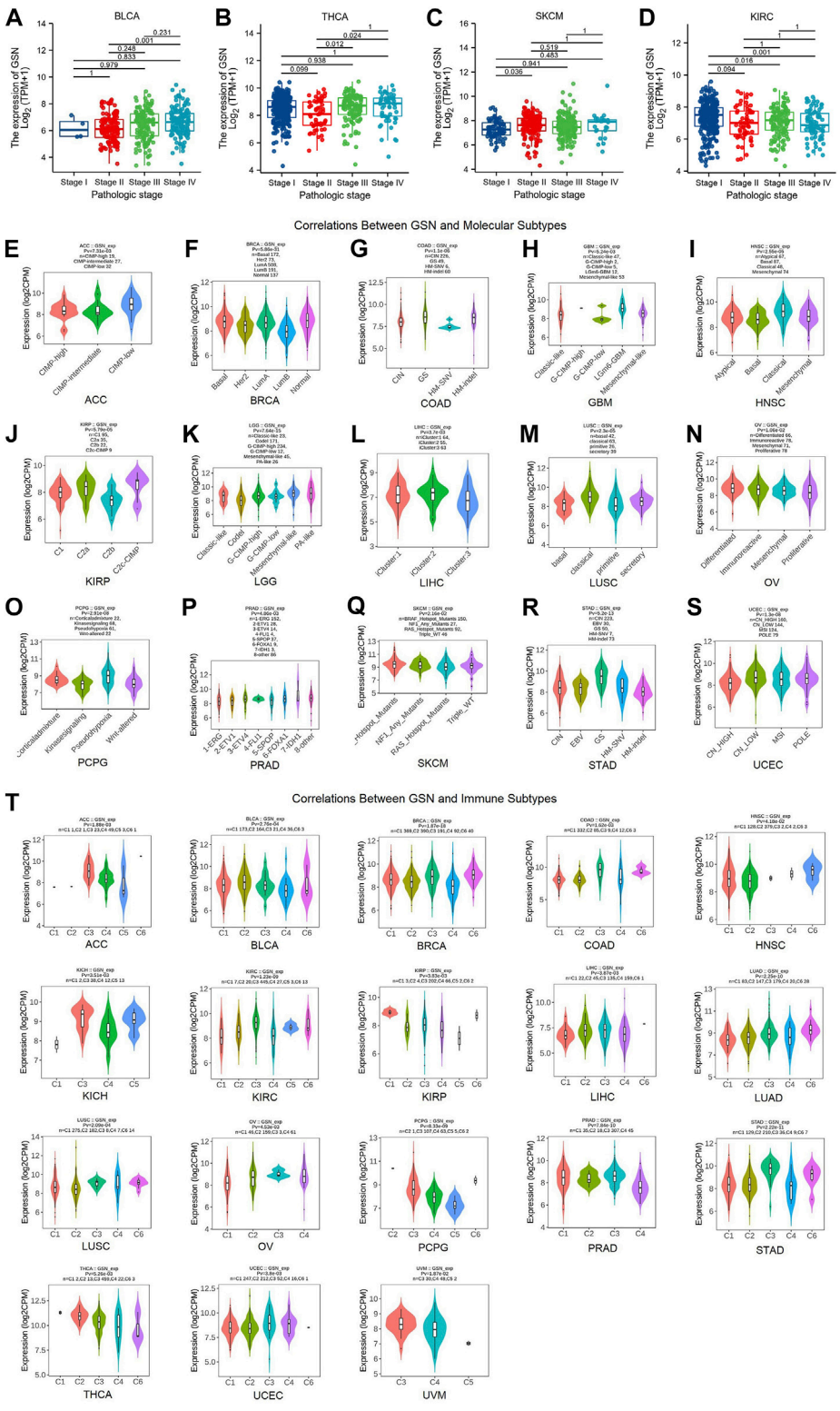


FIGURE 2
Correlation of GSN expression with different tumor stages, molecular subtypes, and immune subtypes. Correlation between GSN expression and different tumor stages, including BLCA **(A)**, THCA **(B)**, SKCM **(C)**, KIRC **(D)**. Correlations between molecular subtypes and GSN expression across TCGA tumors, including **(E)** ACC; **(F)** BRCA; **(G)** COAD; **(H)** GBM; **(I)** HNSC; **(J)** KIRP; **(K)** LGG; **(L)** LIHC; **(M)** LUSC; **(N)** OV; **(O)** PCPG; **(P)** PRAD; **(Q)** SKCM; **(R)** STAD; **(S)** UCEC. **(T)** Correlations between immune subtypes and GSN expression across TCGA tumors, including ACC, BLCA, BRCA, COAD, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, OV, PCPG, PRAD, STAD, THCA, UCEC and UVM. C1 (wound healing), C2 (IFN-g dominant), C3 (inflammatory), C4 (lymphocyte deplete), C5 (immunologically quiet), and C6 (TGF-b dominant).

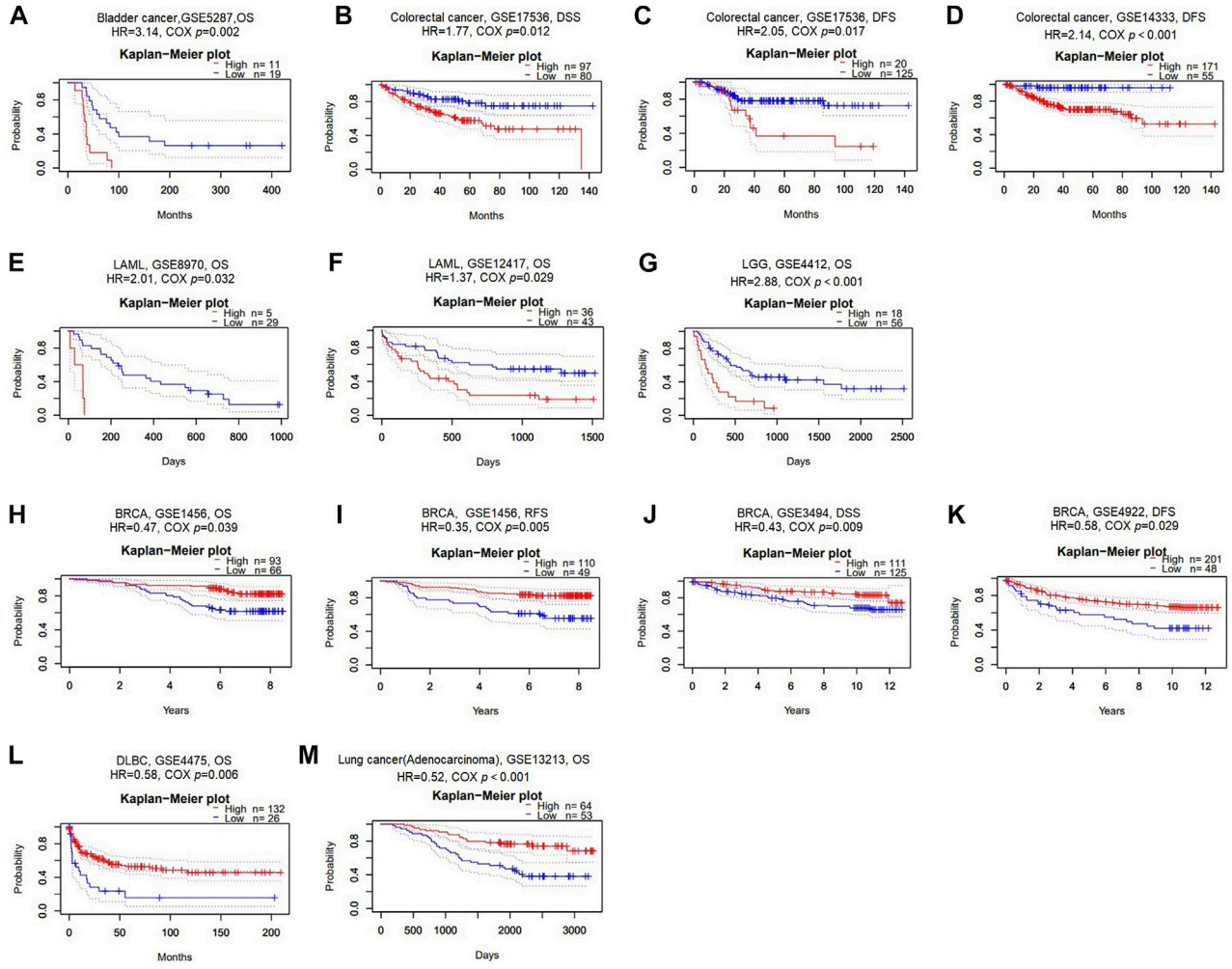


FIGURE 3
Survival analysis of GSN across different cancer types in the GEO and TCGA datasets. Kaplan-Meier plots of GSN in eleven datasets including GSE5287, bladder cancer, OS **(A)**; GSE17536, colorectal cancer, DSS and DFS **(B–C)**; GSE14333, colorectal cancer, DFS **(D)**; GSE8970, LAML, OS **(E)**; GSE12417, LAML, OS **(F)**; GSE4412, LGG, OS **(G)**; GSE1456, BRCA, OS and RFS **(H–I)**; GSE3494, BRCA, DSS **(J)**; GSE4922, BRCA, DFS **(K)**; GSE4475, DLBC, OS **(L)**; GSE13213, LUAD, OS **(M)**.

Figure S1A). For DSS, low GSN expression was a negative factor affecting DSS in patients with CESC, KIRC, SARC, and UCEC, while it is a preventative variable for BLCA, LGG, and STAD patients ($p < 0.05$; Figure 4B and Supplementary Figure S1B). Similarly, low-expression GSN was a detrimental factor affecting PFS in patients with DLBC, KIRC, and UCEC. In contrast, reduced GSN expression was a protective factor affecting PFS in BLCA, LGG, STAD, and UVM patients ($p < 0.05$; Figure 4C and Supplementary Figure S1C). The Venn plot shows that GSN affects three prognoses (OS, DSS, PFS) for patients with BLCA, LGG, and KIRC, revealing that GSN can be a crucial variable in the outcome of such cancers (Figure 4D).

We introduced ROC curve analysis to investigate the possible diagnosis of GSN in pan-cancer. Findings showed that GSN had good diagnostic capabilities (AUC > 0.9), including BLCA (0.945), BRCA (0.981), CESC (0.925), CHOL (0.966), COAD (0.944), DLBC (0.906), ESCA (0.898), LAML (0.917), PAAD (0.975), READ (0.937), UCEC (0.956) and UCS (0.996) (Figure 5A). GSN showed some diagnostic potential (AUC > 0.7) in some tumors, including GBM (0.815), KIRC (0.708), LGG (0.806), LIHC (0.739), LUAD (0.795), LUSC (0.774), OV (0.786), PRAD (0.808) and THYM (0.863) (Figure 5B). GSN showed the highest predictive significance for breast cancer patients by excluding cancers with small sample sizes. Therefore, we collected serum from 37 breast cancer patients and 31 normal people to verify the diagnostic potential of serum GSN for BRCA. Compared to normal people (8.603 ± 3.007 μg/mL), the serum GSN level of breast cancer patients (17.970 ± 5.406 μg/mL) was significantly reduced ($p < 0.001$; Figure 5C). Next, we used ROC curve analysis to examine if serum GSN possesses diagnostic significance of BRCA. The findings revealed that the AUC of serum GSN was 0.947 (cut-off value: 12.883; sensitivity: 97.3%;

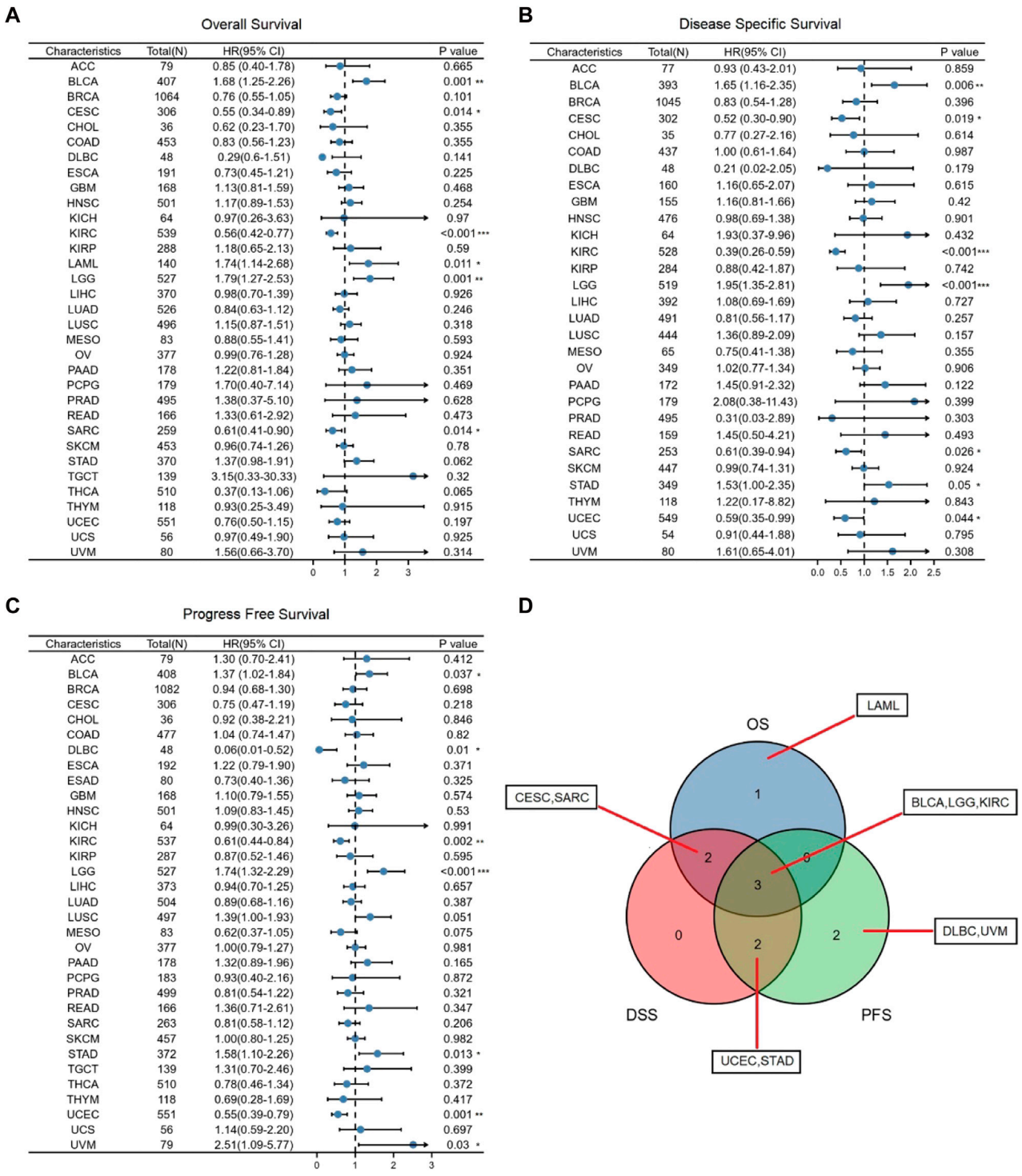


FIGURE 4
Association between GSN expression and prognosis in cancer patients. **(A)** Association between GSN expression and OS in cancer patients. **(B)** Association between GSN expression and DSS in cancer patients. **(C)** Association between GSN expression and PFS in cancer patients. **(D)** The venn diagram shows the intersection of OS, DS, PFS for different cancers. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

specificity: 80.0%; Figure 5D). The results were similar to the above ROC curve analysis data, revealing that serum GSN might be useful in diagnosing BRCA.

Overall, GSN had a modest to the robust ability to differentiate cancer and healthy tissue for most cancers. Serum GSN was validated to have an excellent ability to diagnose breast cancer patients.

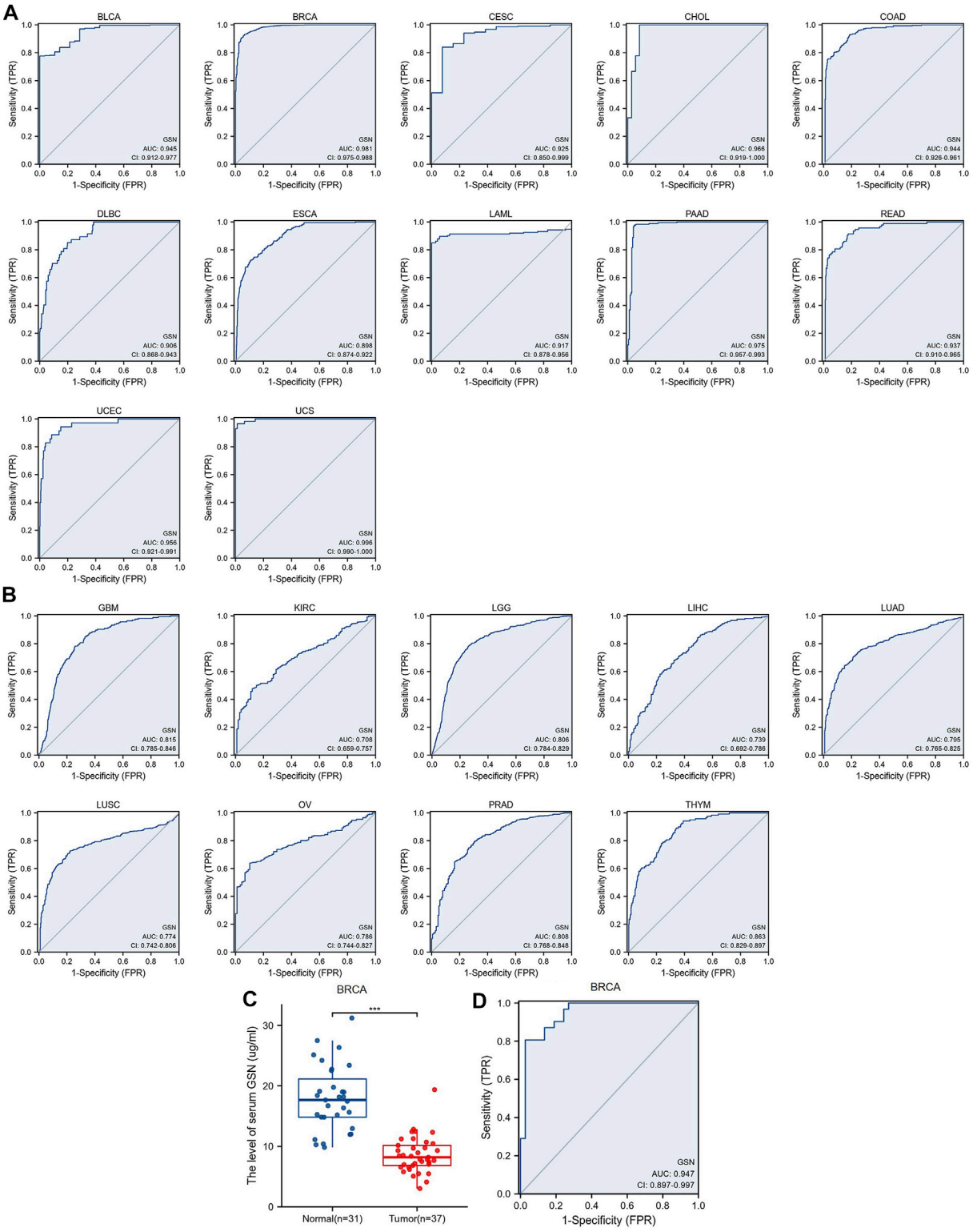


FIGURE 5
Receiver operating characteristic (ROC) curve of GSN expression in pan-carcinoma and determination of serum GSN in breast cancer patients. **(A)** GSN expresses cancers of good diagnostic value (AUC>0.9), including BLCA, BRCA, CESC, CHOL, COAD, DLBC, ESCA, LAML, PAAD, READ, UCEC, UCS. **(B)** GSN expresses cancer with some diagnostic value (AUC>0.7), including GBM, KIRC, LGG, LIHC, LUAD, LUSC, OV, PRAD, THYM. **(C)** GSN protein content in serum of breast cancer patients. **(D)** Diagnostic ROC curve of serum GSN for breast cancer. (***$p < 0.001$).

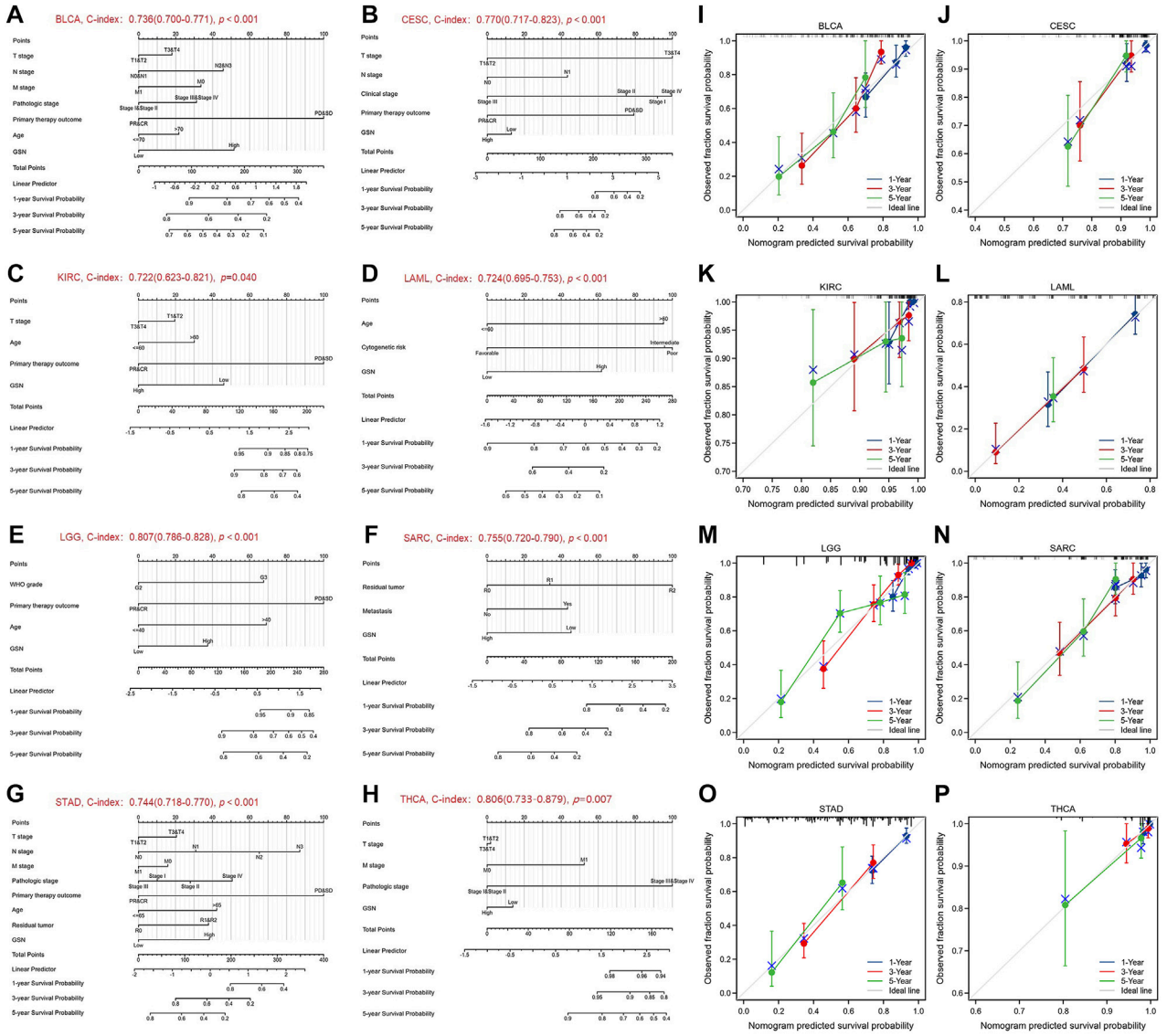


FIGURE 6
Nomograms and calibration curves predicting patient OS in 8 cancers. Nomograms of BLCA **(A)**; CESC **(B)**; KIRC **(C)**; LAML **(D)**; LGG **(E)**; SARC **(F)**; STAD **(G)**; THCA **(H)**. Calibration curves of BLCA **(I)**; CESC **(J)**; KIRC **(K)**; LAML **(L)**; LGG **(M)**; SARC **(N)**; STAD **(O)**; THCA **(P)**. The horizontal and vertical coordinates are the model predicted and actually observed survival probability, respectively. The closer each line is to the ideal line, the better the model.

3.3 GSN is an independent variable in prognosis of some malignancies

We performed univariate and multivariate regression analyses for eight cancer types to investigate the risk factors influencing OS in cancer patients. Herein, the univariate COX regression analysis incorporated cancer types with $p < 0.1$: BLCA, CESC, KIRC, LAML, LGG, SARC, STAD and THCA. For BLCA, multivariate analysis showed that the main treatment result (partial response (PR)/complete response (CR), hazard ratio (HR) = 0.352, $p < 0.001$) and GSN expression (high GSN, HR = 1.712, $p = 0.043$) were independent factors affecting patient OS (Table S1A). For CESC, T stage (T3/T4, HR = 10.091, $p = 0.002$), N stage (N1, HR = 2.722, $p = 0.043$), clinical stage (stage III, HR = 0.119, $p = 0.034$), and main treatment result (PR/CR, HR = 0.160, $p < 0.001$) were independent predictive variables (Supplementary Table S1B). For KIRC, the main treatment result (PR/CR, HR = 0.120, $p = 0.002$) was the only independent predictive variable (Supplementary Table S1C). For LAML, age (>60, HR = 2.751, $p < 0.001$), cytogenetic risk (intermediate, HR = 2.767, $p = 0.005$) (poor, HR = 2.893, $p = 0.009$), and GSN expression (high GSN, HR = 1.928, $p = 0.004$) were independent predictive variables (Supplementary Table S1D). For LGG, WHO grade (G3, HR = 2.871, $p < 0.001$), main treatment result (PR/CR, HR = 0.210, $p < 0.001$), age (>40, HR = 2.939, $p < 0.001$), and GSN expression (high GSN, HR = 1.793, $p = 0.004$) were independent prognostic factors (Supplementary Table S1E). For SARC, residual tumor (R1, HR = 2.192, $p = 0.012$) (R2, HR = 10.143, $p < 0.001$), metastasis (transferred group, HR = 2.738, $p < 0.001$), and GSN expression (high GSN, HR = 0.350, $p < 0.001$) were independent predictive variables (Supplementary Table S1F). For

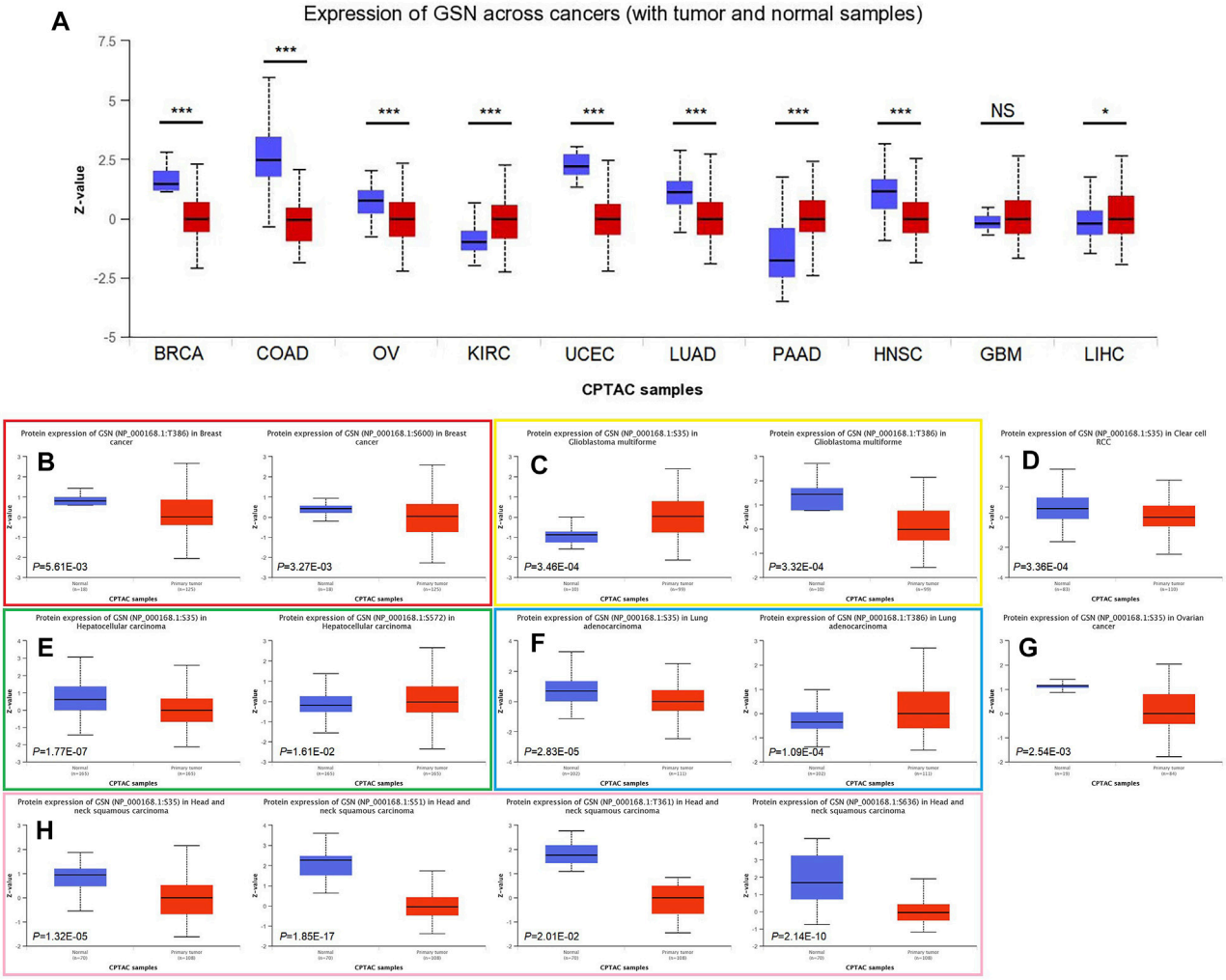


FIGURE 7
Differences in GSN protein content and phosphorylation levels in pan-carcinoma. **(A)** Differences in GSN protein content in pan-carcinoma, including BRCA, COAD, OV, KIRC, UCEC, LUAD, PAAD, HNSC, GBM, and LIHC. Phosphorylation levels of GSN proteins varied in BRCA **(B)**, GBM **(C)**, KIRC **(D)**, LIHC **(E)**, LUAD **(F)**, OV **(G)**, and HNSC **(H)**. (*$p < 0.05$, ***$p < 0.001$, NS: $p > 0.05$).

STAD, N stage (N3, HR = 2.899, $p = 0.043$), primary therapy outcome (PR/CR, HR = 0.295, $p < 0.001$), age (>65, HR = 1.676, $p = 0.019$), and GSN expression (high GSN, HR = 1.597, $p = 0.033$) were independent predictive variables (Supplementary Table S1G). For THCA, the pathologic stage (stage III/IV, HR = 9.573, $p = 0.010$) was the only independent predictive variable (Supplementary Table S1H).

We used the factors that had $p < 0.1$ in univariate COX regression analysis to construct predictive nomograms and calibrations. The findings revealed that the C-index of nomogram in BLCA was 0.736 (0.700–0.771, Figure 6A), in CESC, was 0.770 (0.717–0.823) (Figure 6B), in KIRC was 0.722 (0.623–0.821, Figure 6C), in LAML was 0.724 (0.695–0.753, Figure 6D), in LGG was 0.807 (0.786–0.828, Figure 6E), in SARC, was 0.755 (0.720–0.790, Figure 6F), in STAD was 0.744 (0.718–0.770, Figure 6G), in THCA, was 0.806 (0.733–0.879, Figure 6H). We then calibrated each nomogram to assess the reliability of this model. Except for THCA, the calibration curves for the remaining seven cancer types were close to the ideal line (Figures 6I–P). Therefore, GSN can be used to predict patient outcomes for these tumors independently.

3.4 Differences in protein content, phosphorylation and methylation modification levels of GSN in pan-cancer

We explored protein expression and phosphorylation levels of GSN using the UALCAN database. We found lower GSN protein expression in BRCA, COAD, OV, UCEC, LUAD and HNSC compared to healthy tissues, as well as higher GSN protein expression in KIRC, PAAD and LIHC, but no difference in GSN protein expression in GBM. Further, we utilized the HPA database to observe immunohistochemical photos to measure protein expression levels of GSN (Figure 7A). We observed that the protein expression of GSN in BLCA, BRCA, CESC, COAD, OV

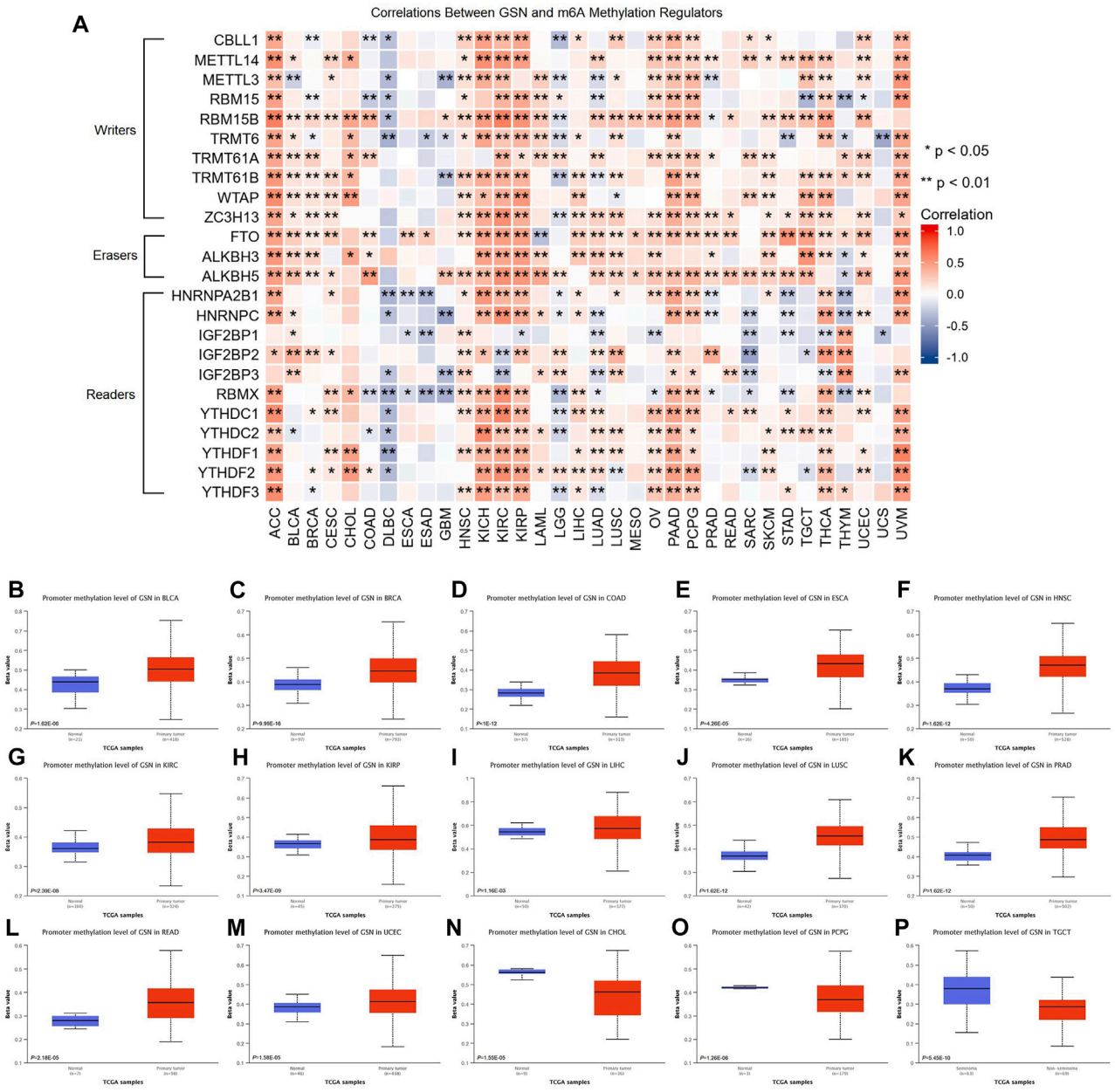


FIGURE 8
Epigenetic methylation analysis of GSN. **(A)** The correlation between the expression of GSN mRNA and m6A methylation regulatory factors in multiple cancers. Differential promoter methylation level (beta values) of GSN in normal tissues and tumors based on UALCAN, including BLCA **(B)**, BRCA **(C)**, COAD **(D)**, ESCA **(E)**, HNSC **(F)**, KIRC **(G)**, KIRP **(H)**, LIHC **(I)**, LUSC **(J)**, PRAD **(K)**, READ **(L)**, UCEC **(M)**, CHOL **(N)**, PCPG **(O)**, and TGCT **(P)**.

and UCEC was significantly reduced contrasted with that of the matching healthy tissue (Supplementary Figure S2). The protein expression of GSN in LGG, LIHC and PAAD was significantly higher compared to the matching healthy tissue. Next, we explored the phosphorylation levels of GSN protein. We observed variations in GSN protein phosphorylation levels in seven malignancies: BRCA, GBM, KIRC, LIHC, LUAD, OV, and HNSC (Figures 7B–H). Among them, S35 was the most crucial phosphorylation modification site, and except for GBM, the phosphorylation level of S35 in other cancers was decreased compared to that of healthy tissues (Figures 7C–H). In HNSC, we found that the GSN protein had the most phosphorylation modification sites, and the phosphorylation was reduced compared to that of healthy tissue (Figure 7H).

We examined the link between GSN mRNA expression and m6A methylation controllers in several tumors because m6A methylation plays a significant part in carcinogenesis and development. In total, 24 essential m6A methylation controllers were chosen: 10 writers (CBLL1, METTL14, METTL3, RBM15, RBM15B, TRMT6, TRMT61A, TRMT61B, WTAP, ZC3H13), 3 erasers (FTO, ALKBH3, ALKBH5), and 11 readers (HNRNPA2B1, HNRNPC, IGF2BP1, IGF2BP2, IGF2BP3, RBMX,

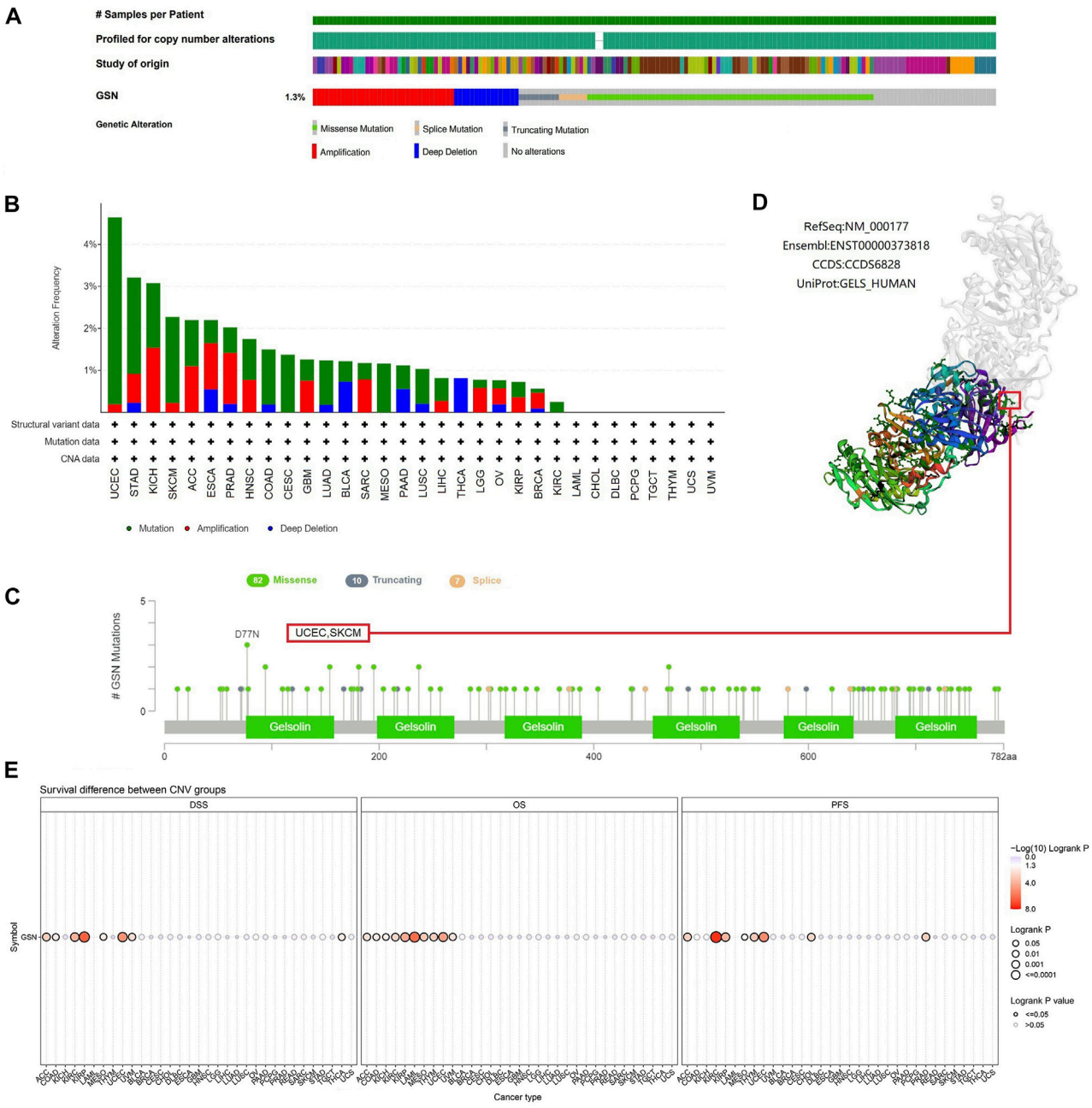


FIGURE 9
Mutated features of GSN in different tumors. **(A)** Summary of changes in GSN expression in different tumors. **(B)** Bar plot of GSN alteration frequency and types across different cancer types. **(C)** The landscape of GSN mutation with the location, types, and number and their relationship with protein domains. **(D)** Some GSN mutations are shown on the 3D structure of the protein. **(E)** Correlation between CNV in GSN and prognosis in cancer patients.

YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3). The heatmap showed that in ACC, KICH, KIRC, KIRP, PAAD, PCPG, THCA, as well as UVM, GSN expression was positively linked to the expression of many m6A methylation regulators (Figure 8A). Moreover, we contrasted promoter methylation contents of GSN in healthy and tumor tissues. The findings declared that the GSN promoter was hypermethylated in several malignancies: BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUSC, PRAD, READ and UCEC (Figures 8B–M). In contrast, the GSN promoter was hypomethylated in CHOL, PCPG, and TGCT contrasted with healthy tissues (Figures 8N–P).

Certain DNA methylations play a massive role in tumor immunogenicity (Hogg et al., 2020). Subsequently, we employed the TISDIB database to explore the connection between GSN methylation patterns and immune cell infiltration. Moreover, the heat map exhibited that GSN methylation levels were adversely linked to the infiltration degree of most immunity cells in ACC, BLCA, BRCA, COAD, GBM, KIRP, LGG, LIHC, PCPG, and PRAD,

while were positively linked to the infiltration degree of most immune cells in CESC, ESCA, HNSC, LUAD, LUSC, and SARC (Supplementary Figure S3A). We utilized the GSCA database to examine the association between GSN methylation patterns and GSN mRNA expression and their influence on the outcome of cancer patients. Accordingly, GSN methylation levels were adversely related to GSN mRNA expression in most cancers except for CHOL and DLBC (Supplementary Figure S3B). Moreover, the hypomethylation level of GSN was an adverse variable influencing the outcome of LGG and BLCA patients (Supplementary Figure S3C).

In conclusion, the GSN protein exhibited low phosphorylation levels, and the GSN promoter exhibited hypermethylation and affected immune cell invasion and patient outcomes in most cancers.

3.5 Genetic changes characteristics of GSN

Cancer is driven by many genetic changes, some of which are potential molecular therapeutic targets (Ben-David and Amon, 2020). Novel therapeutics targeting highly mutated transcription factor TP53 gene products have performed well in pan-carcinoma (Stephenson Clarke et al., 2022). Therefore, we explored its genetic alterations to investigate whether GSN can be used as a target for molecular therapy. We found that 139 of the 10,443 samples (1.3%) developed GSN mutations, and missense mutations were the most common in GSN (Figure 9A). Among all mutations, 45.36% of mutations belonged to missense substitution, and 19.20% of mutations belonged to synonymous substitution (Supplementary Figure S4B). Additionally, the most dominant SNV categories were G > A (35.53%), followed by C > T (30.92%) (Supplementary Figure S4C). The five cancer types that had the greatest mutation frequency were: UCEC (4.64%), STAD (3.21%), KICH (3.08%), SKCM (2.27%), and ACC (2.20%, Figure 9B). D77N in the gelsolin-like 1 domain was the greatest site with mutations, which occurred in two patients with UCEC and one with SKCM (Figure 9C). We exhibited it in the 3D structure of GSN protein (Figure 9D). GSN changes were significantly linked to higher OS (p = 0.0340) and PFS (p = 0.0107) in UCEC patients (Supplementary Figures S4E–F).

Subsequently, we utilized the GSCA database to study the association between GSN mutation and GSN mRNA expression and GSN mutation and outcome of cancer patients. CNV mutations in GSN were adverse factors affecting the outcome of ACC, KIRC, KIRP, MESO and UCEC patients (Figure 9E). CNV pie chart results showed that heterozygous amplification and heterozygous deletion occurred in most cancers. In contrast, rare homozygous amplification occurred mainly in ACC, PRAD and SARC, and rare homozygous deletion occurred mainly in BLCA, READ and THCA (Supplementary Figure S4A). We identified a positive association between GSN mutations and GSN mRNA expression in LUSC, BLCA, HNSC, OV, ESCA, UCEC, KICH, SKCM, BRCA, LUAD, SARC and PAAD (Supplementary Figure S4D). Genetic alterations in GSN occurred in most cancers and were linked to the outcome of cancer patients.

3.6 GSN is related to immune invasion and immune response in pan-cancer

TMB and MSI can respond to the state of immunotherapy as predictive biomarkers of tumor treatment (Filipovic et al., 2020). The radar chart revealed that GSN expression was adversely linked to TMB in 11 cancer types: BRCA, CESC, HNSC, LIHC, LUAD, LUSC, MESO, PRAD, STAD, THCA, and UVM, but only positively associated with TMB of THYM (Figure 10A). Moreover, GSN expression was positively linked to MSI in BLCA, COAD, GBM, LUSC, and SKCM. In contrast, GSN expression was adversely connected to MSI in CESC, PCPG, and STAD (Figure 10B). We measured the connection between stromal and immunological scores and GSN expression in pan-carcinoma. We found that GSN expression was positively linked to StromalScore, ImmuneScore, and ESTIMATEScore in BLCA, BRCA, CHOL, COAD, ESAD, GBM, KIRP, LAML, LGG, LIHC, LUAD, OV, PAAD, PCPG, PRAD, READ, STAD, TGCT, THCA, UCEC, UCS and UVM. In contrast, GSN expression was positively related to StromalScore and ESTIMATEScore in DLBC, ESCA, HNSC, KIRC, MESO, and THYM (Figure 10C).

We studied the connection between GSN and immune checkpoints. GSN expression was positively linked to the expression of many immune checkpoints in BLCA, BRCA, COAD, GBM, KIRP, LGG, LIHC, OV, PAAD, READ, and UVM (Figure 10D). Subsequently, we evaluated the connection between GSN expression and immune-related gene expression, including 43 immune activation-related genes, 22 immunosuppression-related genes, 21 MHC-related genes, 41 chemokines, and 18 chemokine receptors. GSN expression was positively associated with many immune-related genes in BLCA, BRCA, COAD, KIRP, LGG, LIHC, OV, PAAD, PCPG, PRAD, READ, TGCT, THCA, THYM, and UCM (Supplementary Figures S5A–E). In conclusion, in most cancers, GSN expression was significantly associated with the immune score, immune checkpoints, and immune-related genes.

TIICs were a crucial component of the tumor microenvironment (TME) and were closely related to the aggressiveness of cancer. We employed the ssGSEA method to evaluate the association between GSN expression and the 24 immune cells infiltration level. The heat map findings indicated that GSN expression was positively linked to the invasion degree of most immune cells in most cancers like BLCA, BRCA, COAD, KIRP, LIHC, LUAD, PRAD, READ, THCA, and the degree of invasion of some of these immune cells was significantly linked to GSN expression, such as DC cells, immature DC cells (iDC), macrophages, mast cells, neutrophils, eosinophils, NK cells, T effector memory cells (Tem), T follicular helper cells (TFH, Figure 11A). Furthermore, we examined the link between GSN expression levels and various tumor immunological cell infiltration utilizing EPIC, MCP-COUNTER, CIBERSORT, and TIDE algorithms through the Timer2.0 database. In many cancers, GSN expression was positively linked to the CAFs infiltration level, with GSN expression significantly linked to the degree of CAFs infiltration in BLCA (r = 0.486), BRCA (r = 0.463), COAD (r = 0.467), DLBC (r = 0.542), PCPG (r = 0.551), PRAD (r = 0.815), STAD (r = 0.586), TGCT (r = 0.517), THCA (r = 0.520), and THYM (r = 0.742) (Figure 11B). In many cancers, GSN expression was

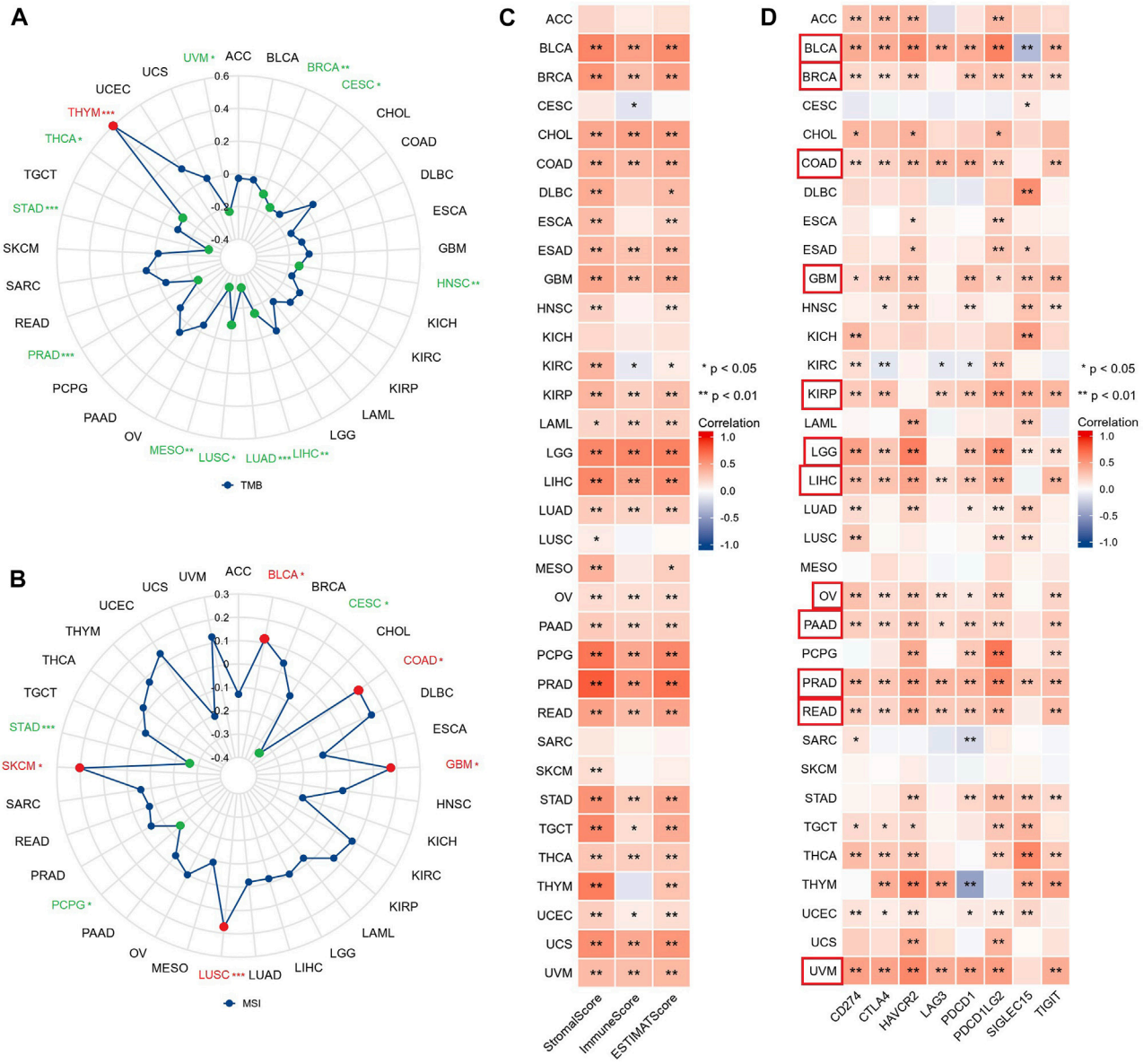


FIGURE 10
GSN expression is associated with TMB, MSI, TME, and immune checkpoints in 33 cancer types. Relationship between GSN expression and TMB **(A)**, MSI **(B)** in 33 cancers. **(C)** Relationship between GSN expression and StromalScore, ImmuneScore, and ESTIMATEScore in 33 cancers. **(D)** Relationship between GSN expression and immune checkpoint expression in 33 cancers.

significantly linked to mDCs invasion (Supplementary Figure S6A). We also found that in THYM, GSN expression was adversely linked to the infiltration level of CD8+ T cells, CD4+ T cells, Tregs, and mDCs, and positively related to the infiltration level of neutrophils, monocytes, macrophages, and NK cells (Supplementary Figure S6A). We studied the association between GSN expression and biomarkers of CAFs and mDCs. The findings declared that GSN expression was positively linked to the biomarker expression of CAFs in nearly all malignancy forms (Supplementary Figure S6B). GSN expression was positively linked to the expression of all biomarkers of mDCs in BLCA, BRCA-LumA, LIHC, PRAD, STAD, and TGCT (Supplementary Figure S6C). In most cancers, GSN expression was positively connected to the invasion degree of multiple immunological cells, particularly CAFs and mDCs.

3.7 Functional enrichment analysis of GSN

To fully comprehend the possible molecular pathways of GSN in tumor initiation and establishment, we explored the enrichment analysis of GSN co-expressed genes. We used the GEPIA2 database to obtain the first 100 GSN co-expressed genes (Supplementary Table S2). The 50 GSN-binding proteins acquired from the STRING database were utilized to build a PPI network (Figure 12A). The

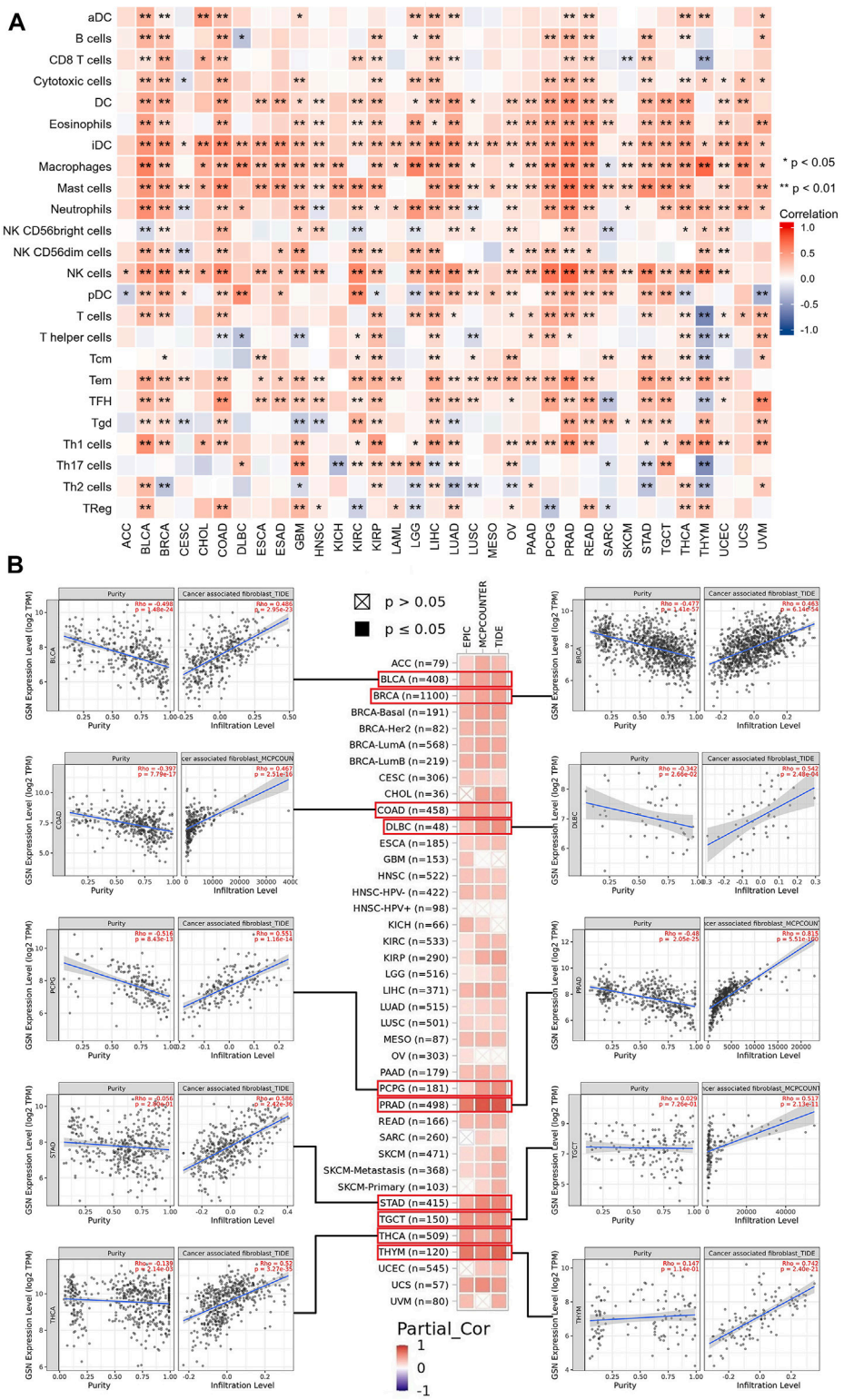


FIGURE 11
Associations between immune cell infiltration levels and GSN expression in pan-cancer. **(A)** The correlation of GSN expression and immune infiltration using the ssGSEA algorithm. **(B)** GSN expression correlation analysis with immune infiltration of CAF cells based on Timer2.0 database, scatter plots including BLCA, BRCA, COAD, DLBC, PCPG, PRAD, STAD, TGCT, THCA, and THYM.

10.3389/fgene.2023.1093163

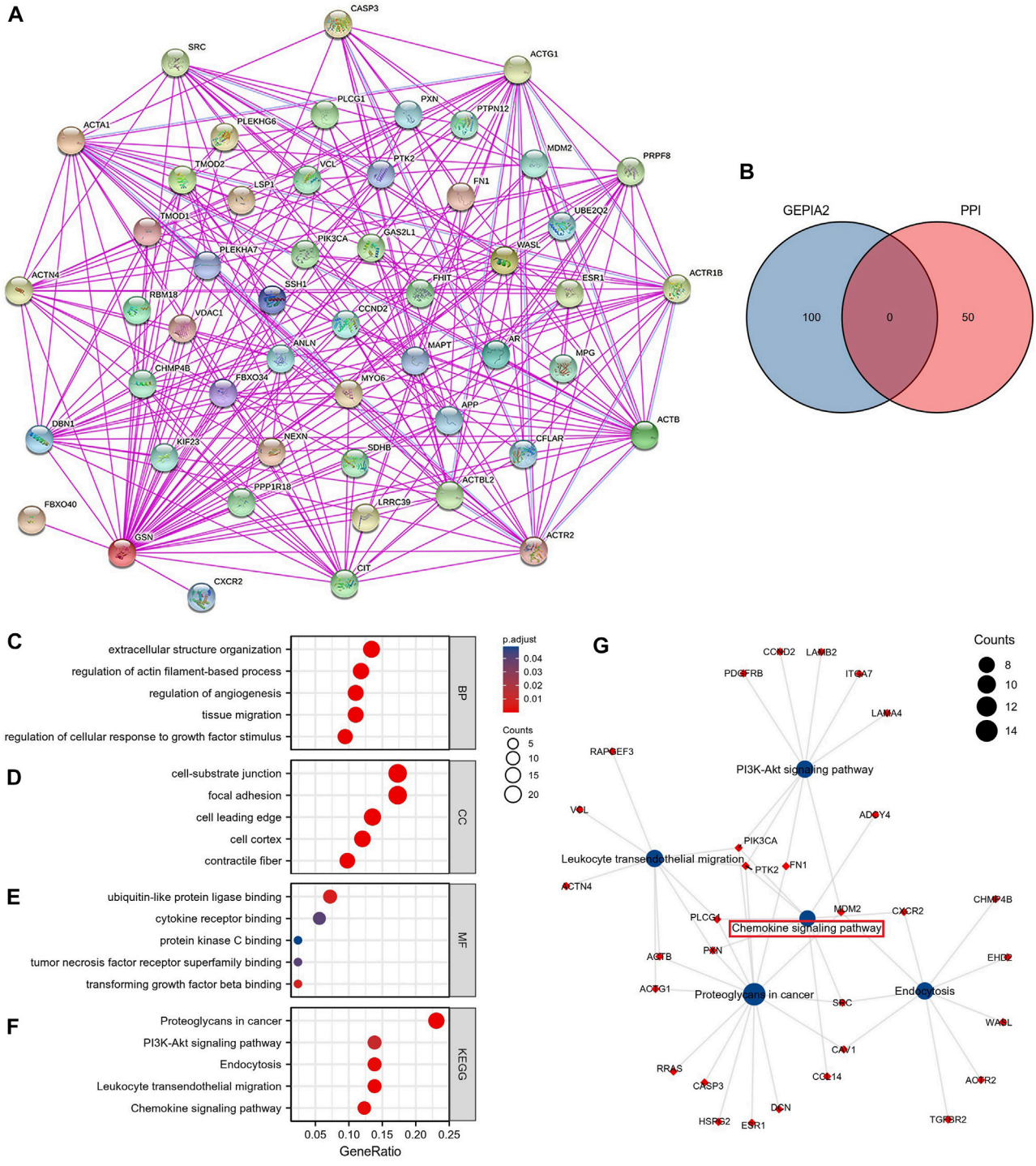


FIGURE 12
GSN-related genes, interacting proteins and functional enrichment analysis. **(A)** PPI Network for GSN. **(B)** The intersection of GSN-binding and interacting genes after selection by Venn diagram analysis. GO analyses, including biological process **(C)**, cellular component **(D)**, molecular function **(E)**, and KEGG pathway **(F)**. **(G)** Visual network of KEGG analyses.

venn plot shows no genes in common (Figure 12B). Subsequently, the first 100 GSN co-expressed genes and 50 GSN-binding proteins were involved in the functional enrichment analysis. Eventually, 357 GO categories were noticed, including 337 biological processes (BP), 55 cellular components (CC), and 44 molecular functions (MF), aside from 49 KEGG pathways (Supplementary Table S3). We presented the top five cancer-related items in each GO entry. The results showed that BP was mainly involved in the extracellular structure organization, actin filament-based process regulation, angiogenesis, tissue migration, and cellular response to growth

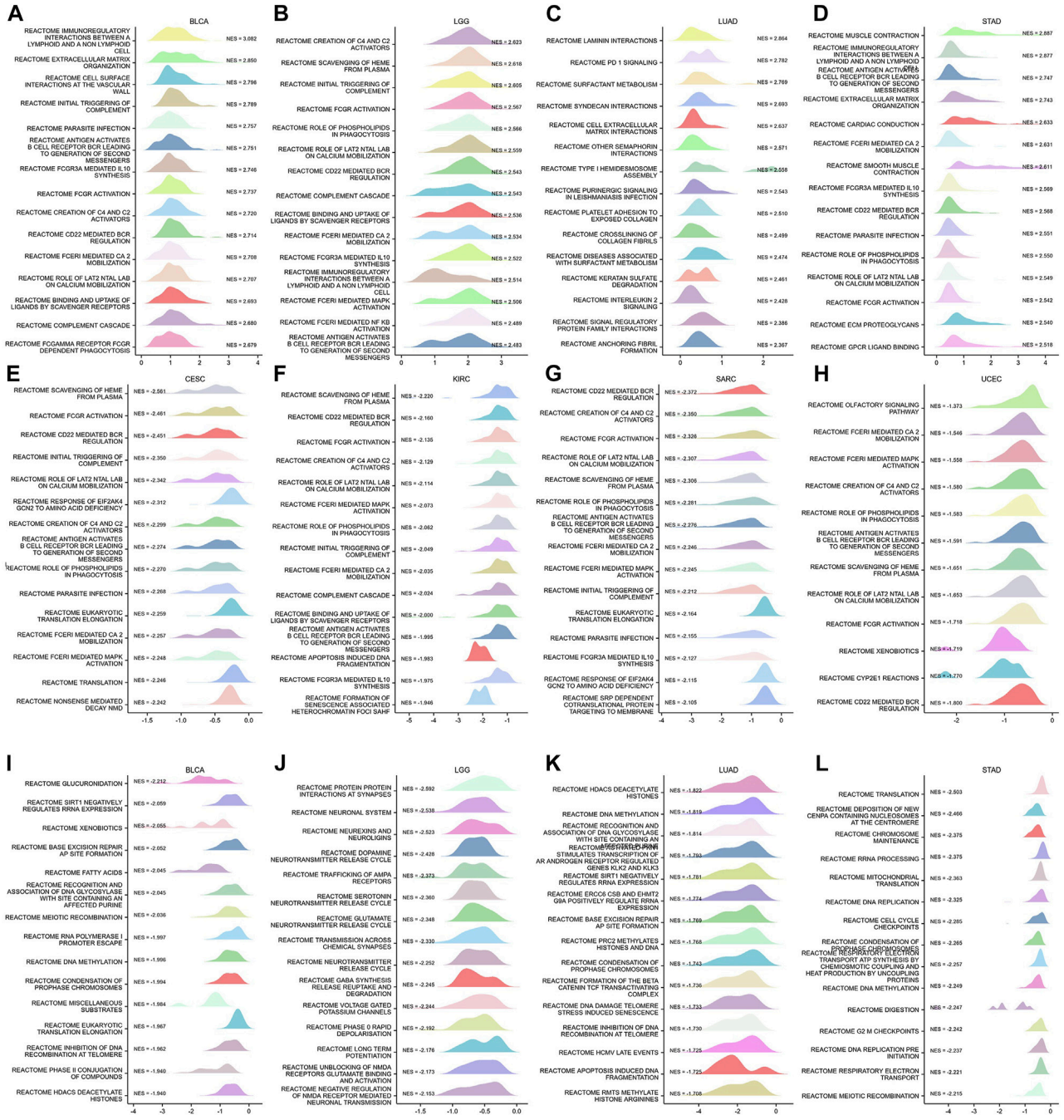


FIGURE 13
GSEA functional enrichment analysis of GSN in 8 cancers. In BLCA **(A)**, LGG **(B)**, LUAD **(C)**, and STAD **(D)**, the first 15 reaction pathways were positively correlated with GSN expression. In CESC **(E)**, KIRC **(F)**, SARC **(G)**, UCEC **(H)**, BLCA **(I)**, LGG **(J)**, LUAD **(K)**, and STAD **(L)**, the first 15 reaction pathways negatively correlated with GSN expression.

factor stimulus (Figure 12C). The CC was mainly enriched in the focal adhesion, cell-substrate junction, cell leading edge, cell cortex, and contractile fiber (Figure 12D). The MF contained ubiquitin-like protein ligase binding, cytokine receptor binding, protein kinase C binding, tumor necrosis factor receptor superfamily binding, and transforming growth factor beta binding (Figure 12E). KEGG pathway analysis found that proteoglycans may mediate GSN in cancer, PI3K-Akt signaling pathway, endocytosis, leukocyte transendothelial migration, and chemokine signaling pathway (Figure 12F). We visualized the KEGG pathway and found that the Chemokine signaling pathway had the most overlapping genes, suggesting that the Chemokine signaling pathway may be a critical GSN-mediated pathway (Figure 12G).

To identify the possible mechanisms for GSN involvement in pan-cancer, we subsequently performed GSEA analysis according to the reactome pathway database. Our analysis included eight cancer

types: BLCA, LGG, LUAD, and STAD, whose outcome was positively linked to GSN expression, and CESC, KIRC, SARC, and UCEC, prognosis was negatively correlated with GSN expression. Our GSEA results showed that among cancer types with a favorable outcome associated with GSN expression, genes positively related to GSN expression were mainly enriched in the immune-associated reactome pathway (Figures 13A–H). Similarly, in cancer types with a prognosis negatively associated with GSN expression, genes adversely linked to GSN expression were also mainly enriched in the immune-related reactome pathway. The enriched pathways mainly included immunomodulatory interactions between lymphocytes and non-lymphocytes, activation of complement signaling pathways, antigen activation of B cell receptors BCR resulting in the production of second messengers, and Cd22-mediated BCR regulation. Furthermore, the enriched pathways suggested that GSN mediated the activation of FCGR pathway and FCERI pathway, mediated downstream IL-10 synthesis, Ca2 mobilization, MAPK activation and other biological functions (Figures 13A–H). Moreover, we found in BLCA, LGG, LUAD and STAD that genes negatively associated with GSN expression are mainly enriched in DNA methylation, protein post-translational modification, mediating cell cycle and other functions, including histone deacetylation and methylation, cell cycle checkpoints, G2M checkpoints and other reactome pathways (Figures 13I–L).

In conclusion, we inferred that GSN played an essential role in cancer primarily by influencing immune-related pathways and regulating biological functions such as DNA methylation.

4 Discussion

This research mainly investigated the influence of GSN on carcinogenesis and progression and its molecular mechanism, including cancer cell proliferation, metastasis, and the means of GSN-mediated EMT. We are the first to examine the involvement of GSN in pan-cancer. The findings revealed that GSN expression varied across 33 malignancies, and GSN expression was significantly elevated in 11 cancer forms and reduced significantly in 17 cancer forms. Notably, we found more pronounced differences in GSN in the early stages of BLCA and KIRC, suggesting that GSN may have the potential to be an early diagnostic marker in both tumor types. In most cancers, we found differences in GSN expression in various subtypes. It should be noted that GSN tends to be highest expressed in the C3 immune subtype and lowest expressed in the C4 immune subtype. Accordingly, the C3 immune subtype has the best survival outcome, whereas the C4 immune subtype has the worst survival outcome, which is consistent with previous studies (Thorsson et al., 2018).

Next, our study determined the predictive value of GSN in pan-cancer and discovered that GSN also possessed various predictive significances in various cancer forms. Notably, GSN was highly expressed in LAML and LGG, while patients who showed elevated GSN expression showed worse outcomes. GSN was low in BRCA, LUAD, CESC, and UCEC, while patients with low GSN expression had a poor prognosis. Upregulated UHRF1 silences GSN to suppress the death of early cervical cancer cells in CESC (Lee et al., 2020). Breast cancer is the first cancer afflicting women worldwide and has replaced lung cancer as the most common cancer worldwide (Sung et al., 2021). TGF-β1 upregulation can increase GSN expression, inhibit cancer cell growth and progression, as well as promote cancer cell migration (Chen et al., 2015). In OV, it has been reported that the OS and PFS of GSN-positive patients were significantly lower than GSN-negative patients, which may be because high GSN expression conferred chemical resistance to cancer cells by altering GSN-FLICE-like inhibitory protein (FLIP)-Itch interaction. Pgsn can be released and transmitted through exosomes (Ex-pGSN), Autocrine upregulation of HIF1α-mediated chemical resistance (Abedini et al., 2014; Asare-Werehene et al., 2020). However, we unobserved such results in OV, and we speculated that this was due to the large number of advanced cancer samples included in the above studies and the insufficient sample size. Wu et al. (2022) declared that GSN expression was dropped in STAD, reduced GSN expression was linked to reduced survival in patients with STAD, and GSN expression was significantly related to STAD tumor purity and degree of DC cell invasion. However, we discovered that GSN high expression was a negative variable affecting the outcome of STAD patients. Therefore, the difference in selected data and the threshold difference in split patients might cause a difference in results. So far, the researchers have unpassed the experimental report, and more data and functional trials are still needed to verify in the future. We observed GSN expression to be an independent predictive variable for BLCA, LAML, LGG, STAD, and SARC, which greatly enriched traditional predictive models, but no related studies have been reported. In CESC, GSN was used as a characteristic gene to construct an effective tool for predicting OS (Li et al., 2022). Finally, GSN is a promising marker for future cancer management.

Next, we examined the diagnostic significance of GSN in pan-cancer. We found that GSN had the best diagnostic efficacy in BRCA, which was validated using serum from breast cancer patients (AUC = 0.947). Serum GSN (AUC = 0.932) levels are superior to common tumor biomarkers, carcinoembryonic antigen (CEA), or carbohydrate antigen 19–9 (CA199) for colon cancer (Chen et al., 2019b). Randall et al. (Brock et al., 2012) reported that for the training set containing 321 COAD samples, 6 serum proteins containing GSN achieved a diagnostic value of AUC = 0.9003 for COAD and AUC = 0.8989 in the validation set containing 110 samples. Further, in esophageal adenocarcinoma, the diagnostic efficacy AUC of serum GSN alone was stabilized at around 0.7 (Shah et al., 2018). The diagnostic value of serum GSN protein alone for pancreatic cancer in diabetic patients was also good (AUC = 0.75) (Peng et al., 2020). In conclusion, GSN may have good predictive potential in many cancer types. However, few relevant studies exist, and more extensive investigations are further required in the future to explore the feasibility of GSN as a diagnostic marker.

Epigenetic modifications play a vital role in tumors through various mechanisms (Sun et al., 2022). VEERLE et al. (De Corte et al., 1999) identified Tyr438 as the most prominent site of GSN phosphorylation by mass spectrometry. We found that S35 was the most common phosphorylation modification site of GSN in most cancers, but whether it is a functional site needs further research. m6A methylation was strongly linked to cancer cell growth, metastasis, immune response and other processes and

affected the sensitivity and resistance of anti-cancer treatment drugs (Lan et al., 2021). Therefore, we indirectly explored the level of GSN methylation modification in pan-carcinoma and its role. In most malignancies, GSN expression is positively correlated with m6A methylation-related gene expression, therefore, we hypothesize that GSN has lower levels of m6A methylation, and the GSN promoter was hypermethylated. In addition, in most cancers, the level of DNA methylation of GSN is negatively correlated with the invasion of immune cells in TME. There are literature reports on raised GSN expression patterns in CESC cells subjected to DNA-hypomethylating agent 5-aza-2′-deoxycytidine (Lee et al., 2020). Tumor-associated macrophages (TAMs) were co-cultured with gastric cancer cells, and DNA methyltransferase 1 (DNMT1) expression increased; however, GSN expression decreased in gastric cancer cells (Wang et al., 2017). In breast cancer, GSN downregulation was triggered *via* hypermethylation of essential DNA methylation sites. A risk score model with excellent prognosis reliability was developed using three methylation probes based on CAV2 and GSN genes (Cao et al., 2022). In conclusion, epigenetic modifications of GSN played an essential role in pan-cancer, but more functional experimental verification mechanisms are needed in the future.

Missense mutations are the most frequent among GSN mutations, the most occurring GSN mutations in UCEC, and D77N in the gelsolin-like 1 domain is the site with the most frequent mutations. In some cancers, the CNV status of GSN was positively correlated with GSN expression and was an adverse factor affecting patient outcomes. We are the first to reveal the significance of GSN mutations in pan-cancer, but more research is required to identify the pathway.

With the advent of immunotherapy, cytokine and immune checkpoint inhibitor (ICI) therapies have gradually proven as medications for several malignancies (Havel et al., 2019). TMB and MSI are predictors of the anti-tumor efficacy of ICIs (Xu et al., 2020). Higher TMB and MSI mean a better response to ICI and a better prognosis for cancer patients (Chen et al., 2019a; Samstein et al., 2019). Our study observed that GSN expression was adversely linked to TMB in LIHC, while GSN was highly expressed in LIHC. GSN expression was positively related to MSI in BLCA, COAD, LUSC, as well as SKCM, whereas GSN was low expressed in these cancers. Therefore, we speculated that GSN is responsible for the low TMB and MSI in the above cancers, predicting that GSN may play a role in immunotherapy. Moreover, in BLCA, BRCA, COAD, GBM, KIRP, LGG, LIHC, OV, PAAD, READ, and UVM, GSN expression was positively linked to immunological scores, most immune checkpoints, and expression of immune-related genes. According to our results, GSN could regulate cancer immunity, and targeting GSN might become a new strategy for tumor immunotherapy.

No correlation analysis has involved GSN and the TME. We observed that GSN expression was positively correlated with the infiltration level of most immunity cells, like DC cells, macrophages, NK cells, Tem, TFH and other immune cells involved in anti-tumor immune effects, and CAF involved in tumor immune evasion or suppression. GSN was low expressed in most cancers, so we speculated that GSN is mainly involved in immune effects in the TME through anti-tumor immune invasion rather than immune escape or immunosuppression. These immune cells can participate in tumor immunity through various mechanisms, including secreting multiple cytokines and chemokines and antigen presentation, mediating the enrollment and functional development of innate and adaptive immunity cells (Ghesquiere et al., 2014; Monteran and Erez, 2019; Shimasaki et al., 2020). Our KEGG analysis showed that GSN might mediate proteoglycans in cancer, PI3K-Akt signaling pathway, endocytosis, leukocyte transendothelial migration, and chemokine signaling pathway. The chemokine signaling pathway was identified as the most critical pathway mediated by GSN. In addition to regulating inflammatory responses, promoting cancer cell metastasis, and regulating apoptosis, glycosylation changes may also regulate inflammatory reactions (Reily et al., 2019).

Similarly, our GSEA analysis showed that in cancer types where GSN expression was negatively associated with prognosis, GSN was positively correlated with immune function. Among the cancer types whose expression was positively associated with prognosis, GSN was negatively correlated with immune function. Although GSN had different effects on the outcome of patients with several malignancies, they all showed the same immune trend; GSN negatively regulated the prognosis of cancer patients by mediating the immune effect in pan-cancer. In cancer forms in which GSN expression was adversely associated with patient outcomes, GSN was negatively correlated with processes such as DNA methylation and cell cycle, reconfirming our previous discussion about the effects of GSN methylation and GSN expression on cell proliferation, invasion, and migration.

5 Conclusion

We conducted the first pan-cancer study of GSN, including expression, prognostic and diagnostic, epigenetics, methylation, immunoassay, and enrichment analyses, indicating that GSN was a potential therapeutic biomarker for malignancy. However, this research has certain restrictions, like a small sample size and lack of experimental validation. In the future, the research sample should be expanded to study the detailed carcinogenic mechanism of GSN in pan-carcinoma through *in vitro* and *in vivo* experiments.

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

Conceiving and designing: YW and CG. Analysis and visualization: YW, ZL, XB, and HW. Writing (original draft

preparation): YW. Review and editing: CG, DI, XB, and HW. Funding acquisition and research supervision: CG and DI. All authors read and approved the final manuscript.

Funding

This study was sponsored by the National Natural Science Foundation of China (32260186), Natural Science Foundation of Xinjiang Province (2022D01A140), Regional Collaborative Innovation Special Project (Science and Technology Assistance to Xinjiang Program) (2022E02136) and The Province and The Ministry Jointly Established the State Key Laboratory (SKL-HIDCA-2021-41).

Acknowledgments

We wish to thank everyone who participated in this research and the databases involved in the 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1093163/full#supplementary-material

References

Abedini, M. R., Wang, P. W., Huang, Y. F., Cao, M., Chou, C. Y., Shieh, D. B., et al. (2014). Cell fate regulation by gelsolin in human gynecologic cancers. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14442–14447. doi:10.1073/pnas.1401166111

Asare-Werehene, M., Nakka, K., Reunov, A., Chiu, C. T., Lee, W. T., Abedini, M. R., et al. (2020). The exosome-mediated autocrine and paracrine actions of plasma gelsolin in ovarian cancer chemoresistance. *Oncogene* 39, 1600–1616. doi:10.1038/s41388-019-1087-9

Ben-David, U., and Amon, A. (2020). Context is everything: Aneuploidy in cancer. *Nat. Rev. Genet.* 21, 44–62. doi:10.1038/s41576-019-0171-x

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Brock, R., Xiong, B., Li, L., Vanbogelen, R. A., and Christman, L. (2012). A multiplex serum protein assay for determining the probability of colorectal cancer. *Am. J. Cancer Res.* 2, 598–605.

Cao, Y. N., Li, Q. Z., and Liu, Y. X. (2022). Discovered Key CpG sites by analyzing DNA methylation and gene expression in breast cancer samples. *Front. Cell Dev. Biol.* 10, 815843. doi:10.3389/fcell.2022.815843

Chandrashekar, D. S., Karthikeyan, S. K., Korla, P. K., Patel, H., Shovon, A. R., Athar, M., et al. (2022). Ualcan: An update to the integrated cancer data analysis platform. *Neoplasia* 25, 18–27. doi:10.1016/j.neo.2022.01.001

Chen, M. H., Chang, S. C., Lin, P. C., Yang, S. H., Lin, C. C., Lan, Y. T., et al. (2019a). Combined microsatellite instability and elevated microsatellite alterations at selected tetranucleotide repeats (EMAST) might Be a more promising immune biomarker in colorectal cancer. *Oncologist* 24, 1534–1542. doi:10.1634/theoncologist.2019-0171

Chen, Z., Li, K., Yin, X., Li, H., Li, Y., Zhang, Q., et al. (2019b). Lower expression of gelsolin in colon cancer and its diagnostic value in colon cancer patients. *J. Cancer* 10, 1288–1296. doi:10.7150/jca.28529

Chen, Z. Y., Wang, P. W., Shieh, D. B., Chiu, K. Y., and Liou, Y. M. (2015). Involvement of gelsolin in TGF-beta 1 induced epithelial to mesenchymal transition in breast cancer cells. *J. Biomed. Sci.* 22, 90. doi:10.1186/s12929-015-0197-0

Choe, H., Burtnick, L. D., Mejillano, M., Yin, H. L., Robinson, R. C., and Choe, S. (2002). The calcium activation of gelsolin: Insights from the 3A structure of the G4-G6/actin complex. *J. Mol. Biol.* 324, 691–702. doi:10.1016/s0022-2836(02)01131-2

de Corte, V., Demol, H., Goethals, M., van Damme, J., Gettemans, J., and Vandekerckhove, J. (1999). Identification of Tyr438 as the major *in vitro* c-src phosphorylation site in human gelsolin: A mass spectrometric approach. *Protein Sci.* 8, 234–241. doi:10.1110/ps.8.1.234

Deng, J., Liang, L., Yi, H., Su, T., Yang, Z., Nie, L., et al. (2020). USP7 inhibition inhibits proliferation and induces megakaryocytic differentiation in MDS cells by upregulating gelsolin. *Br. J. Haematol.* 190, 418–429. doi:10.1111/bjh.16549

Filipovic, A., Miller, G., and Bolen, J. (2020). Progress toward identifying exact proxies for predicting response to immunotherapies. *Front. Cell Dev. Biol.* 8, 155. doi:10.3389/fcell.2020.00155

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

Ghesquiere, B., Wong, B. W., Kuchnio, A., and Carmeliet, P. (2014). Metabolism of stromal and immune cells in health and disease. *Nature* 511, 167–176. doi:10.1038/nature13312

Havel, J. J., Chowell, D., and Chan, T. A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* 19, 133–150. doi:10.1038/s41568-019-0116-x

Hogg, S. J., Beavis, P. A., Dawson, M. A., and Johnstone, R. W. (2020). Targeting the epigenetic regulation of antitumour immunity. *Nat. Rev. Drug Discov.* 19, 776–800. doi:10.1038/s41573-020-0077-5

Huang, C. Y., Lee, K. C., Tung, S. Y., Huang, W. S., Teng, C. C., Lee, K. F., et al. (2022). 2D-DIGE-MS proteomics approaches for identification of gelsolin and peroxiredoxin 4 with lymph node metastasis in colorectal cancer. *Cancers (Basel)* 14, 3189. doi:10.3390/cancers14133189

Kim, J. C., Ha, Y. J., Tak, K. H., Roh, S. A., Kwon, Y. H., Kim, C. W., et al. (2018). Opposite functions of GSN and OAS2 on colorectal cancer metastasis, mediating perineural and lymphovascular invasion, respectively. *PLoS One* 13, e0202856. doi:10.1371/journal.pone.0202856

Lan, Q., Liu, P. Y., Bell, J. L., Wang, J. Y., Huttelmaier, S., Zhang, X. D., et al. (2021). The emerging roles of RNA m(6)A methylation and demethylation as critical regulators of tumorigenesis, drug sensitivity, and resistance. *Cancer Res.* 81, 3431–3440. doi:10.1158/0008-5472.CAN-20-4107

Lee, H. J., Kim, M. J., Kim, Y. S., Choi, M. Y., Cho, G. J., and Choi, W. S. (2020). UHRF1 silences gelsolin to inhibit cell death in early stage cervical cancer. *Biochem. Biophys. Res. Commun.* 526, 1061–1068. doi:10.1016/j.bbrc.2020.03.185

Li, G. H., Arora, P. D., Chen, Y., Mcculloch, C. A., and Liu, P. (2012). Multifunctional roles of gelsolin in health and diseases. *Med. Res. Rev.* 32, 999–1025. doi:10.1002/med.20231

Li, N., Yu, K., Lin, Z., and Zeng, D. (2022). Identifying a cervical cancer survival signature based on mRNA expression and genome-wide copy number variations. *Exp. Biol. Med. (Maywood)* 247, 207–220. doi:10.1177/15353702211053580

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407

Liu, C. J., Hu, F. F., Xia, M. X., Han, L., Zhang, Q., and Guo, A. Y. (2018). GSCALite: A web server for gene set cancer analysis. *Bioinformatics* 34, 3771–3772. doi:10.1093/bioinformatics/bty411

Liu, X., Huang, X., Ma, J., Li, L., Hu, H., Feng, J., et al. (2021). 3'untranslated regions (3'UTR) of Gelsolin mRNA displays anticancer effects in non-small cell lung cancer (NSCLC) cells. *Am. J. Cancer Res.* 11, 3857–3876.

Monteran, L., and Erez, N. (2019). The dark side of fibroblasts: Cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front. Immunol.* 10, 1835. doi:10.3389/fimmu.2019.01835

Nag, S., Ma, Q., Wang, H., Chumnarnsilpa, S., Lee, W. L., Larsson, M., et al. (2009). Ca2+ binding by domain 2 plays a critical role in the activation and stabilization of gelsolin. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13713–13718. doi:10.1073/pnas.0812374106

Peng, H., Pan, S., Yan, Y., Brand, R. E., Petersen, G. M., Chari, S. T., et al. (2020). Systemic proteome alterations linked to early stage pancreatic cancer in diabetic patients. *Cancers (Basel)* 12, 1534. doi:10.3390/cancers12061534

Reily, C., Stewart, T. J., Renfrow, M. B., and Novak, J. (2019). Glycosylation in health and disease. *Nat. Rev. Nephrol.* 15, 346–366. doi:10.1038/s41581-019-0129-4

Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., et al. (2019). Tisidb: An integrated repository portal for tumor-immune system interactions. *Bioinformatics* 35, 4200–4202. doi:10.1093/bioinformatics/btz210

Samstein, R. M., Lee, C. H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* 51, 202–206. doi:10.1038/s41588-018-0312-8

Shah, A. K., Hartel, G., Brown, I., Winterford, C., Na, R., Cao, K. L., et al. (2018). Evaluation of serum glycoprotein biomarker candidates for detection of esophageal adenocarcinoma and surveillance of barrett's esophagus. *Mol. Cell Proteomics* 17, 2324–2334. doi:10.1074/mcp.RA118.000734

Shimasaki, N., Jain, A., and Campana, D. (2020). NK cells for cancer immunotherapy. *Nat. Rev. Drug Discov.* 19, 200–218. doi:10.1038/s41573-019-0052-1

Stephenson Clarke, J. R., Douglas, L. R., Duriez, P. J., Balourdas, D. I., Joerger, A. C., Khadiullina, R., et al. (2022). Discovery of nanomolar-affinity pharmacological chaperones stabilizing the oncogenic p53 mutant Y220C. *ACS Pharmacol. Transl. Sci.* 5, 1169–1180. doi:10.1021/acsptsci.2c00164

Sun, L., Zhang, H., and Gao, P. (2022). Metabolic reprogramming and epigenetic modifications on the path to cancer. *Protein Cell* 13, 877–919. doi:10.1007/s13238-021-00846-7

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 47, W556–W560. doi:10.1093/nar/gkz430

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830.e14. doi:10.1016/j.immuni.2018.03.023

Thul, P. J., and Lindskog, C. (2018). The human protein atlas: A spatial map of the human proteome. *Protein Sci.* 27, 233–244. doi:10.1002/pro.3307

Wang, H. C., Chen, C. W., Yang, C. L., Tsai, I. M., Hou, Y. C., Chen, C. J., et al. (2017). Tumor-associated macrophages promote epigenetic silencing of gelsolin through DNA methyltransferase 1 in gastric cancer cells. *Cancer Immunol. Res.* 5, 885–897. doi:10.1158/2326-6066.CIR-16-0295

Wang, P. W., Abedini, M. R., Yang, L. X., Ding, A. A., Figeys, D., Chang, J. Y., et al. (2014). Gelsolin regulates cisplatin sensitivity in human head-and-neck cancer. *Int. J. Cancer* 135, 2760–2769. doi:10.1002/ijc.28928

Wu, Y., Zheng, J., Yan, Y., Liu, J., and Zhou, Y. (2022). Gelsolin can Be a prognostic biomarker and correlated with immune infiltrates in gastric cancer. *Int. J. Gen. Med.* 15, 927–936. doi:10.2147/IJGM.S339940

Xu, Q., Junttila, S., Scherer, A., Giri, K. R., Kivela, O., Skovorodkin, I., et al. (2017). Renal carcinoma/kidney progenitor cell chimera organoid as a novel tumorigenesis gene discovery model. *Dis. Model Mech.* 10, 1503–1515. doi:10.1242/dmm.028332

Xu, Y., Fu, Y., Zhu, B., Wang, J., and Zhang, B. (2020). Predictive biomarkers of immune checkpoint inhibitors-related toxicities. *Front. Immunol.* 11, 2023. doi:10.3389/fimmu.2020.02023

Yang, J. L., Wang, C. C. N., Cai, J. H., Chou, C. Y., Lin, Y. C., and Hung, C. C. (2020). Identification of GSN and LAMC2 as Key prognostic genes of bladder cancer by integrated bioinformatics analysis. *Cancers (Basel)* 12, 1809. doi:10.3390/cancers12071809

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

Yuan, X., Yu, L., Li, J., Xie, G., Rong, T., Zhang, L., et al. (2013). ATF3 suppresses metastasis of bladder cancer by regulating gelsolin-mediated remodeling of the actin cytoskeleton. *Cancer Res.* 73, 3625–3637. doi:10.1158/0008-5472.CAN-12-3879

Zhang, J., Furuta, T., Sabit, H., Tamai, S., Jiapaer, S., Dong, Y., et al. (2020a). Gelsolin inhibits malignant phenotype of glioblastoma and is regulated by miR-654-5p and miR-450b-5p. *Cancer Sci.* 111, 2413–2422. doi:10.1111/cas.14429

Zhang, Y., Luo, X., Lin, J., Fu, S., Feng, P., Su, H., et al. (2020b). Gelsolin promotes cancer progression by regulating epithelial-mesenchymal transition in hepatocellular carcinoma and correlates with a poor prognosis. *J. Oncol.* 2020, 1980368. doi:10.1155/2020/1980368

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 30 March 2023
DOI 10.3389/fgene.2023.1111816

OPEN ACCESS

EDITED BY
Minglun Li,
LMU Munich University Hospital, Germany

REVIEWED BY
Denggang Fu,
Indiana University, United States
Gengming Cai,
Fujian Medical University, China

*CORRESPONDENCE
Guolin Tan,
guolintan@csu.edu.cn

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

RECEIVED 30 November 2022
ACCEPTED 16 March 2023
PUBLISHED 30 March 2023

CITATION
Wang Q, Zhao Y, Wang F and Tan G (2023), Clustering and machine learning-based integration identify cancer associated fibroblasts genes' signature in head and neck squamous cell carcinoma. *Front. Genet.* 14:1111816. doi: 10.3389/fgene.2023.1111816

Clustering and machine learning-based integration identify cancer associated fibroblasts genes' signature in head and neck squamous cell carcinoma

Qiwei Wang[1], Yinan Zhao[2], Fang Wang[3] and Guolin Tan[4]*

[1]Department of Otolaryngology Head and Neck Surgery, Third Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]Xiangya School of Nursing, Central South University, Changsha, Hunan, China, [3]Department of Otorhinolaryngology/Head and Neck Surgery, University Hospital Rechts der Isar, Technical University of Munich, Munich, Bavaria, Germany, [4]Third Xiangya Hospital, Central South University, Changsha, China

Background: A hallmark signature of the tumor microenvironment in head and neck squamous cell carcinoma (HNSCC) is abundantly infiltration of cancer-associated fibroblasts (CAFs), which facilitate HNSCC progression. However, some clinical trials showed targeted CAFs ended in failure, even accelerated cancer progression. Therefore, comprehensive exploration of CAFs should solve the shortcoming and facilitate the CAFs targeted therapies for HNSCC.

Methods: In this study, we identified two CAFs gene expression patterns and performed the single-sample gene set enrichment analysis (ssGSEA) to quantify the expression and construct score system. We used multi-methods to reveal the potential mechanisms of CAFs carcinogenesis progression. Finally, we integrated 10 machine learning algorithms and 107 algorithm combinations to construct most accurate and stable risk model. The machine learning algorithms contained random survival forest (RSF), elastic network (Enet), Lasso, Ridge, stepwise Cox, CoxBoost, partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), generalised boosted regression modelling (GBM), and survival support vector machine (survival-SVM).

Results: There are two clusters present with distinct CAFs genes pattern. Compared to the low CafS group, the high CafS group was associated with significant immunosuppression, poor prognosis, and increased prospect of HPV negative. Patients with high CafS also underwent the abundant enrichment of carcinogenic signaling pathways such as angiogenesis, epithelial mesenchymal transition, and coagulation. The MDK and NAMPT ligand–receptor cellular crosstalk between the cancer associated fibroblasts and other cell clusters may mechanistically cause immune escape. Moreover, the random survival forest prognostic model that was developed from 107 machine learning algorithm combinations could most accurately classify HNSCC patients.

Conclusion: We revealed that CAFs would cause the activation of some carcinogenesis pathways such as angiogenesis, epithelial mesenchymal transition, and coagulation and revealed unique possibilities to target glycolysis pathways to enhance CAFs targeted therapy. We developed an unprecedentedly

stable and powerful risk score for assessing the prognosis. Our study contributes to the understanding of the CAFs microenvironment complexity in patients with head and neck squamous cell carcinoma and serves as a basis for future in-depth CAFs gene clinical exploration.

KEYWORDS

HNSCC, cancer-associated fibroblasts (CAFs), machine learning, the tumor microenvironment, prognosis

Introduction

Head and Neck Squamous Cell Carcinoma (HNSCC) is an aggressive tumor associated with poor prognosis. There are more than 600,000 new cases are diagnosed worldwide each year (Bray et al., 2018). A hallmark signature of the tumor microenvironment (TME) in HNSCC is abundantly infiltration of cancer-associated fibroblasts (CAFs), which facilitate HNSCC progression (Custódio et al., 2020). CAFs could secrete exosomes which assist cell to-cell communication with TME thereby remodeling extracellular matrix (ECM) (Custódio et al., 2020). Biologically, the characteristics of cell stromal appears to be no difference between HNSCC patients, suggesting there is likely to exist a common weakness in stromal compartment which could be potential CAFs treatment targets (Puram et al., 2017).

Previous studies revealed an immunosuppressive role of CAFs, which could strongly induce the dysfunction of T cells and macrophages (Thomas and Massagué, 2005; Tauriello et al., 2018). The reason is attributable to CAFs secrete ECM components hence developing a dense fibrotic barrier in the tumor (Drifka et al., 2016). Benefit from bulk and single-cell RNA sequencing, A lots of new CAFs biomarkers have been figured out. Targeted therapy for CAFs also has made a breakthrough in hepatocellular carcinoma (Yin et al., 2019). However, some targeted CAFs clinical trials ended in failure, even accelerated cancers progression (Catenacci et al., 2015; Van Cutsem et al., 2020). Therefore, more comprehensive exploration of CAFs should solve the shortcoming and facilitate the targeted therapies in HNSCC.

Here, we collect 868 HNSCC samples from multi-dimensional common datasets, using clustering and machine learning method to detect the correlation between CAFs biological functions and clinical characteristics in HNSCC. We hope to find some specific molecular mechanisms to understand tumor progression and improve clinical management in head and neck squamous cell carcinoma.

Methods

HNSCC dataset source and processing

We summarized 31 CAFs genes from Kürten, C. H. L. et al. (Kürten et al., 2021) single-cell RNA sequencing research and Chakravarthy, A.et al. (Chakravarthy et al., 2018) bulk-RNA sequencing research. Total 868 samples from The Cancer Genome Atlas HNSCC (TCGA-HNSC) cohort and Gene Expression Omnibus cohort (GSE65858, GSE41613) were involved in our study. We constructed a Combined cohort by filtering common genes from GSE41613, GSE65858, and TCGA cohorts. We used "combat" R software package to remove the batch effects. The results were validated using internal and external cohort.

Construction of molecular types and score system based on the CAFs genes

We used R package "ClassDiscovery" to distinguish CAFs genes' expression pattern in the Combined cohort. The single-sample gene set enrichment analysis (ssGSEA) method was used to construct CAFs related score system CafS.

Estimation of immune infiltration

We used ssGSEA method and CIBERSORT algorithm (Newman et al., 2015) to evaluated absolute abundance of multiple immune cell populations. R package "ESTIMATE" was performed to calculated stromal score.

Single-cell analysis

We downloaded GSE164690 single cell cohort from Gene Expression Omnibus database. R package "Seurat" (Butler et al., 2018) was used to analysis single cell database. We filtered mitochondrial genes with parameter <10%. We selected highly variable genes with parameter nfeatures = 2000. These variable genes were used as inputs for PCA. Dims = 1:15 was used to FindNeighbors and resolution = 0.5 were used for FindClusters. We identified 18 primary clusters, and cluster analysis were performed by the RunUMAP function. We found differentially expressed genes (DEGs) for each cluster with parameters min.pct = 0.25 & thresh.use = 0.25. We compared DEGs and annotated CAFs (FAP, MMP11, PDGFRA, PDGFRB, ADAMTS2, SFPR2); Endothelial cell (PLVAP, KDR, PTPRB) in clusters. "Single R" package was used to annotate remaining clusters. MuSic deconvolution method (Wang et al., 2019) was used to calculate the proportion of CAFs. The CellChat method (Jin et al., 2021) was used to analysis cellular communication.

Construction and verification of the prognostic model

LASSO algorithm was first used to filter the candidate prognostic CAFs genes. We then integrated 10 machine learning algorithms and 107 algorithm combinations to construct most accurate and stable risk model. The machine learning algorithms contained random survival forest (RSF), elastic network (Enet), Lasso, Ridge, stepwise Cox, CoxBoost, partial least squares regression for Cox (plsRcox), supervised principal components

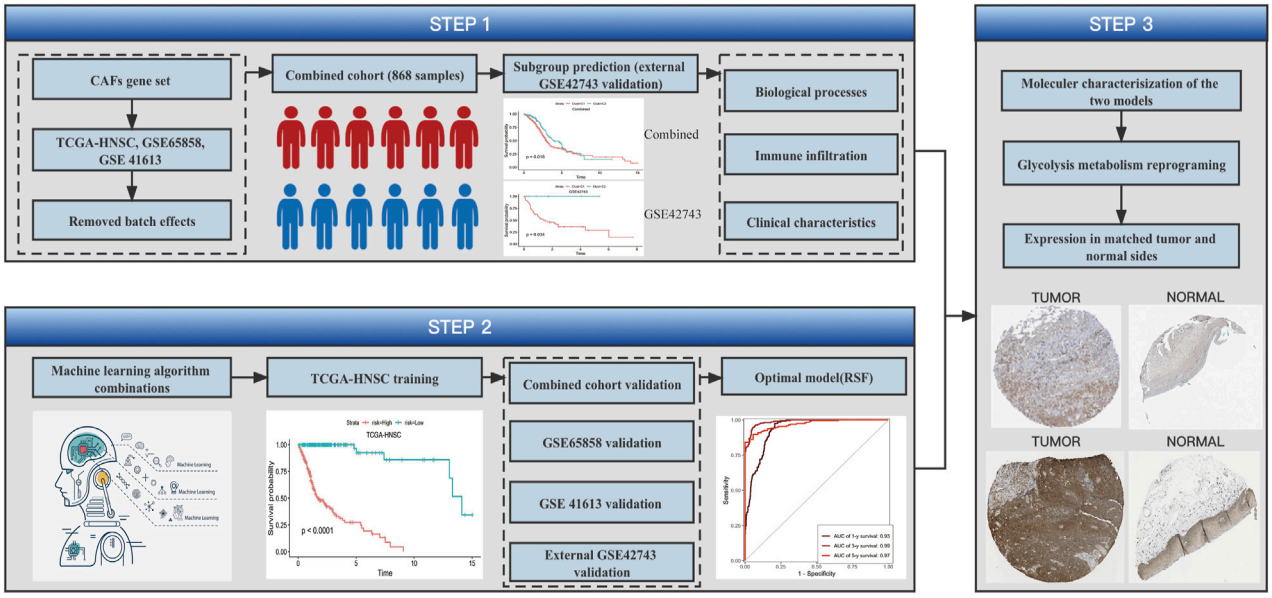


FIGURE 1
Workflow. The workflow of HNSCC cancer-associated fibroblasts signature analysis.

(SuperPC), generalised boosted regression modelling (GBM), and survival support vector machine (survival-SVM). All models were detected in four datasets (GSE41613, GSE65858, TCGA-HNSC, and Combined cohort). We calculated the concordance index (C-index) across all datasets, and the model with the highest average C-index was considered optimal. We used the optimal average C-index model machine learning algorithm to validate the robustness of prognostic model in the external cohort GSE42743.

Cell lines and quantitative real-time PCR assay

HNSCC cell lines CAL-27, FaDu and normal nasopharyngeal epithelial cell line (NP69) were obtained from National Collection of Authenticated Cell Cultures. For reverse transcription, 2 μg of total RNA was used to synthesize cDNA with a cDNA Synthesis Kit. β-actin was used as an internal control. The PGAM1 forward sequence of primer was 5-AAACGCAGGACAGTCTGATGC-3, and reverse sequence of primer was 5-CCGTCTGCAGCTACAACTCA-3. The ENO1 forward sequence of primer was 5-CGAGACCCAGTGGCT AGAAGTT-3, and reverse sequence of primer was 5-AAGTGCCAC CCAGAGAGGAC-3. The β-actin forward sequence of primer was 5-CATTAAGGAGAAGCTGTGCT-3, and reverse sequence of primer was 5-GTTGAAGGTAGTTTCGTGGA-3.

Statistical analysis

All statistical analysis and bioinformatics methods used R (V4.1.2, https://www.r-project.org/) or GraphPad Prism 9. 4 software. The correlation analysis was conducted using Pearson method. The Wilcoxon test were performed to compare continuous variables and ordered categorical variables.

Data and code availability statements

All datasets used in this study are available in public database. The codes supporting the conclusions of this article could provide by reasonable request to corresponding author.

Result

Workflow of this study

The study of HNSCC cancer-associated fibroblasts signature analysis is listed in Figure 1 workflow.

Enrichment analysis in 31 CAFs genes

We collected 31 CAFs genes from Kürten, C. H. L. et al. (Kürten et al., 2021) single-cell RNA sequencing research and Chakravarthy, A.et al. (Chakravarthy et al., 2018) RNA sequencing research. They are GAPDH, ENO1, ITGA6, PGK1, TGFBI, ACTN1, FTH1, KDELR2, CD82, SSR3, A2M, PHLDA1, TSC22D1, ISG15, PRSS23, PGAM1, SFRP2, PDGFRB, CEBPB, TNFRSF12A, MMP9, SNA12, ADAMTS2, MMP11, MMP12, COL4A6, STEAP1, ITGAX, ADAMTS14, TLL1 and COL4A4. Some of them have long been recognized as CAFs biomarker. For example, CAFs marker matrix metalloproteinase 11 (MMP11) can be delivered into gastric cancer cells to promote migration (Xu et al., 2019). In addition, CAFs could express MMP9 to enhance

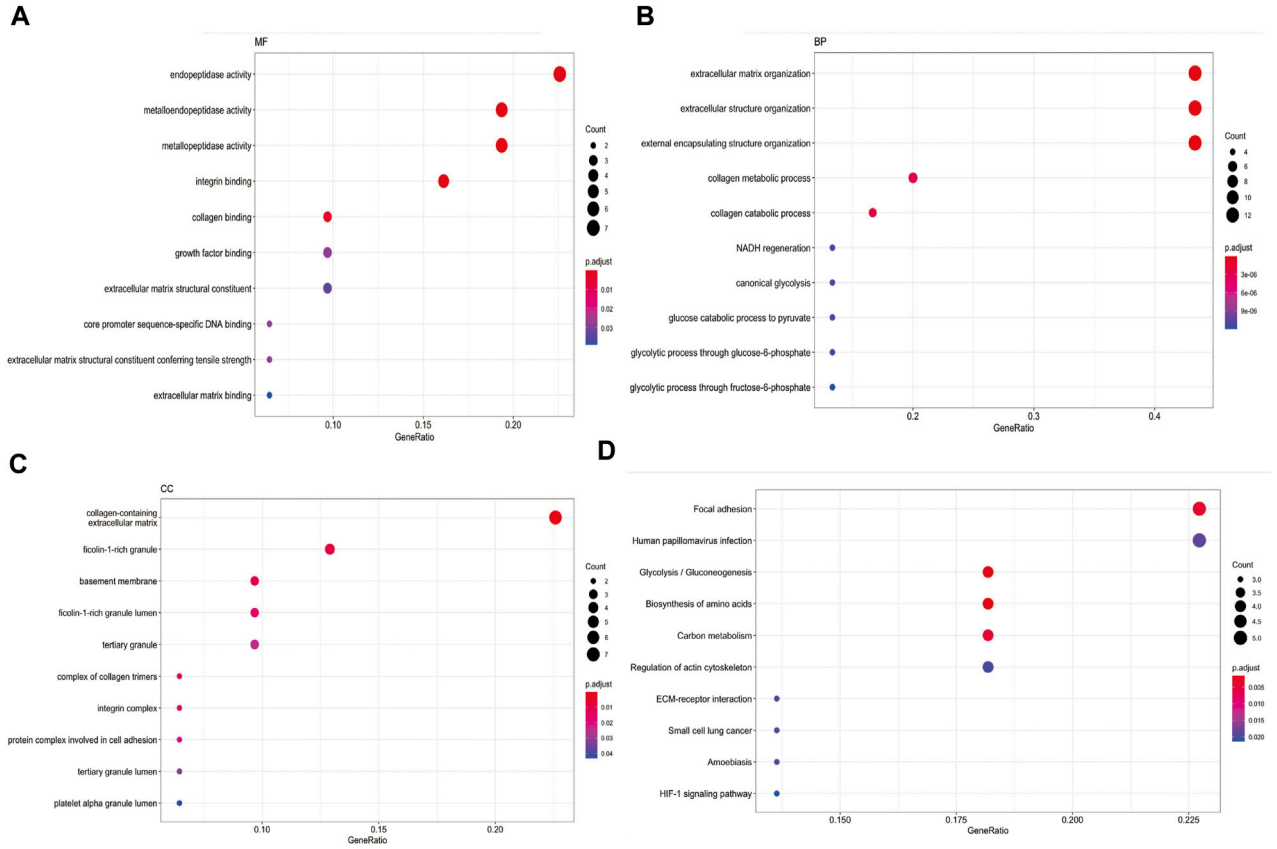


FIGURE 2
Enrichment analysis in 31 CAFs genes **(A)** MF function analysis for 31 CAFs genes. **(B)** BP function analysis for 31 CAFs genes. **(C)** CC function analysis for 31 CAFs genes. **(D)** KEGG pathway analysis for 31 CAFs genes.

proangiogenic phenotype thereby facilitating cancer cell invasion ability in HNSCC (Li et al., 2022). We used "clusterProfiler" R package (Yu et al., 2012) to plot enrichment landscape(Figures 2A–D). Gene Ontology (GO) analysis revealed 31 CAFs genes were mainly enriched in functions such as endopeptidase activity, extracellular matrix organization and collagen-containing extracellular matrix. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed 31 CAFs genes functions were mainly involved in pathway of focal adhesion. These enrichment function results indicated CAFs genes could play a cellular barrier role in medicine effects by regulating extracellular matrix (Lin et al., 2022).

Clustering analysis identified two CAFs patterns

We collected a total of 868 head and neck squamous cell carcinoma samples from TCGA and GEO cohorts to conjoint analysis, the tabular format of clinical sample information covered by this study, which are presented in Supplementary Figures S1A–C. We used "combat" software package to avoid the batch effects, the gene expression profile of each cohort is dispersive (Figure 3A), after elimination of the batch effects, the profile was agminated (Figure 3B). We identified two different CAFs patterns using R package "ClassDiscovery" and labeled as C1 and C2. We plotted a heat map which showed 31 CAFs genes was differential expression in Clust-C1 and Clust-C2 (Figure 3C). Then, after removing unreliable and incomplete clinical data, we analyzed survival prognosis between these two subtypes. The C1 cluster presented particularly survival disadvantage, conversely, the C2 cluster showed exceedingly survival benefit (log-rank, $p = 0.018$; Figure 3D). Similarly, this modification pattern also was observed in the external cohort of GSE42743, clustering analysis identified two similar CAFs related subtypes reminiscent of those observed in the previous combined cohort (Figure 3E). Clust_C1 also exhibits shorter survival than Clust_C2 (log rank $p = 0.034$, Figure 3F). These results suggested there might exist two CAFs related subtypes which could classify HNSCC patients' survival time.

Construct a score system CafS to evaluate 31 CAFs genes expression and classify HNSCC clinical characteristics

To further explored 31 CAFs genes expression functions in head and neck squamous cell carcinoma, we used the single-

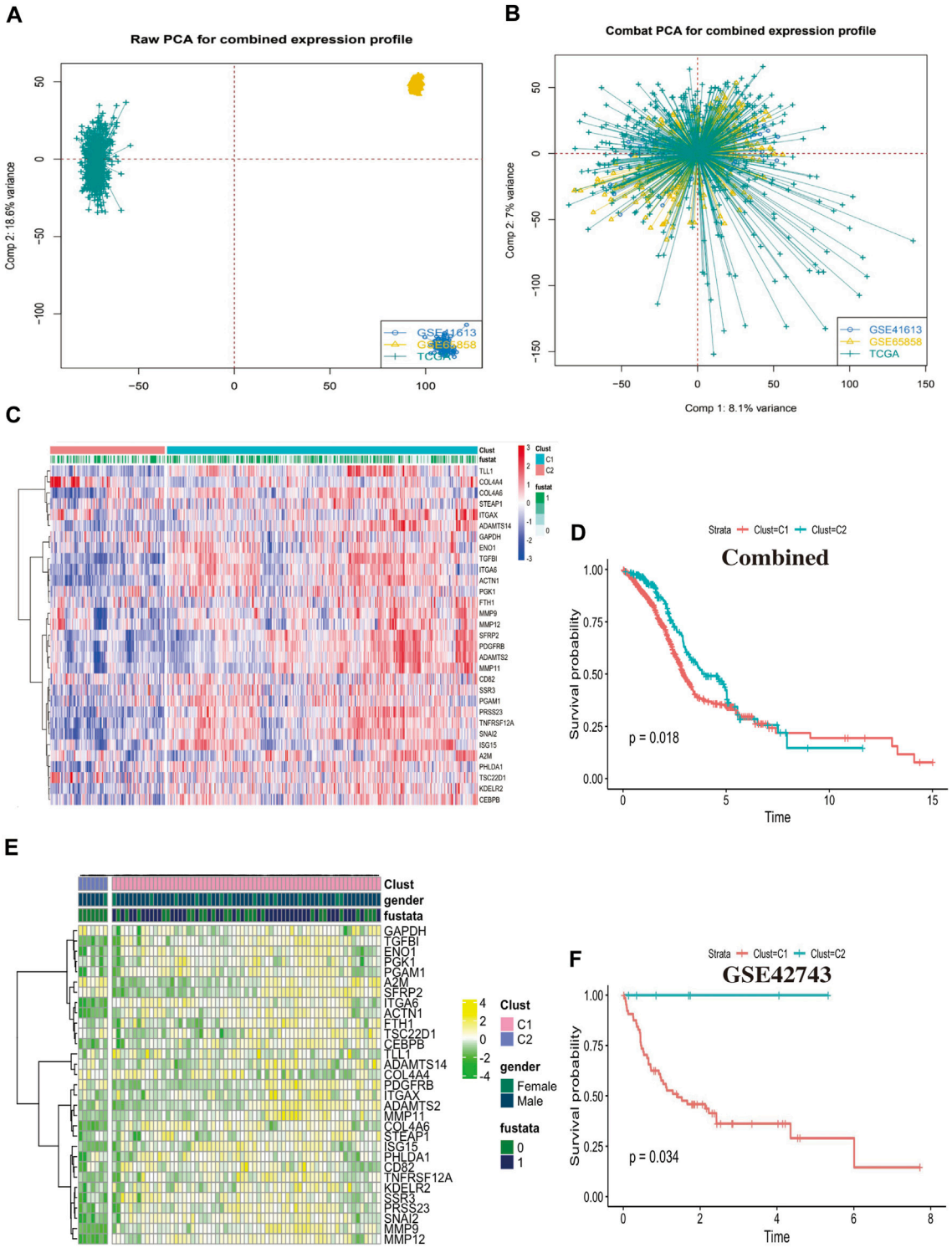


FIGURE 3
Clustering analysis identified two CAFs patterns. **(A)** Principal component analysis (PCA) showed the gene expression profile in the combined cohort, before elimination of the batch effects. **(B)** Principal component analysis (PCA) showed the gene expression profile in the combined cohort after elimination of the batch effects. **(C)** The heatmap displays the differential expression between the two groups of the 31 cancer associated fibroblasts (CAFs) genes, C1 cluster, C2 cluster, "1" means dead, "0" means alive, "fustat" means survival status. **(D)** The Kaplan-Meier plot displays significant differences survival rate among the two kinds of CAFs phenotypes in the Combined cohorts. C1 was worse than C2 (log rank p = 0.018), unit of Time (years). **(E)** The heatmap displays the differential expression between the two groups of the 31 cancer associated fibroblasts (CAFs) genes in the external cohort GSE42743. **(F)** The Kaplan-Meier plot displays the same trend of significant differences survival rate among the two kinds of CAFs phenotypes in the external cohort GSE42743 (log rank p = 0.034).

10.3389/fgene.2023.1111816

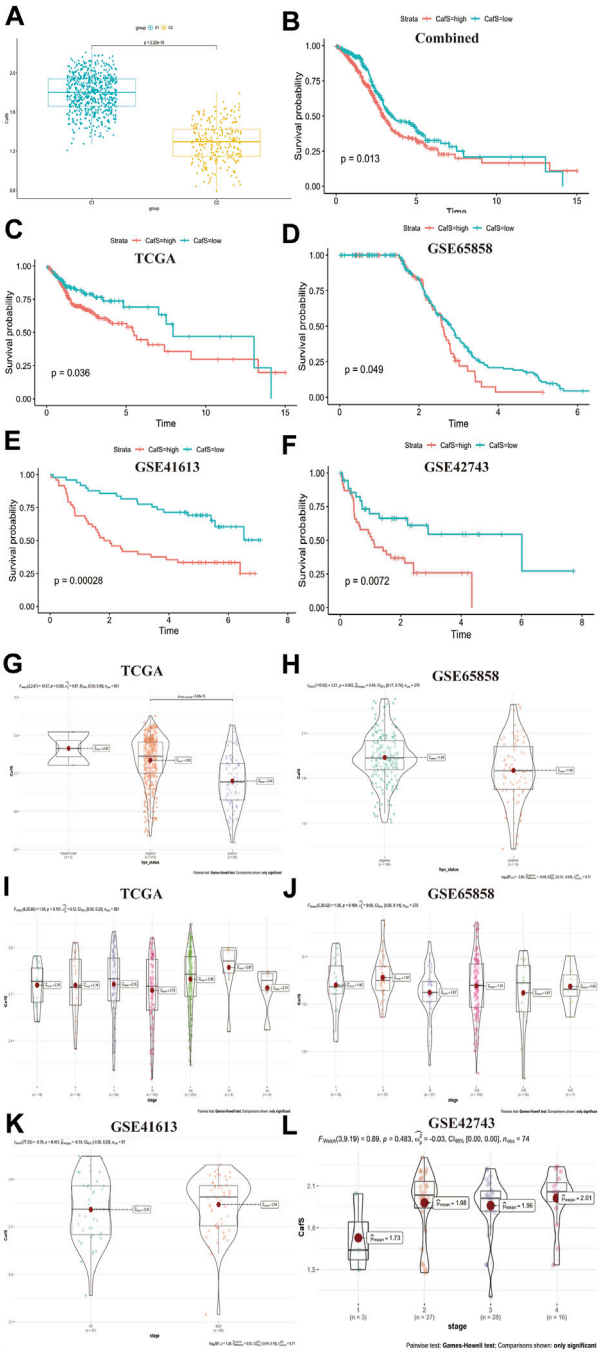


FIGURE 4
Construct a score system named CafS to evaluate 31 CAFs genes expression and classify HNSCC clinical characteristics **(A)** CafS in the groups of C1 and C2; combined database; $p < 2.2e\text{-}16$. **(B–F)** The Kaplan-Meier plot displays significant differences of survival time among the high-CafS and low-CafS groups in the Combined, TCGA, GSE65858, GSE41613, and GSE42743 cohort, respectively. High group was worse than low group, log rank $p = 0.013$, 0.036, 0.049, 0.00028, 0.0072. **(G)** CafS in TCGA cohort among the group of HPV positive and HPV negative, $p = 6e\text{-}10$. **(H)** CafS in GSE65858 cohort among the group of HPV positive and HPV negative, $p = 0.002$. **(I–L)** CafS in TCGA, GSE65858, GSE41613, and external cohort GSE42743; among the group of stages; respectively, ($p = 0.101$, 0.194, 0.451, 0.483).

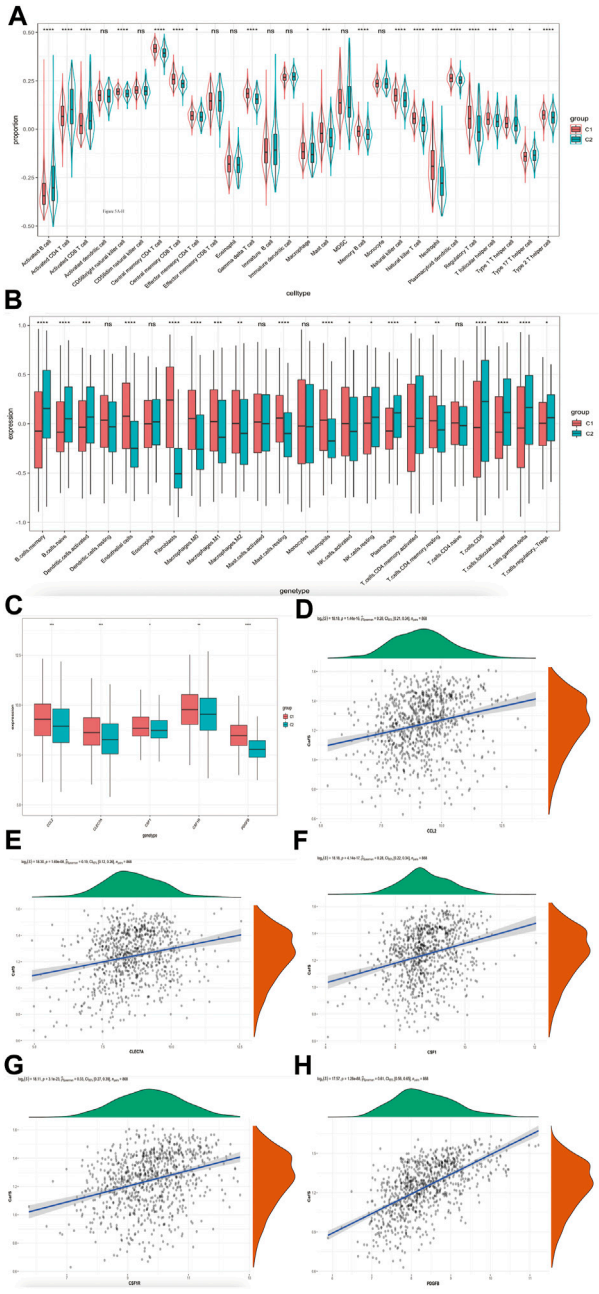


FIGURE 5
CafS changes the tumor immune microenvironment and is related to tumor associated macrophage (TAM) **(A)** Enrichment of 28 immune cell types infiltrating in the groups of CafS; combined database; the asterisk represents the different p values (* <0.05; ** <0.01; *** <0.001, **** <0.0001). **(B)** Boxplot of 24 immune cell types infiltrating by CIBERSORT algorithm in the groups of CafS; combined database; the asterisk represents the different p values (* <0.05; ** <0.01; *** <0.001, **** <0.0001). **(C)** Five genes' expression of tumor associated macrophage (TAM) in the groups of CafS; the asterisk represents the different p values (* <0.05; ** <0.01; *** <0.001, **** <0.0001). **(D–H)** Correlation between CafS and CCL2, CLEC7A, CSF1, CSF1R and PDGFB (CafS and CCL2: r = 0.28, $p = 1.44e\text{-}16$; CafS and CLEC7A: r = 0.19, $p = 1.69e\text{-}08$; CafS and CSF1: r = 0.28, $p = 4.14e\text{-}17$; CafS and CSF1R: r = 0.33, $p = 3.1e\text{-}23$; CafS and PDGFB: r = 0.61, $p = 1.28e\text{-}88$).

sample gene set enrichment analysis (ssGSEA) method to construct a score system CafS which represented the quantification of these 31 CAFs genes. We found CafS in C1 cluster was significantly higher than C2 (*t*-test, $p < 2.22e\text{-}16$; Figure 4A). According the CafS, we divided survival cohort samples into high and low group by the optimal cut-off value, we found CafS was a prognostic factor (log-rank, $p = 0.013$; Figure 4B). In the internal TCGA, GSE41613 and GSE65858 cohorts, high CafS indicated worse survival (log-rank, $p = 0.036$; $p = 0.00028$; $p = 0.049$; respectively, Figures 4C–E). In the external cohort, high CafS also predicted bad outcoming (Figure 4F). These results proved poor prognosis for patients with high CafS. Ang, K. K. et al. (Ang et al., 2010) found there was significant survival advantages in HNSCC patients with HPV (+) comparing to HPV (-). In our study, we found that the CafS level in the HPV (-) group was significantly higher than that in the HPV (+) group (p = 6e-10; $p = 0.002$; Figures 4G, H). We subsequently investigated the tumor mutation burden (TMB) in the groups C1 and C2 from TCGA database, but no statistically significant difference was found between them ($p = 0.22$, Supplementary Figures S2A, B).Tobacco use may contribute to the distribution of CafS, but we did not observe a significant difference between the smoking and non-smoking groups (TCGA, $p = 0.76$, GSE65858, $p = 0.84$, Supplementary Figures S3A, B). Moreover, we evaluated the CafS levels across all stages of head and neck squamous cell carcinoma (HNSCC), the results showed no statistical difference in CafS among different HNSCC stages (Figures 4I–L, $p = 0.101, 0.194, 0.451, 0.483$).

High CafS changes the tumor immune microenvironment and is related to tumor associated macrophage (TAM)

We explored the relevence between the immune cell infiltration and CafS in the groups of C1 and C2. According to Bindea, G. et al. (Bindea et al., 2013) study, we calculated 28 immune cells value by the method of ssGSEA. Our results showed the proportion of activated CD4 T cell and activated CD8 T cell were significantly higher in the C2 cluster, on the contrary, the related macrophages infiltration was exceedingly higher in the C1 cluster (Figure 5A). We used CIBERSORT algorithm (Newman et al., 2015) to further detect the differential immune infiltration in the clusters C1 and C2. The results showed the expression of activated CD4 T cell and CD8 T cell were higher but macrophages (including M0, M1 and M2 status) were lower in C2 (Figure 5B). It reflected that high CafS may prevent immune cell cytotoxic effects but promote immune cell inflammatory effects in head and neck squamous cell carcinoma. M2 status macrophage often referred as tumor associated macrophage (TAM) which promote tumor growth, invasion, and metastasis (Ovais et al., 2019). So, we collected TAM markers from previous studies (Ngambenjawong et al., 2017; Jiang et al., 2022), they are CCL2, CLE7A, CSF1, CSF1R and PDGFB. We detected these TAM markers expression in the combined cohort, found all of them were significantly higher expression in high CafS group (Figure 4C). The correlation plots showed CafS was significantly positive correlated with these five TAM markers in combined database (CCL2: r = 0.28, $p = 1.44e\text{-}16$; CLEC7A: r = 0.19, $p = 1.69e\text{-}08$; CSF1: r = 0.28, $p = 4.14e\text{-}17$; CSF1R: r = 0.33, $p = 3.1e\text{-}23$; PDGFB: r = 0.61, $p = 1.28e\text{-}88$; respectively, Figures 5D–H).

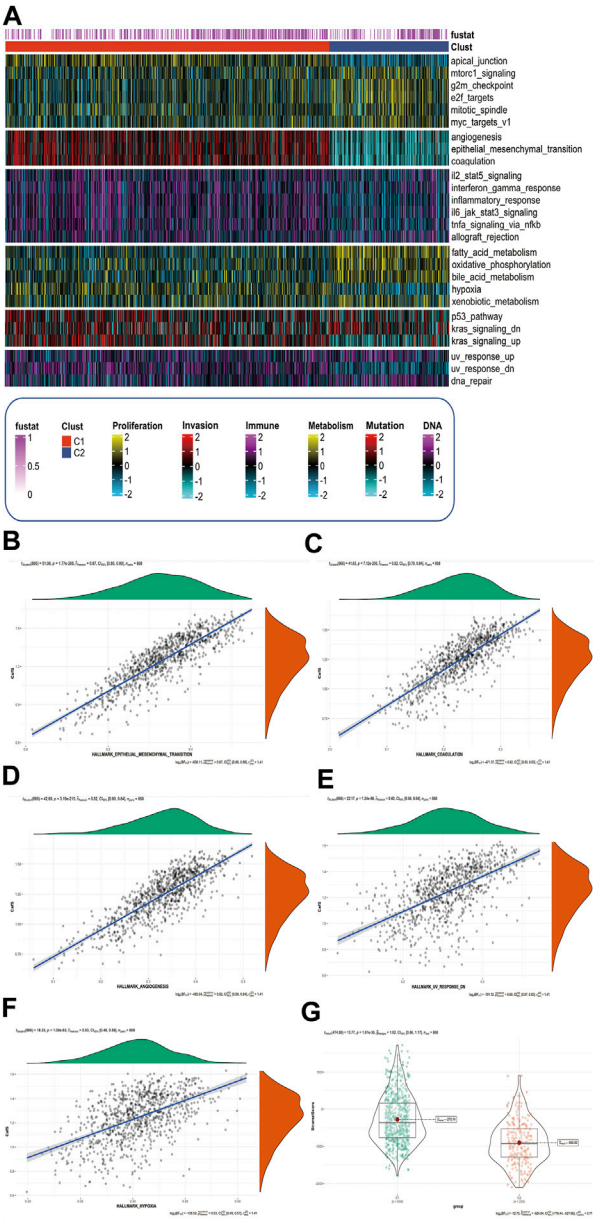


FIGURE 6
CafS changes hallmark signaling pathway and promotes the ability of tumor invasion **(A)** Complex-heatmap display the landscape in the combined cohort; the panel display the expression of hallmark signaling pathway involved in different CafS group; Proliferation, Invasion, Immune, Metabolism, Mutation and DNA represented six different pathway modules which arrange from the top down in complex-heatmap and they are labeled at the bottom of it; C1 represented high Cafs group and C2 represented low Cafs group; "fustat" means survival status, "1" means dead, "0" means alive. **(B–F)** Correlation between CafS and EMT, coagulation, angiogenesis, hypoxia, and uv-response-down pathway (CafS and EMT: r = 0.87, $p = 1.77e\text{-}268$; CafS and Coagulation: r = 0.82, $p = 7.12e\text{-}209$; CafS and Angiogenesis: r = 0.82, $p = 3.19e\text{-}215$; CafS and Hypoxia: r = 0.53, $p = 1.09e\text{-}63$; CafS and Uv-response-down: r = 0.60, $p = 1.24e\text{-}86$). **(G)** StromalScore in the groups of CafS; combined database; $p = 1.61e\text{-}36$.

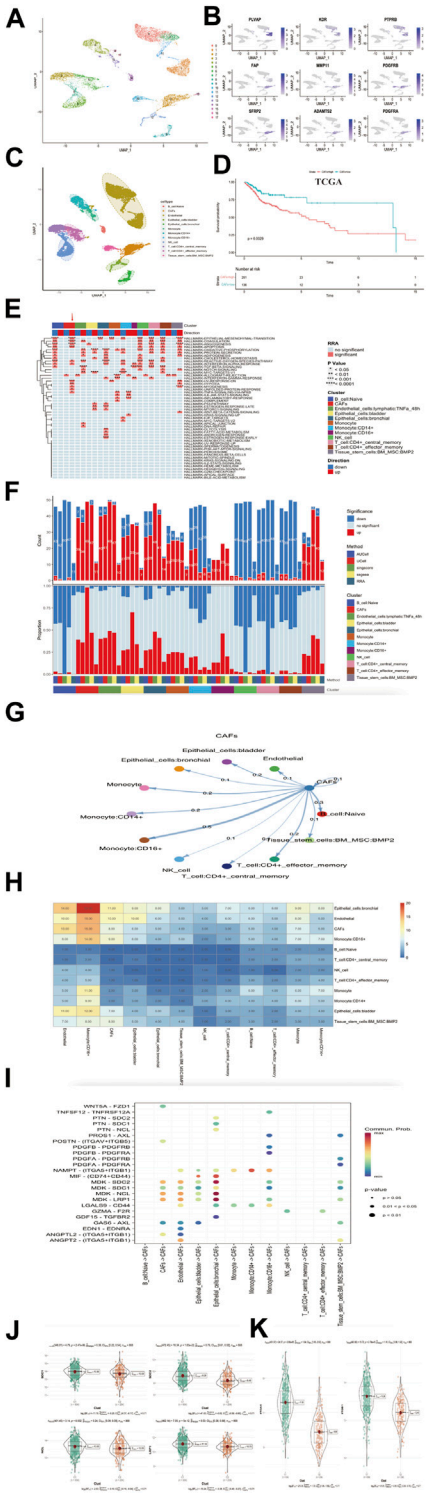


FIGURE 7
Verification of CafS clinical characteristic and biological function in single-cell RNA sequencing database **(A)** UMAP plot of selected 10244 single cells in tumor and non-immune stromal cells (CD45 negative). Different colors represent different cell types. **(B)** UMAP plot showed the expression of endothelial cell and cancer associated fibroblasts cell. **(C)** UAMP plot of selected 10244 single cells in tumor and non-immune stromal cells (CD45 negative). 18 cell clusters were divided into 12 cell types. **(D)** The Kaplan-Meier plot displays significant differences of survival time among the high-CAFs proportion and low-CAFs proportion in TCGA cohorts. Deconvolution method. High proportion group had worse overall time than low. **(E)** Display of the landscape of signaling pathways in different cell clusters; the panel display the hallmark signaling pathway involved in different cell clusters; Cluster (red module represent CAFs cell type), *p*-value and Direction of up or down are labeled at the right of plot; RRA represent significance. **(F)** Display of the landscape of gene set up-regulation or down-regulation in different cell clusters; Cluster (red module represent CAFs gene set), Method and Significance are labeled at the right of plot. **(G)** The differential cell–cell cellular communication shows CAFs weight coefficient between all cell types. **(H)** The heatmap of cell–cell cellular communication shows the counts of CAFs between all cell types. **(I)** Communication network of the significant ligand-receptor pairs between CAFs and other cell types, which contribute to the signaling from CAFs to Naive B cell, Endothelial Cell, Epithelial cell, Monocyte cell, NK cell, $CD4^+$ T cell and Tissue stem cell subpopulations. Dot color reflects communication probabilities and dot size represents computed *p*-values. Empty space means the communication probability is zero. *p*-values are computed from one-sided permutation test. **(J)** SDC1, SDC2, NCL, LRP1 in the groups of Combined cohorts; $p = 2.47e\text{-}06$; $p = 1.03e\text{-}22$; $p = 0.002$; p = 5e-12. **(K)** ITGA5 and ITGB1 in the groups of combined cohorts; $p = 2.88e\text{-}85$; $p = 2.78e\text{-}45$.

CafS changes hallmark signaling pathway and promotes the ability of tumor invasion

We used GSVA method to analysis the characteristics of the associated signaling pathways in different CafS subtypes. The hallmark signaling pathway gene set was download from The Molecular Signatures Database (Liberzon et al., 2015) (MSigDB, https://www.gsea-msigdb.org/). We found high CafS in C1 had a remarkable enrichment in tumor invasion-related pathways such as angiogenesis, epithelial mesenchymal transition, and coagulation. In addition, we found there were different enrichments in pathways including tumor proliferation-related, tumor immune-related, tumor metabolism-related, tumor mutation-related, and tumor DNA damage-related (Figure 6A). We explored the relationship between CafS and tumor invasion-related pathways to further understand the mechanism of tumor process, we found CafS was significantly positive correlated with these tumor invasion-related, DNA damage-related and metabolism-related signaling pathways (EMT: r = 0.87, $p = 1.77e\text{-}268$; Coagulation: r = 0.82, $p = 7.12e\text{-}209$; Angiogenesis: r = 0.82, $p = 3.19e\text{-}215$; Hypoxia: r = 0.53, $p = 1.09e\text{-}63$; Uv-response-down: r = 0.60, $p = 1.24e\text{-}86$; respectively; Figures 6B–F). Fibroblasts contributed to a dominant component of the tumor stroma (Kalluri and Zeisberg, 2006), so, we used "ESTIMATE" R package to quantify the scores of stromal: the "StromalScore". We found the StromalScore is profoundly higher in C1 than C2 group (Figure 6G).

Verification of CAFs characteristics in single-cell RNA sequencing database

We used R package "Seurat" (Butler et al., 2018) to analysis HNSCC single cell database. We selected CD45 negative as tumor and non-immune stromal cells to elucidate the heterogeneity of head and neck squamous. After quality control and filtering, we identified 10,244 cells from five head and neck squamous cell patients. We distinguished 18 distinct clusters based on a resolution value 0.5

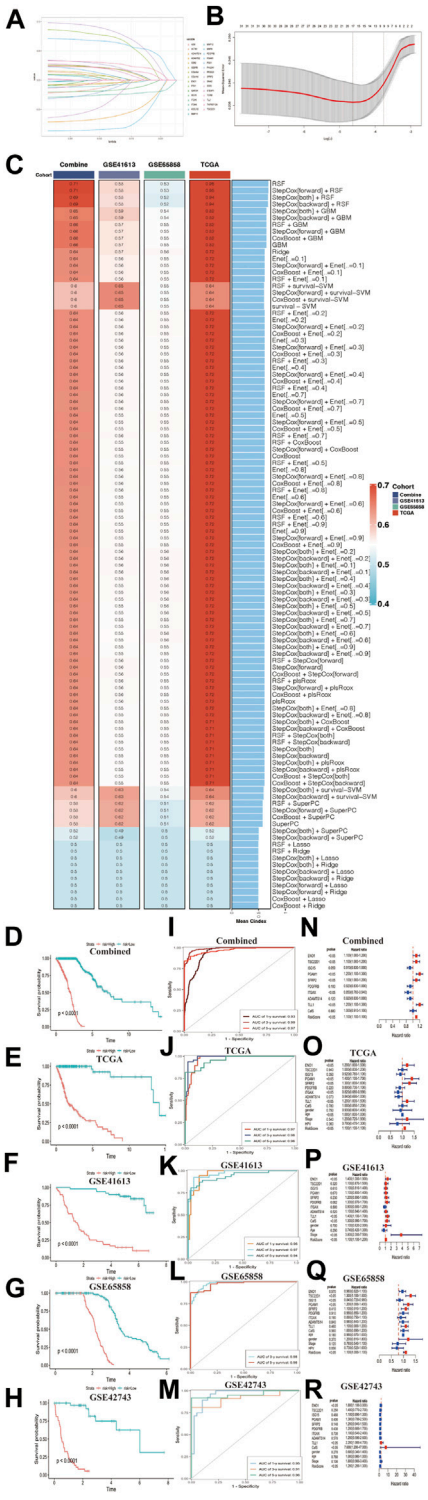


FIGURE 8
Construction and verification to the CAFs risk prediction model using machine learning methods **(A)** LASSO coefficient profiles of 31 cancer associated fibroblasts marker genes in Combined cohort. **(B)** 1000-time cross-validation for tuning parameter selection in the LASSO model; Combined cohort. **(C)** A total of 107 kinds of prediction models *via* machine learning and further calculated the C-index of each model across training and all validation cohorts. **(D–H)** Kaplan–Meier curves of overall survival according to the median risk score in Combined, TCGA-HNSC, GSE41613, GSE65858, and GSE42743 external validation cohorts. All log-rank $p < 0.0001$. (*Continued*)

FIGURE 8 (Continued)
(I–M) Time-ROC value in Combined, TCGA-HNSC, GSE41613, GSE65858, and GSE42743 external validation cohorts. **(I)** Combined, AUC one-three-five years = 0.93, 0.99, 0.97. **(J)** TCGA-HNSC, AUC one-three-five years = 0.97, 0.98, 0.96. **(K)** GSE41613, AUC one-three-five years = 0.96, 0.97, 0.94. **(L)** GSE65858, AUC one-three years = 0.98. **(M)** GSE42743, AUC one-three-five years = 0.95, 0.91, 0.96. **(N–R)** Multivariate Cox regression of risk score regarding to OS in the Combined, TCGA-HNSC, GSE41613, GSE65858 and GSE42743. Statistic tests: two-sided Wald test. Data are presented as hazard ratio (HR) ± 95% confidence interval [CI].

(Figure 7A). We labeled cell types as endothelial cell (Endothelial, gene expression of PLVAP, KDR, PTPRB) and cancer associated fibroblasts cell (CAFs, gene expression of FAP, MMP11, PDGFRB, SFRP2, PDGFRA, ADAMTS2; Figure 7B). In addition to endothelial and cancer associated fibroblasts cell types classified above, we used "Single R" package to identify several other distinct clusters, they were b-cell Naïve, epithelial-cells bladder, epithelial-cells bronchial, monocyte, monocyte:CD14+, monocyte:CD16+, NK cell, $CD4^+$ central memory T cell, $CD4^+$ central effector T cell and tissue-stem-cells:BM_MSC:BMP2 (Figure 7C). To further understand the characteristics for the CAFs, we performed a deconvolution method (Wang et al., 2019) to calculate the bulk tissue proportion of CAFs in TCGA cohort with this single cell RNA sequencing database reference (Supplementary Table S1). Combining clinical data of TCGA-HNSC, we found higher CAFs proportion indicated significant poorer prognosis for head and neck squamous cell carcinoma (log rank $p = 0.0029$, Figure 7D). This result validated the high CAFs proportion may be identical to the high CafS as a prognostic indicator in HNSCC.

We next used irGSEA (https://github.com/chuiqin/irGSEA) package to analysis the associated signaling pathways in different cell clusters and focused more on CAFs type. The results exhibited CAFs cluster had a remarkable up-enrichment in tumor progression-related pathways such as angiogenesis, epithelial mesenchymal transition, coagulation, hypoxia and uv-response-down (Figure 7E), which were observed as the same to high CafS in group C1 (Figure 6A). We also detected CAFs cluster gene set expression, undoubtedly, CAFs cluster gene set was up-regulation (Figure 7F). These results illustrated up-regulation of CAFs genes could play a precondition role in activating specific signaling pathways such as angiogenesis, epithelial mesenchymal transition, coagulation, hypoxia, and uv-response-down, etc. This alternation influenced the tumor microenvironment and leaded to poorer prognosis in head and neck squamous cell carcinoma patients.

To further detect the enrichment of CAFs populations in HNSCC cells, we hypothesized that those CAFs populations might be functionally distinct across other different cell type. We hence performed the ligand–receptor-based cell-cell cellular cross-talk analysis (Jin et al., 2021). The plot showed the different weight coefficient distribution and counts frequency of CAFs to others cellular cross-talk (Figures 7G–H). These results suggested that HNSCC CAFs cells could preferentially reprogram and induce their specific functional status-likely explained by the specificity between genes' differential expression, which could directly impact TME. We used the same method (Jin et al., 2021) to distinguish the signaling of ligand–receptor interactions network in HNSCC cells.

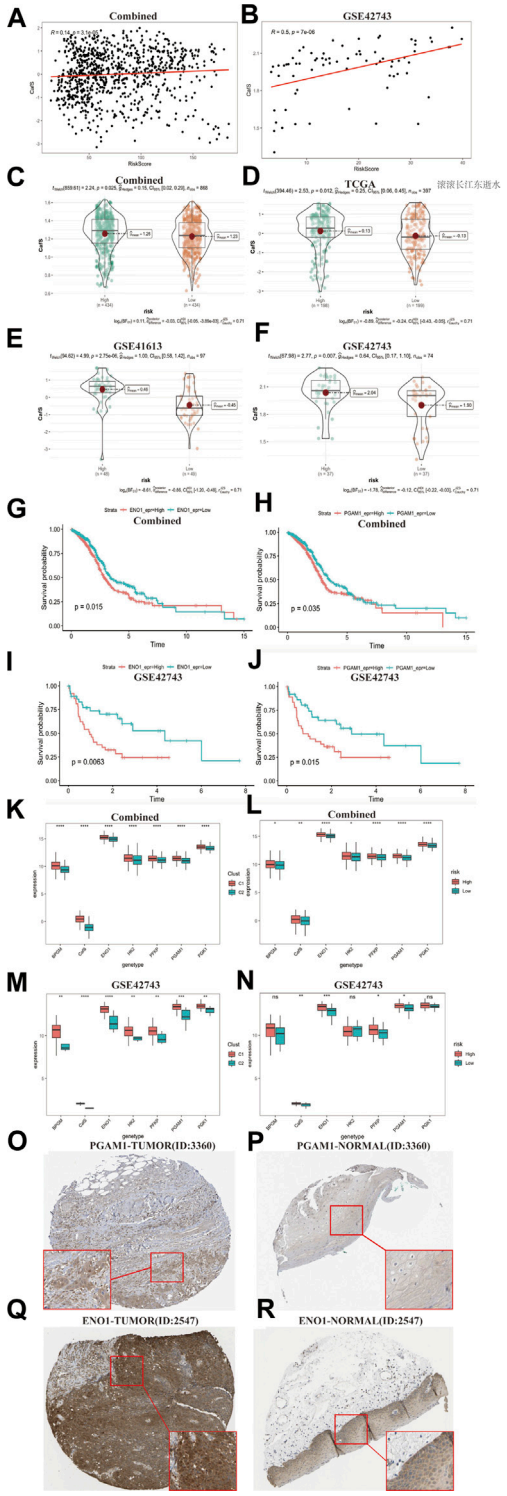


FIGURE 9
Relationship between CafS classification pattern and CAFs risk score model **(A, B)** The correlation between CafS and risk score in the combined and external cohort GSE42743; Combined, r = 0.14, p = 3.1e-05; GSE42743, r = 0.5, p = 7e-06. **(C–F)** CafS in the high and low risk groups in the combined, TCGA-HNSC, GSE41613 and GSE42743 cohort p = 0.025, 0.012, 2.57e-06, 0.007, respectively. **(G–J)** Kaplan–Meier curves of overall survival according to the median of ENO1 and PGAM1 expression in the Combined and GSE42743 external validation cohorts. Combined, ENO1 and PGAM1, log-rank p = 0.015, 0.035, respectively. GSE42743, ENO1 and PGAM1, log-rank p = 0.0063, 0.015, respectively. **(K–N)** ENO1, PGAM1, HK2, PFKP, BPGM, PGK1 and CafS expression in the CafS classification model and risk score model; Combined and external GSE42743 cohorts; the asterisk represents the different p values (* <0.05; ** <0.01; *** <0.001, **** <0.0001). **(O–R)** Immunohistochemical of PGMA1 and ENO1 in the matched tumor and normal side.

We identified MDK and Nicotinamide phosphoribosyl transferase (NAMPT) ligand–receptor pairs contributing to the most communication from CAFs to each HNSCC cell type (Figure 7I). In combined bulk cohort, we further vitrificated high expression of these related ligand–receptor genes in high CafS group (Figures 7J–K). Therefore, these ligand–receptor pairs specifically enriched in HNSCC TME maybe provide a clue for targeted therapy.

Construction and verification to the CAFs risk prediction model using machine learning methods

Several studies have proved that CAFs genes were biomarker for prognostic in many types of cancer (Wen et al., 2019; Li et al., 2021a; Shelton et al., 2021). So, we used the Lasso algorithm to filter the most candidate prognostic CAFs genes from the classification model (Figures 8A, B). Considering the convenience for the future clinical testing, we selected 9 CAFs genes (including ENO1, TSC22D1, ISG15, PGAM1, SFEP2, PDGFRB, ITGAX, ADAMTS14 and TLL1) and CafS to construct risk model. We set TCGA-HNSC database as training cohort; the combined cohort, GSE65858, and GSE41613 cohorts as validation datasets. GSE42743 cohort was used to external verification. In TCGA-HNSC cohort, we first fitted 107 kinds of prediction models *via* the 10 machine learning algorithms and further calculated the C-index value of each model across all validation datasets (Figure 8C). Interestingly, the optimal training model with the highest C-index value (0.95) was designed by random survival forest (RSF) algorithm, and this model also present highest average C-index value (0.61) in all validation cohorts (Figure 8C). Next, a risk score for each patient was calculated using the "predict" function in this RSF model, according to their median risk score, all patients were divided into high- and low-risk groups. The Kaplan–Meier curve showed patients in the high-risk group had significantly dismal overall survival (OS) compared to the low-risk group in the TCGA-HNSC training dataset and four validation datasets (Figures 8D–G, all log rank p < 0.0001). The trend of this finding was validated in the external cohort GSE42743 using the same method (Figure 8H, log rank p < 0.0001). Time receiver operating characteristic (ROC) method was applied to verify the sensitivity and specificity to the risk model. As we had expected, the results demonstrated that all datasets had remarkably delight Time-ROC values (combined, AUC one-three-five years = 0.93, 0.99, 0.97; TCGA-HNSC, AUC one-three-five years = 0.97, 0.98, 0.96; GSE41613, AUC one-three-five years = 0.96, 0.97, 0.94; GSE65858, AUC one-three years = 0.98; GSE42743, AUC one-three-five years = 0.95, 0.91, 0.96; Figures 8I–M). Those results

indicated this CAFs related risk model had considerably predictive significance. Multivariate Cox regression demonstrated that risk score remained statistically significant (all p-value < 0.05) in all cohorts after adjusting for available clinical traits, such as age (less than 60 vs. be equal or greater than 60); gender (Female vs. Male); stage and HPV status, the results suggested that risk score is an independent predict factor for overall survival (Figures 8N–R).

Relationship between CafS classification pattern and CAFs risk score model

As the results showed above: the high CafS and CAFs related high-risk score both indicate worse survival, we assumed that those HNSCC populations with high-risk score seem to combine with high CafS. Hence, we generated a correlation map which showed CafS was significantly positive correlated with the risk score in the combined and external validation cohort (r = 0.19; p = 3.1e-05; r = 0.5, p = 7e-06; Figures 9A, B). We further calculated the CafS in the groups of risk model and found that the model risk score might fit CafS distribution in the combined and validation cohorts (combined, p = 0.025; TCGA-HNSC, p = 0.012; GSE41613, p = 2.75e-06; GSE42743, p = 0.007; Figures 9C–F). These results confirmed our hypothesis that there is a certain degree of interaction between CafS and risk score thereby bringing dark survival to HNSCC patients. In the Multivariate Cox regression model, we found ENO1 and PGAM1were glycolysis enzyme markers (Huang et al., 2022a; Yang et al., 2022) may contribute to bad outcomes. According to the gene median expression, the ENO1 and PGAM1 were significant predictors in both combined and validation cohorts (combined, ENO1, log-rank p = 0.015, PGAM1, log-rank p = 0.035; GSE42743, ENO1, log-rank p = 0.0063, PGAM1, log-rank p = 0.015; Figures 9G–J). As the results described above, we infer that high CafS and risk score could induce glycolysis reprograming, we hence collected other four crucial glycolysis enzyme biomarkers from previous study (Warmoes and Locasale, 2014), they are BPGM, HK2, PFKP and PGK1. In the CafS classification model, the results showed these genes' expression presented higher in the group of C1 (Figures 9K, L). In the risk model, glycolysis enzyme marker genes expression also was observed increased trend in the combined cohort, but not all were observed high expression in the GSE42743 external validation cohort, possibly due to the sample size (Figures 9M, N). Based on matched tumor and normal tissues from the patients in Human Protein Atlas, we found both PGAM1 and ENO1 were up-regulation in tumor side (Figures 9O–R). The qPCR assay also validated that PGAM1 and ENO1 were over expression in HNSCC cell lines compare to nasopharyngeal epithelial cell line (Supplementary Figures S4A–D). Hence, the above results provide us a clue for targeted glycolysis reprograming therapy might make breakthroughs in CAFs treatment.

Discussion

At present, a lot of studies only use TCGA or single GSE cohort as data sources to analysis malignant tumor, which are short of sample size and beyond to the accuracy and effectiveness for medical practice. In our study, we collected 868 cases to explore the molecular actions of CAFs in head and neck squamous cell carcinoma. We constructed a robust CAFs related classification and score system using 31 CAFs marker genes, which could effectively predict the prognosis of the HNSCC patients. We found there were remarkable discrepancy in clinical and biological characteristics such as HPV status and TAMs among different CafS clusters. HPV-negative head and neck tumors patients was confirmed with terrible prognoses (Johnson et al., 2020). In our study, high CafS patient was more likely to be HPV-negative, indicated that CafS could exert adverse impact on clinical outcomes. Tumor associated macrophages marker genes played a crucial role in tumor process, for example, high expression CCL2 in macrophages could promote HNSCC invasion and metastasis (Ling et al., 2022). CLEC7A also called Dectin1, Daley, D. et al. (Daley et al., 2017) found it activated macrophages and promotes pancreatic ductal adenocarcinoma progression. CSF1/CSF1R signaling axis had been proved induced macrophages to M2 polarization and promoted tumor growth and lung metastasis (Fujiwara et al., 2021). PDGFB as a platelet activation factor for promoting tumor metastasis by recruitment of TAMs (Yang et al., 2016). In our study, we showed these five TAMs markers not only were exceedingly high expression in C1 group but closely associated with CafS, these results illustrated high CafS was associated with abundant M2 macrophages enrichment and provided us an expanded knowledge for the CAFs genes' role in the tumor microenvironment.

Analysis for the associated signaling pathways in different CafS groups revealed interesting findings. First, high CafS represented extensively activation in pathways such as angiogenesis, epithelial mesenchymal transition, and coagulation, all these pathways enhanced tumor cell malignancy (Nash et al., 2001; Zhang et al., 2021a; Huinen et al., 2021). UV-response-down pathway was a process that organism undergo UV-B or UV-A radiation may generate genomic mutations and instability leading to tumorigenesis (Tan et al., 2019). In this study, we found uv-response-down was up-regulation in C1 group, it broadened our horizons of CAFs carcinogenesis. Hypoxia pathway could activate multiple genes' expression which participate in iron metabolism, glucose transport, cell proliferation thereby resulting in poor prognosis of treatment (Nordgren and Tavassoli, 2011). Our result showed high CafS aggravate head and neck squamous cell carcinoma hypoxic condition and displayed a remarkable correlation between these five tumor-related signaling pathways and CafS, which enhanced our comprehension for CAFs promoting HNSCC proliferation and metastasis. In addition, high CafS was characterized by leading stromal score, hence are more likely to lead to tumor capped by extracellular matrix and induce immunosuppression (Buchheit et al., 2014).

Single cell RNA sequencing revealed head and neck squamous cell complexity and heterogeneity. We identified CAFs clusters and other 11 distinct cell types. Similar to the bulk tissue sequencing results, our results validated CAFs population possess the characteristics of strong cancer-promoting signatures, it indicated angiogenesis, epithelial mesenchymal transition, coagulation, uv-

response-down and hypoxia pathways were up-regulated in this cell cluster. The up regulation of CAFs gene set profile contribute to the activation of these related signaling pathways. Moreover, the deconvolution result showed high CAFs proportion, just like high Cafs, robustly correlated with poor survival in TCGA cohort, suggesting a prospective adoption to CAFs biomarkers for HNSCC treatment.

Comprehensive investigations of intercellular communications are essential for understanding interactions and spatial proximity between CAFs and other cell types. Midkine (MDK) belong to a group of heparin-binding growth factors that has been shown to have pleiotropic functions in various biological processes during development and disease (Cui and Lwigale, 2019). It has been reported to overlap with the expression of SCD1 and LRP1 and promote epidermal growth factor receptor (EGFR) signaling by interacting with surface nucleolin (NCL) in hypoxic condition (Cui and Lwigale, 2019; Kinoshita et al., 2020). In addition, overexpression of SDC1 and SDC2 were associated with more aggressive in prostate cancer and MDK-LRP1 will induce the differentiation of immunosuppressive macrophages (Zhang et al., 2021b; Santos et al., 2021). In our study, we first identified those MDK related ligand-receptor pairs as the dominant signaling facilitate to the cellular cross-talk between CAFs and other cell types. We further contextualize this finding in our combined bulk cohort, thus those ligand-receptor analysis of the putative interactions displayed here can be pursued further to better understand the ecosystem cultivated by intercellular communication in the HNSCC tumor microenvironment. Nicotinamide phosphoribosyl transferase (NAMPT) played a crucial role in cancer cell metabolism, often overexpressed in tumor tissues and was an effective target for antitumor treatments (Garten et al., 2015). NAMPT inhibitor was proved to effectively repress cell growth in head and neck squamous cell carcinoma (Cai et al., 2022). In our study, we revealed the extensive enrichment of NAMPT ligand-receptor pair (ITGA5, ITGB1) communication, we also found ITGA5 and ITGB1 were overexpressed in high CafS group, thus providing an explanation for the complex pro tumorigenic mechanism of CAFs.

With the expression profiles of these CAFs genes, we developed an integrative pipeline to construct a predictive model according to the CafS classifier. We first used Lasso algorithm to screen the contents of model container. In total, 9 CAFs related genes and 107 kinds of models were fitted to the training datasets *via* machine learning. Further validation in independent cohorts revealed that the optimal model was random survival forest (RSF). In contrast to the former studies, the advantage of this model with consensus performance on the prognosis of HNSCC is based on a variety of machine learning algorithms and their combinations, which further make this model more convincing to accurate prognosis. TIME-ROC curve suggested that risk score calculated by this model maintained the high precision and high stable performance in all datasets, which indicated great potential for the future clinical application using this risk score. In addition, compared to the conventional tools such as age, gender, stage and HPV status for evaluating clinical outcomes, the risk score signature worked independently of these factors and had significantly superior efficiency in predicting prognosis in training and validation cohorts. We also reviewed previous published HNSCC-related risk models which including different genes' combination (Liu et al., 2020; Li et al., 2021b; He et al., 2021; Wang et al., 2022a; Huang et al., 2022b; Wang et al., 2022b; Chen et al., 2022; Chi et al., 2022; Du et al., 2022; Han et al., 2022; Peng et al., 2022), among these, none of them presented better AUC value performance than our model. Therefore, our risk score signature could be a promising surrogate for evaluating the prognosis of HNSCC in clinical practice.

Combining the Multivariate Cox regression model and Kaplan–Meier curve, the result revealed glycolysis enzyme biomarkers ENO1 and PGAM1 might be important predictors of overall survival in HNSCC. They have been verified to promote cancer cell proliferation and progression (Ishikawa et al., 2014; Qiao et al., 2021). Another study proved CAFs could secrete cytokines and chemokines thus triggering mobilization of glycogen in cancer cells and induce glycolysis reprograming, this CAFs-mediated glycolysis reprograming then results in the invasion and metastasis enhanced in ovarian cancer (Curtis et al., 2019). In our study, we found ENO1 and PGAM1 both were up-regulation in C1 and high-risk score group. In addition, the same trend was observed in the other four glycolysis enzyme markers, and we validate ENO1 and PGAM1 were overexpression in matched tumor part compared to normal side. Hence, we suggested CAFs could dominate the tumor metabolism microenvironment by inducing glycolysis reprograming in head and neck squamous cell carcinoma. To this end, glycolysis inhibitors present a hopeful method to improve CAFs targeted therapeutic strategy.

Although these promising findings were detected in this study, we acknowledge some limitations. For example, we should verify our results using fresh tumor samples, further biological experiment, including cell and molecular assays need to validate the findings of this study. In addition, we conducted a retrospective study, and future validation should be performed in a prospective multicenter cohort.

In conclusion, we constructed a classification system to distinguish the CAFs-related subtype in head and neck squamous cell carcinoma. We observed the potential mechanism of carcinogenesis to CAFs genes and revealed unique possibilities to target glycolysis pathways to enhance CAFs targeted therapy. We developed an unprecedentedly stable and powerful risk score for assessing the prognosis. Our study contributes to the understanding of the CAFs microenvironment complexity in patients with head and neck squamous cell carcinoma and serves as a basis for future in-depth CAFs gene clinical exploration.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.jianguoyun.com/p/DVCMPk0QuK-dChj-4ugEIAA.

Author contributions

QW and GT designed the study, QW analyzed and interpreted the data, QW, FW, and YZ wrote this manuscript. QW, FW, YZ, and GT edited and revised the manuscript. All authors have seen and approved the final version of the manuscript.

Funding

This work was supported by National Natural Science Foundation of China (No. 81870708).

Acknowledgments

We would like to thank the TCGA, GEO, and other databases for the availability of the data.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1111816/full#supplementary-material

SUPPLEMENTARY FIGURE S1
(A–C) The tabular formats of clinical sample information of TCGA, GSE65858 and GSE41613 cohorts.

SUPPLEMENTARY FIGURE S2
(A) The tumor mutation waterfall plot. **(B)** Tumor mutation burden (TMB) in the groups C1 and C2 from TCGA database, $p = 0.22$.

SUPPLEMENTARY FIGURE S3
(A–B) CafS distribution in the smoking and no smoking groups; TCGA ($p = 0.76$) and GSE65858 ($p = 0.84$).

SUPPLEMENTARY FIGURE S4
(A–D) PGAM1 and ENO1 expression in normal nasopharyngeal epithelial cell line (NP69) and HNSCC cell lines (CAL-27 and FaDu), quantitative real-time PCR assay, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$.

References

Ang, K. K., Harris, J., Wheeler, R., Weber, R., Rosenthal, D. I., Nguyen-Tân, P. F., et al. (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* 363 (1), 24–35. doi:10.1056/NEJMoa0912217

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

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

Buchheit, C. L., Weigel, K. J., and Schafer, Z. T. (2014). Cancer cell survival during detachment from the ECM: Multiple barriers to tumour progression. *Nat. Rev. Cancer* 14 (9), 632–641. doi:10.1038/nrc3789

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36 (5), 411–420. doi:10.1038/nbt.4096

Cai, B. H., Bai, Z. Y., Lien, C. F., Yu, S. J., Lu, R. Y., Wu, M. H., et al. (2022). NAMPT inhibitor and P73 activator represses P53 R175H mutated HNSCC cell proliferation in a synergistic manner. *Biomolecules* 12 (3), 438. doi:10.3390/biom12030438

Catenacci, D. V., Junttila, M. R., Karrison, T., Bahary, N., Horiba, M. N., Nattam, S. R., et al. (2015). Randomized phase ib/II study of gemcitabine plus placebo or vismodegib, a hedgehog pathway inhibitor, in patients with metastatic pancreatic cancer. *J. Clin. Oncol.* 33 (36), 4284–4292. doi:10.1200/JCO.2015.62.8719

Chakravarthy, A., Khan, L., Bensler, N. P., Bose, P., and De Carvalho, D. D. (2018). TGF-β-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nat. Commun.* 9 (1), 4692. doi:10.1038/s41467-018-06654-8

Chen, N., He, D., and Cui, J. (2022). A neutrophil extracellular traps signature predicts the clinical outcomes and immunotherapy response in head and neck squamous cell carcinoma. *Front. Mol. Biosci.* 9, 833771. doi:10.3389/fmolb.2022.833771

Chi, H., Jiang, P., Xu, K., Zhao, Y., Song, B., Peng, G., et al. (2022). A novel anoikis-related gene signature predicts prognosis in patients with head and neck squamous cell carcinoma and reveals immune infiltration. *Front. Genet.* 13, 984273. doi:10.3389/fgene.2022.984273

Cui, R., and Lwigale, P. (2019). Expression of the heparin-binding growth factors Midkine and pleiotrophin during ocular development. *Gene Expr. Patterns* 32, 28–37. doi:10.1016/j.gep.2019.02.001

Curtis, M., Kenny, H. A., Ashcroft, B., Mukherjee, A., Johnson, A., Zhang, Y., et al. (2019). Fibroblasts mobilize tumor cell glycogen to promote proliferation and metastasis. *Cell Metab.* 29 (1), 141–155.e9. doi:10.1016/j.cmet.2018.08.007

Custódio, M., Biddle, A., and Tavassoli, M. (2020). Portrait of a CAF: The story of cancer-associated fibroblasts in head and neck cancer. *Oral Oncol.* 110, 104972. doi:10.1016/j.oraloncology.2020.104972

Daley, D., Mani, V. R., Mohan, N., Akkad, N., Ochi, A., Heindel, D. W., et al. (2017). Dectin 1 activation on macrophages by galectin 9 promotes pancreatic carcinoma and peritumoral immune tolerance. *Nat. Med.* 23 (5), 556–567. doi:10.1038/nm.4314

Drifka, C. R., Loeffler, A. G., Mathewson, K., Keikhosravi, A., Eickhoff, J. C., Liu, Y., et al. (2016). Highly aligned stromal collagen is a negative prognostic factor following pancreatic ductal adenocarcinoma resection. *Oncotarget* 7 (46), 76197–76213. doi:10.18632/oncotarget.12772

Du, P., Chai, Y., Zong, S., Yue, J., and Xiao, H. (2022). Identification of a prognostic model based on fatty acid metabolism-related genes of head and neck squamous cell carcinoma. *Front. Genet.* 13, 888764. doi:10.3389/fgene.2022.888764

Fujiwara, T., Yakoub, M. A., Chandler, A., Christ, A. B., Yang, G., Ouerfelli, O., et al. (2021). CSF1/CSF1R signaling inhibitor pexidartinib (PLX3397) reprograms tumor-associated macrophages and stimulates T-cell infiltration in the sarcoma microenvironment. *Mol. Cancer Ther.* 20 (8), 1388–1399. doi:10.1158/1535-7163.MCT-20-0591

Garten, A., Schuster, S., Penke, M., Gorski, T., de Giorgis, T., and Kiess, W. (2015). Physiological and pathophysiological roles of NAMPT and NAD metabolism. *Nat. Rev. Endocrinol.* 11 (9), 535–546. doi:10.1038/nrendo.2015.117

Han, Y., Ding, Z., Chen, B., Liu, Y., and Liu, Y. (2022). A novel inflammatory response-related gene signature improves high-risk survival prediction in patients with head and neck squamous cell carcinoma. *Front. Genet.* 13, 767166. doi:10.3389/fgene.2022.767166

He, D., Liao, S., Xiao, L., Cai, L., You, M., He, L., et al. (2021). Prognostic value of a ferroptosis-related gene signature in patients with head and neck squamous cell carcinoma. *Front. Cell Dev. Biol.* 9, 739011. doi:10.3389/fcell.2021.739011

Huang, C. K., Sun, Y., Lv, L., and Ping, Y. (2022). ENO1 and cancer. *Mol. Ther. Oncolytics* 24, 288–298. doi:10.1016/j.omto.2021.12.026

Huang, J., Huo, H., and Lu, R. (2022). A novel signature of necroptosis-associated genes as a potential prognostic tool for head and neck squamous cell carcinoma. *Front. Genet.* 13, 907985. doi:10.3389/fgene.2022.907985

Huinen, Z. R., Huijbers, E. J. M., van Beijnum, J. R., Nowak-Sliwinska, P., and Griffioen, A. W. (2021). Anti-angiogenic agents - overcoming tumour endothelial cell anergy and improving immunotherapy outcomes. *Nat. Rev. Clin. Oncol.* 18 (8), 527–540. doi:10.1038/s41571-021-00496-y

Ishikawa, M., Inoue, T., Shirai, T., Takamatsu, K., Kunihiro, S., Ishii, H., et al. (2014). Simultaneous expression of cancer stem cell-like properties and cancer-associated fibroblast-like properties in a primary culture of breast cancer cells. *Cancers (Basel)* 6 (3), 1570–1578. doi:10.3390/cancers6031570

Jiang, J., Mei, J., Yi, S., Feng, C., Ma, Y., Liu, Y., et al. (2022). Tumor associated macrophage and microbe: The potential targets of tumor vaccine delivery. *Adv. Drug Deliv. Rev.* 180, 114046. doi:10.1016/j.addr.2021.114046

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., et al. (2021). Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* 12 (1), 1088. doi:10.1038/s41467-021-21246-9

Johnson, D. E., Burtness, B., Leemans, C. R., Lui, V. W. Y., Bauman, J. E., and Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Prim.* 6 (1), 92. doi:10.1038/s41572-020-00224-3

Kalluri, R., and Zeisberg, M. (2006). Fibroblasts in cancer. *Nat. Rev. Cancer* 6 (5), 392–401. doi:10.1038/nrc1877

Kinoshita, D., Shishido, T., Takahashi, T., Yokoyama, M., Sugai, T., Watanabe, K., et al. (2020). Growth factor midkine aggravates pulmonary arterial hypertension via surface nucleolin. *Sci. Rep.* 10 (1), 10345. doi:10.1038/s41598-020-67217-w

Kürten, C. H. L., Kulkarni, A., Cillo, A. R., Santos, P. M., Roble, A. K., Onkar, S., et al. (2021). Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat. Commun.* 12 (1), 7338. doi:10.1038/s41467-021-27619-4

Li, B., Pei, G., Yao, J., Ding, Q., Jia, P., and Zhao, Z. (2021). Cell-type deconvolution analysis identifies cancer-associated myofibroblast component as a poor prognostic factor in multiple cancer types. *Oncogene* 40 (28), 4686–4694. doi:10.1038/s41388-021-01870-x

Li, X., Jiang, E., Zhao, H., Chen, Y., Xu, Y., Feng, C., et al. (2022). Glycometabolic reprogramming-mediated proangiogenic phenotype enhancement of cancer-associated fibroblasts in oral squamous cell carcinoma: Role of PGC-1α/PFKFB3 axis. *Br. J. Cancer* 127 (3), 449–461. doi:10.1038/s41416-022-01818-2

Li, Y., Weng, Y., Pan, Y., Huang, Z., Chen, X., Hong, W., et al. (2021). A novel prognostic signature based on metabolism-related genes to predict survival and guide personalized treatment for head and neck squamous carcinoma. *Front. Oncol.* 11, 685026. doi:10.3389/fonc.2021.685026

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lin, Y., Cai, Q., Chen, Y., Shi, T., Liu, W., Mao, L., et al. (2022). CAFs shape myeloid-derived suppressor cells to promote stemness of intrahepatic cholangiocarcinoma through 5-lipoxygenase. *Hepatology* 75 (1), 28–42. doi:10.1002/hep.32099

Ling, Z., Li, W., Hu, J., Li, Y., Deng, M., Zhang, S., et al. (2022). Targeting CCL2-CCR4 axis suppress cell migration of head and neck squamous cell carcinoma. *Cell Death Dis.* 13 (2), 158. doi:10.1038/s41419-022-04610-5

Liu, X., Chen, J., Lu, W., Zeng, Z., Li, J., Jiang, X., et al. (2020). Systematic profiling of immune risk model to predict survival and immunotherapy response in head and neck squamous cell carcinoma. *Front. Genet.* 11, 576566. doi:10.3389/fgene.2020.576566

Nash, G. F., Walsh, D. C., and Kakkar, A. K. (2001). The role of the coagulation system in tumour angiogenesis. *Lancet Oncol.* 2 (10), 608–613. doi:10.1016/s1470-2045(01)00518-6

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Ngambenjawong, C., Gustafson, H. H., and Pun, S. H. (2017). Progress in tumor-associated macrophage (TAM)-targeted therapeutics. *Adv. Drug Deliv. Rev.* 114, 206–221. doi:10.1016/j.addr.2017.04.010

Nordgren, I. K., and Tavassoli, A. (2011). Targeting tumour angiogenesis with small molecule inhibitors of hypoxia inducible factor. *Chem. Soc. Rev.* 40 (8), 4307–4317. doi:10.1039/c1cs15032d

Ovais, M., Guo, M., and Chen, C. (2019). Tailoring nanomaterials for targeting tumor-associated macrophages. *Adv. Mater* 31 (19), e1808303. doi:10.1002/adma.201808303

Peng, G., Chi, H., Gao, X., Zhang, J., Song, G., Xie, X., et al. (2022). Identification and validation of neurotrophic factor-related genes signature in HNSCC to predict survival and immune landscapes. *Front. Genet.* 13, 1010044. doi:10.3389/fgene.2022.1010044

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171 (7), 1611–1624.e24. doi:10.1016/j.cell.2017.10.044

Qiao, G., Wu, A., Chen, X., Tian, Y., and Lin, X. (2021). Enolase 1, a moonlighting protein, as a potential target for cancer treatment. *Int. J. Biol. Sci.* 17 (14), 3981–3992. doi:10.7150/ijbs.63556

Santos, N. J., Barquilha, C. N., Barbosa, I. C., Macedo, R. T., Lima, F. O., Justulin, L. A., et al. (2021). Syndecan family gene and protein expression and their prognostic values for prostate cancer. *Int. J. Mol. Sci.* 22 (16), 8669. doi:10.3390/ijms22168669

Shelton, M., Anene, C. A., Nsengimana, J., Roberts, W., Newton-Bishop, J., and Boyne, J. R. (2021). The role of CAF derived exosomal microRNAs in the tumour microenvironment of melanoma. *Biochim. Biophys. Acta Rev. Cancer* 1875 (1), 188456. doi:10.1016/j.bbcan.2020.188456

Tan, Q., Wu, X. G., Zhang, M., Meng, L., Zhong, H., Cai, Y., et al. (2019). Performance analysis of PQDCF-coated silicon image sensor using Monte-Carlo ray-trace simulation. *Opt. Express* 27 (6), 9079–9087. doi:10.1364/OE.27.009079

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., et al. (2018). TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* 554 (7693), 538–543. doi:10.1038/nature25492

Thomas, D. A., and Massagué, J. (2005). TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* 8 (5), 369–380. doi:10.1016/j.ccr.2005.10.012

Van Cutsem, E., Tempero, M. A., Sigal, D., Oh, D. Y., Fazio, N., Macarulla, T., et al. (2020). Randomized phase III trial of pegvorhyaluronidase alfa with nab-paclitaxel plus gemcitabine for patients with hyaluronan-high metastatic pancreatic adenocarcinoma. *J. Clin. Oncol.* 38 (27), 3185–3194. doi:10.1200/JCO.20.00590

Wang, Q., Wang, F., Zhao, Y., and Tan, G. (2022). Necroptosis is related to anti-PD-1 treatment response and influences the tumor microenvironment in head and neck squamous cell carcinoma. *Front. Genet.* 13, 862143. doi:10.3389/fgene.2022.862143

Wang, X., Park, J., Susztak, K., Zhang, N. R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* 10 (1), 380. doi:10.1038/s41467-018-08023-x

Wang, X., Zhao, Y., Strohmer, D. F., Yang, W., Xia, Z., and Yu, C. (2022). The prognostic value of MicroRNAs associated with fatty acid metabolism in head and neck squamous cell carcinoma. *Front. Genet.* 13, 983672. doi:10.3389/fgene.2022.983672

Warmoes, M. O., and Locasale, J. W. (2014). Heterogeneity of glycolysis in cancers and therapeutic opportunities. *Biochem. Pharmacol.* 92 (1), 12–21. doi:10.1016/j.bcp.2014.07.019

Wen, S., Hou, Y., Fu, L., Xi, L., Yang, D., Zhao, M., et al. (2019). Cancer-associated fibroblast (CAF)-derived IL32 promotes breast cancer cell invasion and metastasis via integrin β3-p38 MAPK signalling. *Cancer Lett.* 442, 320–332. doi:10.1016/j.canlet.2018.10.015

Xu, G., Zhang, B., Ye, J., Cao, S., Shi, J., Zhao, Y., et al. (2019). Exosomal miRNA-139 in cancer-associated fibroblasts inhibits gastric cancer progression by repressing MMP11 expression. *Int. J. Biol. Sci.* 15 (11), 2320–2329. doi:10.7150/ijbs.33750

Yang, G. J., Tao, F., Zhong, H. J., Yang, C., and Chen, J. (2022). Targeting PGAM1 in cancer: An emerging therapeutic opportunity. *Eur. J. Med. Chem.* 244, 114798. doi:10.1016/j.ejmech.2022.114798

Yang, Y., Andersson, P., Hosaka, K., Zhang, Y., Cao, R., Iwamoto, H., et al. (2016). The PDGF-BB-SOX7 axis-modulated IL-33 in pericytes and stromal cells promotes metastasis through tumour-associated macrophages. *Nat. Commun.* 7, 11385. doi:10.1038/ncomms11385

Yin, Z., Dong, C., Jiang, K., Xu, Z., Li, R., Guo, K., et al. (2019). Heterogeneity of cancer-associated fibroblasts and roles in the progression, prognosis, and therapy of hepatocellular carcinoma. *J. Hematol. Oncol.* 12 (1), 101. doi:10.1186/s13045-019-0782-x

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, N., Ng, A. S., Cai, S., Li, Q., Yang, L., and Kerr, D. (2021). Novel therapeutic strategies: Targeting epithelial-mesenchymal transition in colorectal cancer. *Lancet Oncol.* 22 (8), e358–e368. doi:10.1016/S1470-2045(21)00343-0

Zhang, Y., Zuo, C., Liu, L., Hu, Y., Yang, B., Qiu, S., et al. (2021). Single-cell RNA-sequencing atlas reveals an MDK-dependent immunosuppressive environment in ErbB pathway-mutated gallbladder cancer. *J. Hepatol.* 75 (5), 1128–1141. doi:10.1016/j.jhep.2021.06.023

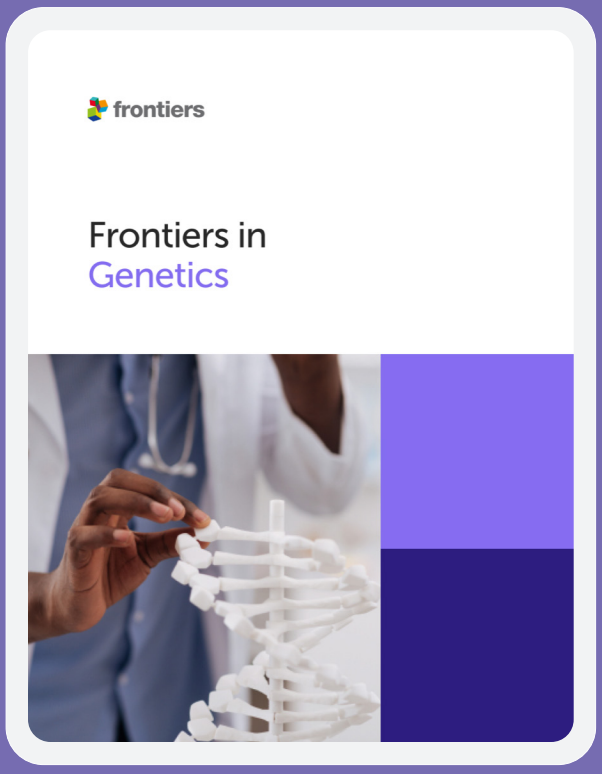
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact