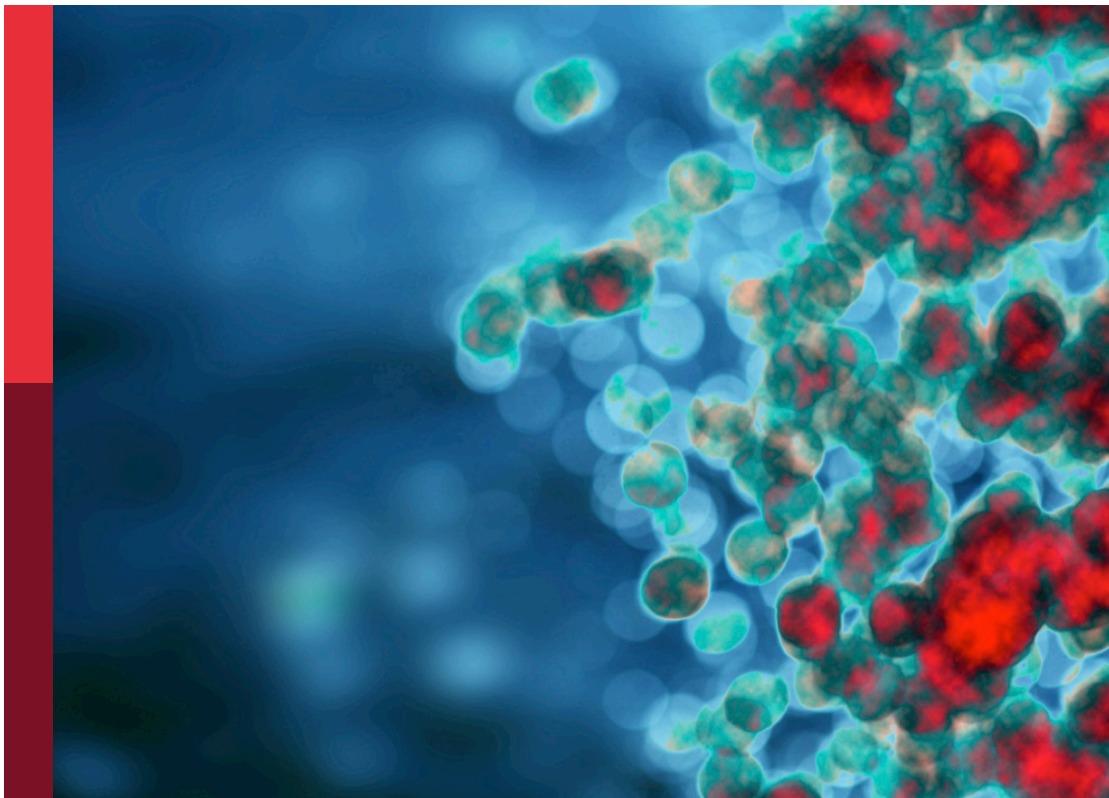
frontiers | Research Topics

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

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

Published in
Frontiers in Immunology
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-3630-8
DOI 10.3389/978-2-8325-3630-8

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

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

Topic editors

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

Citation

Zhang, N., Zhou, M., Zhang, H., Liu, J., Zheng, P., eds. (2023). *Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3630-8

Table of contents

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 07 October 2022
DOI 10.3389/fimmu.2022.988303

OPEN ACCESS

EDITED BY
Meng Zhou,
Wenzhou Medical University, China

REVIEWED BY
Winfried F. Pickl,
Medical University of Vienna, Austria
Bo Li,
Sun Yat-sen Memorial Hospital, Sun Yat-sen University, China
Shangxiang Chen,
Sun Yat-sen University, China

*CORRESPONDENCE
Ming Liu
mingliu35@hrbmu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 07 July 2022
ACCEPTED 20 September 2022
PUBLISHED 07 October 2022

CITATION
Sun Z, Li G, Shang D, Zhang J, Ai L and Liu M (2022) Identification of microsatellite instability and immune-related prognostic biomarkers in colon adenocarcinoma. *Front. Immunol.* 13:988303. doi: 10.3389/fimmu.2022.988303

Identification of microsatellite instability and immune-related prognostic biomarkers in colon adenocarcinoma

Ziquan Sun[1†], Guodong Li[2†], Desi Shang[3], Jinning Zhang[1], Lianjie Ai[1] and Ming Liu[1*]

[1]Colorectal Cancer Surgery Department, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of General Surgery, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China, [3]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Background: Colon adenocarcinoma (COAD) is a prevalent malignancy that causes significant mortality. Microsatellite instability plays a pivotal function in COAD development and immunotherapy resistance. However, the detailed underlying mechanism requires further investigation. Consequently, identifying molecular biomarkers with prognostic significance and revealing the role of MSI in COAD is important for addressing key obstacles in the available treatments.

Methods: CIBERSORT and ESTIMATE analyses were performed to evaluate immune infiltration in COAD samples, followed by correlation analysis for MSI and immune infiltration. Then, differentially expressed genes (DEGs) in MSI and microsatellite stability (MSS) samples were identified and subjected to weighted gene co-expression network analysis (WGCNA). A prognostic model was established with univariate cox regression and LASSO analyses, then evaluated with Kaplan-Meier analysis. The correlation between the prognostic model and immune checkpoint inhibitor (ICI) response was also analyzed.

Results: In total, 701 significant DEGs related to MSI status were identified, and WGCNA revealed two modules associated with the immune score. Then, a seven-gene prognostic model was constructed using LASSO and univariate cox regression analyses to predict survival and ICI response. The high-risk score patients in TCGA and GEO cohorts presented a poor prognosis, as well as a high immune checkpoint expression, so they are more likely to benefit from ICI treatment.

Conclusion: The seven-gene prognostic model constructed could predict the survival of COAD and ICI response and serve as a reference for immunotherapy decisions.

KEYWORDS

colon adenocarcinoma (COAD), microsatellite instability (MSI), immune microenvironment, prognostic biomarkers, drug response

Introduction

Colorectal cancer (CRC) is among the most prevalent cancers globally and is ranked the second most common cause of cancer-related death (1). In developed countries, CRC patients' 5-year survival has been enhanced by early detection, yet 25% present with stage four and additional 25%–50% present in the early stages but progress to metastasis (2). Therefore, further research for effective treatment development is urgently required. Over the past decade, immunotherapy has achieved long-term durable effective responses in treating tumors, including lung cancer and melanoma (2). For CRC, immune checkpoint therapy was approved in 2017 for treating tumors with heavy mutations that have mismatch-repair-deficiency (dMMR) or high levels of microsatellite instability (MSI-H), also known as dMMR-MSI-H tumors. Pembrolizumab obtained FDA approval for treating solid tumors with MSI-Hor dMMR (2).

Colon cancer can be categorized into mismatch-repair-proficient (pMMR), microsatellite stability (MSS), and dMMR microsatellite instability (MSI) subtypes (3). Recent studies have revealed that CRC patients who benefit from immune checkpoint inhibitors mainly have a high mutation burden and mismatch repair deficiency (MSI) (4). In several tumors, the immune cell infiltration biological characteristics and prognostic value have been thoroughly described (3), but the value of MSI as a biomarker remains limited. For example, several clinical trials revealed that metastatic CRC (mCRC) patients with MMR deficiency/MSI-H benefit from the immune checkpoint inhibitor (ICI) treatment (5). However, the efficacy of MSI for drug response and treatment benefit prediction of patients with COAD is unclear. Chen T et al. also developed a lncRNA model to predict gastric cancer's MSI and prognosis (6). Hence, exploring MSI application in COAD therapy and biomarker identification is necessary. This requires identifying accurate predictive biomarkers to comprehend the pathogenesis, predict the clinical outcomes, and subsequently develop a treatment plan for COAD patients.

The tumor microenvironment (TME) and cancer evolution are strongly co-dependent (7, 8). TME comprises several cellular components, such as endothelial cells, fibroblasts, lymph vessels, blood vessels, and immune cells (9). The immune microenvironment has a crucial function in cancer development and therapy, as the immune system components are usually affected by cancers (10–12). Due to the heterogeneity and complexity of tumor immune microenvironment, few patients have benefited from immunotherapy (13), leading to diverse immunotherapy effects among COAD patients (14). The MSI status alone cannot predict the immune checkpoint blockade therapy response because of the complicated interaction between tumor and immune cells (15). Besides, COAD patients' prognosis could be predicted by immune-related parameters (14, 16). Consequently, the immune-related and MSI status for identifying prognosis biomarkers is necessary.

This study used MSI and immune-related gene modules to construct and evaluate a prognostic model. Moreover, the prognostic model and drug sensitivity correlation were analyzed using drug response datasets.

Methods and materials

Colon adenocarcinoma datasets acquisition

The UCSC Xena (https://xenabrowser.net/) was utilized to obtain clinical and gene expression data of samples from COAD patients in the Cancer Genome Atlas (TCGA). MSI or microsatellite stability (MSS) of TCGA COAD samples was obtained from Zaravinos et al. (17). In TCGA cohort, the clinical-pathological stage and microsatellite status were evaluated using the chi-square test and considered statistically significant if the P-value was less than 0.05 (Table 1).

Expression levels were detected using a microarray of two datasets with corresponding clinical information (GSE17536 and

Abbreviations: COAD, Colon adenocarcinoma; MSI, Microsatellite instability; MSS, Microsatellite stability; CRC, Colorectal cancer; dMMR, Mismatch-repair-deficiency; MSI-H, High levels of microsatellite instability; pMMR, Mismatch-repair-proficient; ICI, Immune checkpoint inhibitor; TME, Tumor microenvironment; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; MSI-L, Low levels of microsatellite instability; IOD, Immunohistochemical integral optical density; OS, Overall survival

TABLE 1 Baseline characteristics of patients in TCGA COAD cohort.

Characteristics	Whole Cohort	MSI Group	MSS Group	*P*
TCGA cohort	(n=432)	(n=157)	(n=275)	
Gender				0.0034
Male	230 (53.24%)	69 (43.95%)	161 (58.55%)	
Female	202 (46.76%)	88 (56.05%)	114 (41.45%)	
Age				0.17
<65 years	164 (37.96%)	53 (33.76%)	111 (40.36%)	
>=65 years	268 (62.04%)	104 (66.24%)	164 (59.64%)	
T-stage				1
T1	11 (2.55%)	4 (2.55%)	7 (2.55%)	
T2	73 (16.9%)	26 (16.56%)	47 (17.09%)	
T3	293 (67.82%)	107 (68.15%)	186 (67.64%)	
T4	54 (12.5%)	20 (12.74%)	34 (12.36%)	
N-stage				0.012
N0	250 (57.87%)	105 (66.88%)	145 (52.73%)	
N1	103 (23.84%)	32 (20.38%)	71 (25.82%)	
N2	79 (18.29%)	20 (12.74%)	59 (21.45%)	
M-stage				0.049
M0	314 (72.69%)	118 (75.16%)	196 (71.27%)	
M1	64 (14.81%)	17 (10.83%)	47 (17.09%)	
Stage				0.0079
I	71 (16.44%)	27 (17.2%)	44 (16%)	
II	150 (34.72%)	69 (43.95%)	81 (29.45%)	
III	122 (28.24%)	35 (22.29%)	87 (31.64%)	
IV	64 (14.81%)	17 (10.83%)	47 (17.09%)	

GSE39582), four datasets with corresponding microsatellite stability status (GSE13294, GSE18088, GSE13067, and GSE72969), and gene expression of two datasets (GSE33113 and GSE17537) were obtained using Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo), serving as the validation sets.

Evaluation of the correlation between microsatellite stability status and tumor immune infiltration

Depending on CIBERSORT, the number of each tumor-infiltrating immune cell type was determined (18). CIBERSORT is a tool that estimates specific types of cell abundance based on the gene expression in a mixed cell population, and mRNA expression data were used in this study to compute the range of 22 infiltrating immune cells in TCGA cohort. CIBERSORT score is available on their website (https://cibersort.stanford.edu/index.php) with 1000 permutations. Additionally, the tumor purity score, the stromal cell level, and the level of infiltrated immune cells in TCGA COAD tumor tissues were determined according to ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues) method *via* the "estimate" R package (19).

The expression data of five immune checkpoints were extracted from TCGA cohort; CD274 (code PD-L1), PDCD1 (code PD-1), BTLA, CD47, and CTLA4. A one-sided Wilcoxon rank-sum test was employed for evaluating differences in CIBERSORT 22 immune cells score, ESTIMATE score, and five immune checkpoints expression between the MSI and MSS groups or MSI-H and microsatellite instability low (MSI-L) groups. A P-value less than 0.05 was considered significant.

Differentially expressed genes and functional analysis

TCGA cohort gene expression data were standardized before performing a differential expression analysis using "edgeR" R package for DEGs detection in MSI and MSS samples using a threshold of FDR < 0.05 and |logFC| > 1. In total, 701 DEGs were identified (Table S1), and those DEGs with GO Biological Processes were analyzed using the pathway and process enrichment analysis using Metascape web-based tool (https://metascape.org/gp/index.html) with default settings: terms with

an enrichment factor > 1.5, a minimum count of 3, and P < 0.01. The Metascape data are always up to date.

Weighted gene co-expression network analysis (WGCNA) to identify immune-related modules

WGCNA is a data reduction and unsupervised classification method (20, 21). Subsequently, depending on DEG expression profile, the co-expression network was built using "WGCNA" R package with a parameters set as follows: mergeCutHeight = 0.25, minModuleSize = 20, corType = "Pearson". The module-trait association method was used to determine the co-expression module related to immune infiltration without impact on the clinical characteristics (Table S2). After gene clustering, the modules and phenotype correlation were illustrated by a heatmap. The blue and turquoise modules were eligibly selected.

Construction of a prognostic model

In TCGA cohort, univariate Cox proportional regression was conducted on "blue" as well as "turquoise" module genes linked to OS. Seventy-eight genes with a P value of less than 0.01 were considered for further analysis. In Cox regression model, the considerable prognostic genes were identified by the least absolute shrinkage and selection operator (LASSO) method for variable selection, as well as one standard error (SE) above minimum criteria. The following risk score formula is presented: Risk score = (exp Gene1 * coef Gene1) + (exp Gene2 * coef Gene2) + … +(exp Gene7* coef Gene7), considering the optimized gene expression and the correlation estimated Cox regression coefficients. COAD patients were categorized into two risk groups according to the given risk score median, and their survival time differences were evaluated using a log-rank test. The findings were presented using Kaplan-Meier plots. The risk score differences between MSI and MSS groups or MSI-H and MSI-L groups in TCGA, GSE13294, GSE18088, and GSE13067 cohorts were evaluated using a one-sided Wilcoxon rank-sum and demonstrated statistical significance when the p-value < 0.05.

The risk score and drug response correlation analysis

IMvigor210 was a single-arm phase 2 study to investigate atezolizumab in metastatic urothelial cancer (mUCC) patients (NCT02108652, NCT02951767) (22). The IMvigor210 trial complete expression and clinical data were obtained using "IMvigor210CoreBiologies" R package obtained from http://research-pub.gene.com/IMvigor210CoreBiologies. The risk score difference between the drug response (PD [progressive disease], PR [partial response], SD [stable disease], and CR [complete response]) was assessed. The difference in mutation and neoantigen burdens between the risk groups was evaluated by a one-sided Wilcoxon rank-sum. Differential expression for five immune checkpoints between the two risk groups was evaluated in IMvigor210 and TCGA cohorts. The Genomics of Drug Sensitivity in Cancer (GDSC, http://www.cancerrxgene.org/) was utilized to obtain the drug response measurements as LN_IC50 (natural log of the fitted half-maximal inhibitory concentration) and transcription profiles for about 1000 cancer cell lines and drugs targets/pathways. The drug sensitivity and risk score correlation were calculated using Pearson correlation analysis.

Immunohistochemical verification

Twenty colorectal cancer tissues, including 10 MSI and 10 MSS, were acquired from the Fourth Affiliated Hospital of Harbin Medical University. Immunohistochemistry was performed as previously described (23). Tissues were incubated with anti-CALB2 (ABclonal, dilution 1:100) antibody at 37°C for 1h and with secondary antibodies at room temperature for 30 min. The Olympus BX53 microscope was utilized to capture images, and the immunohistochemical integral optical density (IOD) was analyzed using Image-Pro Plus v6.0. The groups' average optical densities were compared. The groups' average optical densities were compared.

Statistical analysis

Statistical analysis was conducted using GraphPad Prism 8 and R 3.6.3 (https://www.r-project.org/). For comparing the continuous variables in immunohistochemical analysis, the t-test was employed. We applied the Kruskal-Wallis test to compare the continuous variables during the bioinformatic analysis. The subgroups were divided based on the median value. Kaplan-Meier survival analysis was used to generate overall survival curves, and the log-rank test was used to calculate the significance.

Results

Microsatellite stability status affected tumor immune infiltration

The abundance of twenty-two immune cells within TCGA COAD samples was calculated by CIBERSORT to evaluate the immune cell infiltration (Figure 1). Immune profile for the evaluation of the immune cell infiltration was shown in Table S3.

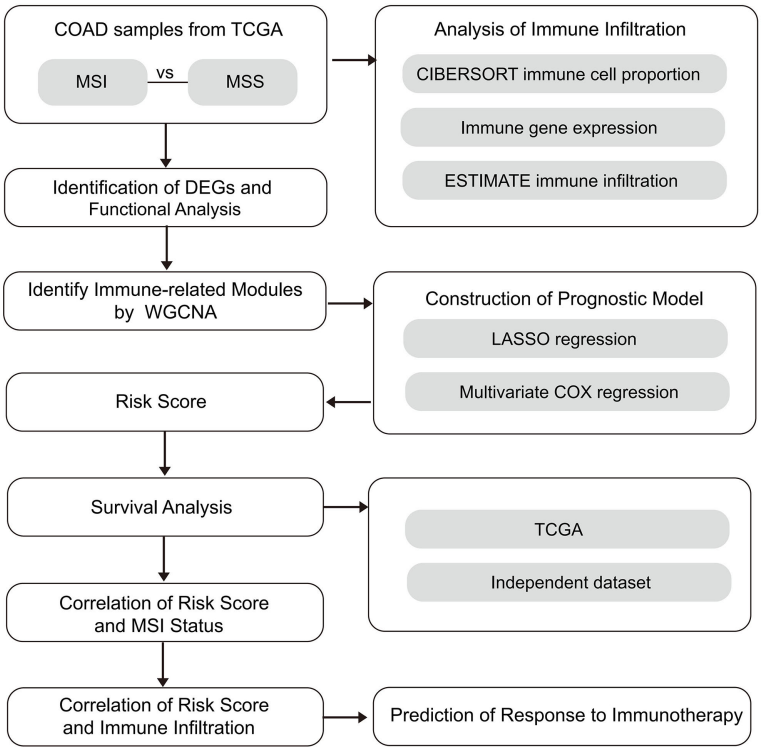


FIGURE 1
The study design schematic diagram.

Then, the proportions of different subpopulations of tumor-infiltrating immune cells were explored in TCGA COAD (Figure 2A). "CD4 memory resting T cells" and "M0 macrophages" represent a significant proportion of COAD immune cell infiltration. Next, we assessed the differentially infiltrated immune cells between MSI and MSS groups (Figure 2B), with the infiltration of "follicular helper T cells" (P = 2.4E-04), "M1 macrophages" (P = 4.5E-04), and "neutrophils" (P = 1.7E-02) in MSI group higher than in MSS group, and infiltration of "CD4 naive T cells" (P = 1.8E-02), "naive B cells" (P = 5.5E-03) and "plasma cells" (P = 1.3E-04) in MSI group lower than in MSS group. Besides, we assessed the immune cell infiltration in MSI-H and MSI-L groups (Figure 2C), showing that the infiltration of "follicular helper T cells" (P = 5.2E-05), "M1 macrophages" (P = 2.3E-08), and "neutrophils" (P = 2.4E-04) in MSI-L group was lower than in MSI-H group, and the infiltration of "CD4 naive T cells" (P = 6.2E-03), "naive B cells" (P = 2.7E-02) and "plasma cells" (P = 9.4E-06) in MSI-H group was lower compared to MSI-L group. Besides, infiltration of CD8 T cells in MSI group was considerably higher than in MSS group (Figure S1A; P = 1.8E-02), and infiltration in MSI-L group was lower than in MSI-H group (Figure S1B; P = 8.8E-02).

To explore the tumor purity distinction between different microsatellite stability statuses in TCGA COAD tumor tissues, the ESTIMATE method was applied to evaluate the level of stromal cells and the immune cell infiltration, and these are the basis for ESTIMATE score. The ESTIMATE score (P = 7.8E-03), Stromal score (P = 2.6E-01) and Immune score (P = 1.7E-05) in MSI group were higher than MSS group (Figure 2D). The ESTIMATE score (P = 3.8E-06), Stromal score (P = 1.8E-03) and Immune score (P = 5.4E-09) in MSI-L group were lower than in MSI-H group (Figure 2E).

The differential expression of five immune checkpoints in the microsatellite instability groups were then analyzed, showing that expressions of BTLA (P = 1.4E-02), PD-L1 (P = 1.2E-08), CD47 (P = 1.5E-02), CTLA-4 (P = 9.7E-07) and PD-1 (P = 4.3E-06) in MSI group were higher than in MSS group (Figure 2F), with BTLA expression (P = 1.2E-03), PD-L1 (P = 3.1E-15), CD47 (P = 5.6E-04), CTLA-4 (P = 1.5E-06) and PD-1 (P = 7.3E-07) in MSI-L group lower than in MSI-H group (Figure 2G).

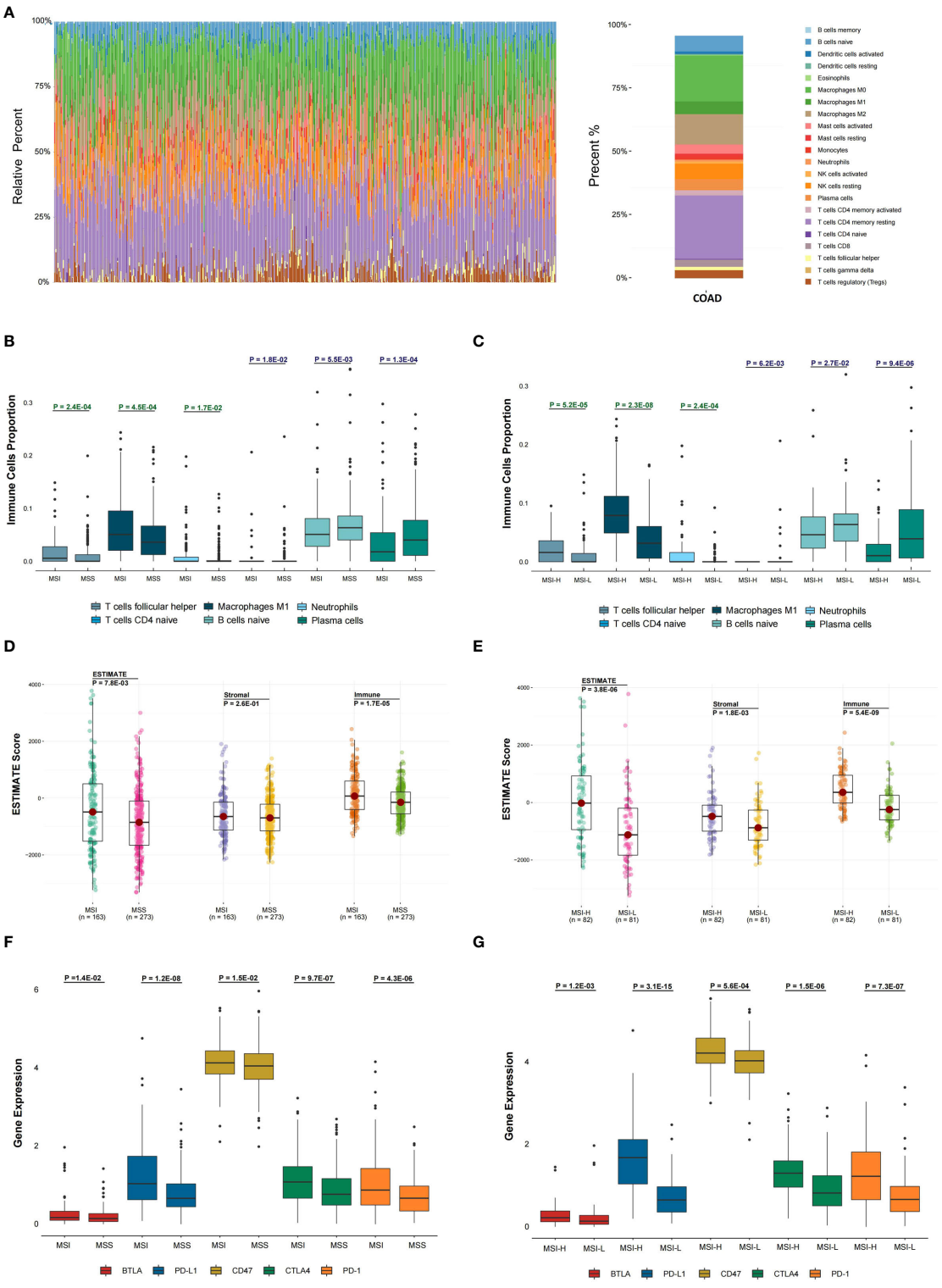


FIGURE 2
Evaluation of the association between microsatellite stability status and tumor immune infiltration. **(A)** In the TCGA COAD cohort, 22 immune cell proportion and distribution using CIBERSORT are shown. **(B)** The six immune cells infiltration difference between MSI and MSS groups. **(C)** The difference of six immune cells infiltration between MSI-L and MSI-H groups and **(D)** The difference in ESTIMATE score between MSI and MSS groups was analyzed. **(E)** The difference in ESTIMATE score between MSI-L and MSI-H groups was analyzed. **(F, G)** The differential expression status of five immune checkpoints between different MSI groups was analyzed. The one-sided Wilcoxon rank-sum test was utilized to compute P-values.

Identifying differentially expressed genes between different microsatellite stability status

Differential expression analysis of the MSI and MSS samples was performed to identify genes that have a pivotal function in the microsatellite stability status, identifying 701 genes (Figures 1, 3A), the top 50 of which are shown in the heatmap (Figure 3B). Pathway and process enrichment analyses using Metascape were used to detect the functional processes regulated by these 701 DEGs. Significantly enriched in GO biological processes were "Cellular component organization or biogenesis", "Negative regulation of biological process", "Developmental process", "Metabolic process", and so on (Figure 3C).

Weighted gene co-expression network construction and immune-related modules identification

Co-expressed networks were built by WGCNA according to 701 DEGs expressions in TCGA COAD cohort to identify the co-expression modules associated with immune traits (Figure 1). The module power value between 1 and 30 was evaluated to assure the average connectivity and high independence. To ensure a scale-free network, the power value was set to 3 when the scale-free R2 reached 0.9 as the soft-thresholding parameter (Figure 4A). The number of genes in each of the six modules identified was as follows: 179 in blue, 273 in turquoise, 122 in brown, 52 in green, 18 in gray, and 57 in the yellow module. The cluster tree is displayed in Figure 4B. The blue module was

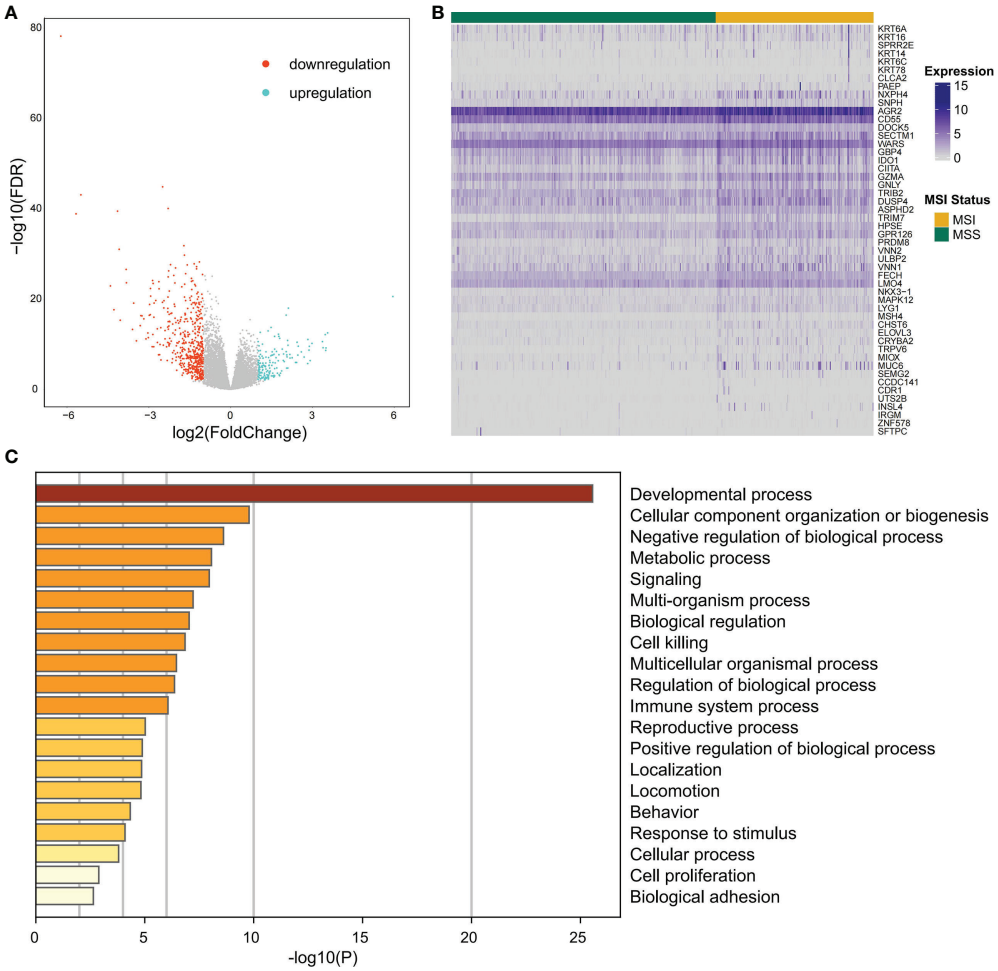


FIGURE 3
The differentially expressed gene identification and functional analysis. **(A)** Volcano plot for differential expression genes (DEGs) between MSI and MSS samples from TCGA COAD cohort. Blue points mean upregulation, and orange points mean downregulation. **(B)** The heatmap shows the expression of the top 50 DEGs in samples from TCGA COAD cohort. **(C)** For 701 significant DEGs, pathway and process enrichment analysis has been done with GO Biological Processes. The graphical graph revealed the top 20 enrichments having $P < 0.01$. A P-value was multi-test adjusted in log 10.

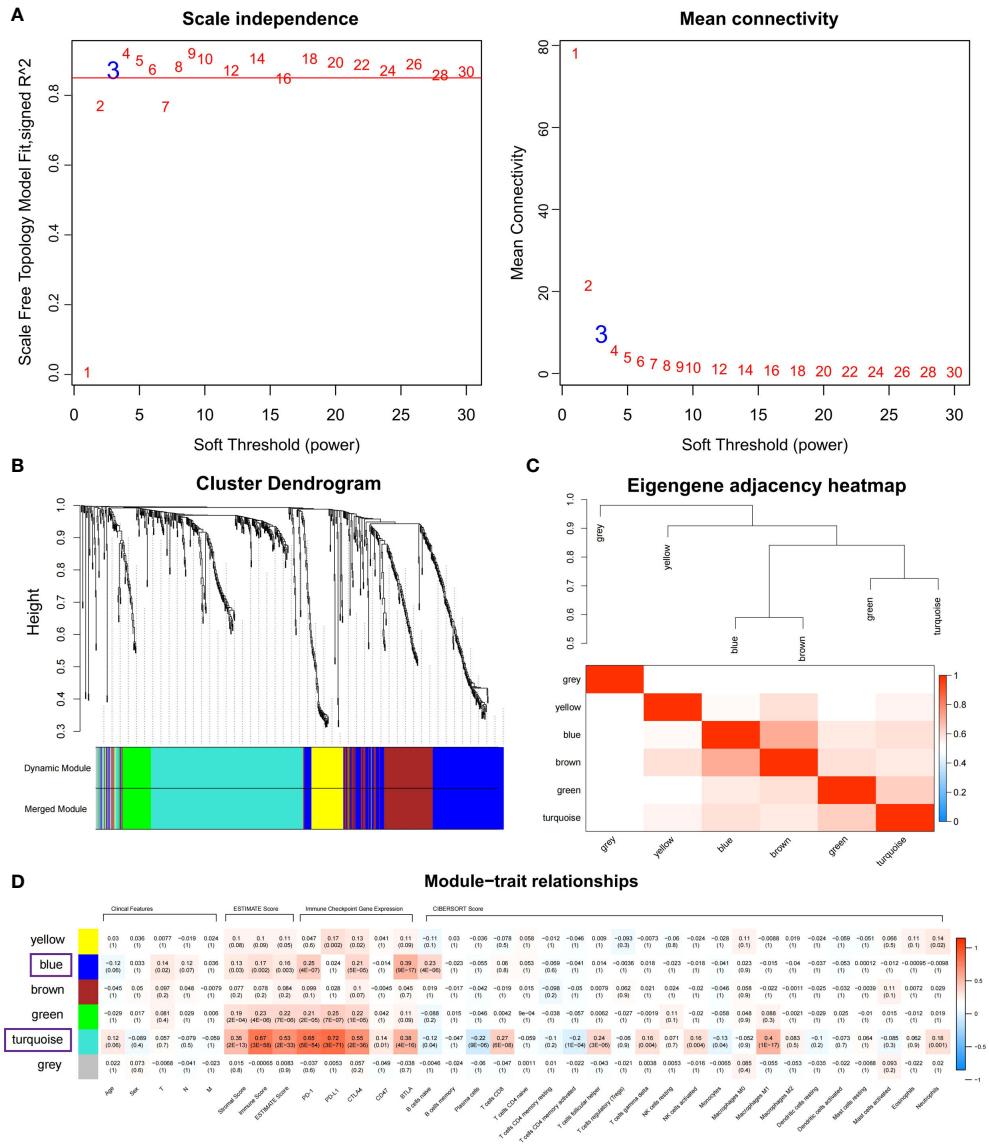


FIGURE 4
Identifying immune-related modules by WGCNA. **(A)** The scale-free fit index analysis and the mean connectivity for various soft-thresholding powers (β). **(B)** Dendrogram for clustering all differentially expressed genes relies on a measure of dissimilarity (1-TOM). **(C)** Clustering correlations among WGCNA modules. The color red represents a positive correlation, and blue represents a negative correlation. **(D)** Heatmap revealing the relationship between modules, clinical features, and immune factors, including ESTIMATE score, the immune checkpoints expression, and CIBERSORT 22 immune cell score. The red refers to a positive correlation, while the blue indicates a negative correlation.

positively linked to brown and turquoise modules, and the turquoise module was positively correlated with green and blue modules (Figure 4C).

The module-trait association method was applied to detect the high co-expression modules relevant to the immune factors but did not affect clinical features. After gene clustering, the correlation between modules and phenotype was illustrated by heatmaps (Figure 4D). According to correlation analysis, blue and turquoise modules were identified as the immune-related modules highly correlated with ESTIMATE score (Figures S2A, B; Cor = 0.6, P = 4.4E-28 for turquoise; Cor = 0.15, P = 4.5E-02 for blue), the expression of PD-1 (Figures S2C, D; Cor = 0.72, P = 6.8E-45 for turquoise; Cor = 0.63, P = 3.5E-21 for blue) and BTLA (Figures S2E, F; Cor = 0.53, P = 3.6E-21 for turquoise; Cor = 0.85, P = 3.8E-51 for blue). Besides, the turquoise modules were highly linked to CD8 T cells (Figure S2G; Cor = 0.58, P = 6.1E-26) and M1 macrophage (Figure S2H; Cor = 0.6, P = 4.4E-28) infiltration.

The co-expression network of blue and turquoise modules is displayed in Figure S3.

The MSI-related prognostic model construction

Univariate Cox proportional regression analysis was conducted to determine the prognostic value of selected MSI-related co-expression module genes, displaying that 78 MSI-related co-expression genes were statistically considerably linked to the overall survival (OS) (Figures 1, 5A; P < 0.01). Next, LASSO analysis was utilized to identify the most effective prognostic genes in addition to one SE over the minimum threshold selected, leading to a model having seven MSI-related co-expression prognostic genes: SMC1B, MAGEA1, LHX8, KHDC1L, HOXC9, GABRG2, and CALB2 (Figures 5B, C). Next, a predictive model was developed according to TCGA training set: risk score = (0.09433 * SMC1B expression) + (0.02362 * MAGEA1 expression) + (0.02937 * LHX8 expression) + (0.1195 * KHDC1L expression) + (0.02567 * HOXC9 expression) + (0.08978 * GABRG2 expression) + (0.01932 * CALB2 expression) (Figure 5D). In TCGA training set, every patient's risk score was determined per the previous formula. The patients were categorized per the median risk score as the cutoff value into two risk groups, with the high-risk group having considerably poorer OS (Figure 5E; P = 2.1E-03; log-rank test).

In the validation set GSE17536, the survival analysis revealed that the high-risk group had a poorer prognosis in OS (Figures 6A; P = 7.4E-03; log-rank test) and disease-specific survival (DFS) (Figure S4; P = 4.2E-02; logrank test), and more patients survived in the low risk group, whereas in the validation set GSE39582, the high-risk group had a poorer prognosis (Figures 6B; P = 5.2E-02; log-rank test), and more patients survived in the low-risk group. Further investigations were performed to confirm if the risk score indicates prognosis for distinct subgroups of clinical characteristics. In TCGA cohort, females, older patients, T3 stage, N1 stage, pathological stage (Stages III and IV), and M subgroups (M0 and M1), the high-risk group patients, presented a considerably poorer OS (Figures 6C–J; P < 0.05; log-rank test). We also found that the risk score of T3+T4 group was higher than that of T1+T2 group (Figure 6K). APC gene had the most mutations in COAD (Figure 7A) and TMB is higher in the high-risk group (Figure 7B, Wilcoxon test, P<0.0001). During the comparison of patient prognosis of low risk and low TMB group, low risk and high TMB group, high risk and low TMB group, and high risk and high TMB group, the patients of the four groups had different outcomes (Figure 7C; P=0.041).

Then, we compared the two risk groups' genetic mutation status. In TCGA cohort, Figures S5A, B revealed the top 20 mutations in the two risk samples. The top five mutations and prognosis correlation were analyzed in the high-risk group, showing that KARS mutation was linked to a poor prognosis (Figure S5C; P = 0.072). However, no difference was observed between the low-risk group and the entire TCGA COAD cohort (Figures S5D, E).

Risk scores were related to immune features and microsatellite stability status

The risk score potential in predicting COAD's immune features was determined by first illustrating the expression status of immune-related genes in the two risk groups of the two data sets (Figures 8A, B). The analysis of the linkage between the immune cell infiltration and expression levels of risk score component genes indicated that HOXC9 and CALB2 are significantly correlated with most immune cell infiltration levels (Figures 8C, D). To confirm the associations between risk score and microsatellite stability status, we analyzed the risk score difference of different microsatellite stability statuses (Figure 1). In TCGA COAD cohort, MSI group had a higher risk score than MSS group (Figure 9A; P = 1.1E-08; one-sided Wilcoxon rank-sum test), and the risk score in MSI-L group was lower than MSI-H group (Figure 9B; P = 1.9E-05; one-sided Wilcoxon rank-sum test). Besides, the risk score in MSI group was higher than MSS group in GSE13294 (Figure 9C; P = 1.8E-05; one-sided Wilcoxon rank-sum test), GSE18088 (Figure 9D; P = 9.0E-03; one-sided Wilcoxon rank-sum test) and GSE13067 (Figure 9E; P = 6.4E-02; one-sided Wilcoxon rank-sum test) cohorts. Next, Pearson correlation analysis was conducted to analyze the correlation between the risk score and expression level of MLH1 and MSH4 (Figure 9F). MutS homologues are the major conductor of the correction of errors introduced in microsatellites. MLH1, MSH3, PMS2, MSH4, MLH3 are five component genes of MutS homologues which can recognize mismatched nucleotides to initiate the repair process (24, 25). Thus, here we analyzed the differential expression status of the five genes in high risk score group and low risk score group. Differences in the five MMR gene expression levels in the high-risk and low-risk score groups are presented in Figure 9G.

The risk score and drug response correlation

In IMvigor210 cohort, the risk score differences among the immunotherapy responsive groups were evaluated to determine if the risk score can predict patients' immunotherapy response (Figure 1). The risk scores in SD and PD were significantly higher than in CR (Figure 10A; P < 0.05), while the risk score in PR was higher compared to CR (Figure 10A; P = 0.062). The immunotherapy responsive group risk score was higher compared to non-response group (Figure 10B; P = 0.063). The high-risk patients with PD or SD responses were less than low-risk patients, and the high-risk group patients with PR or CR

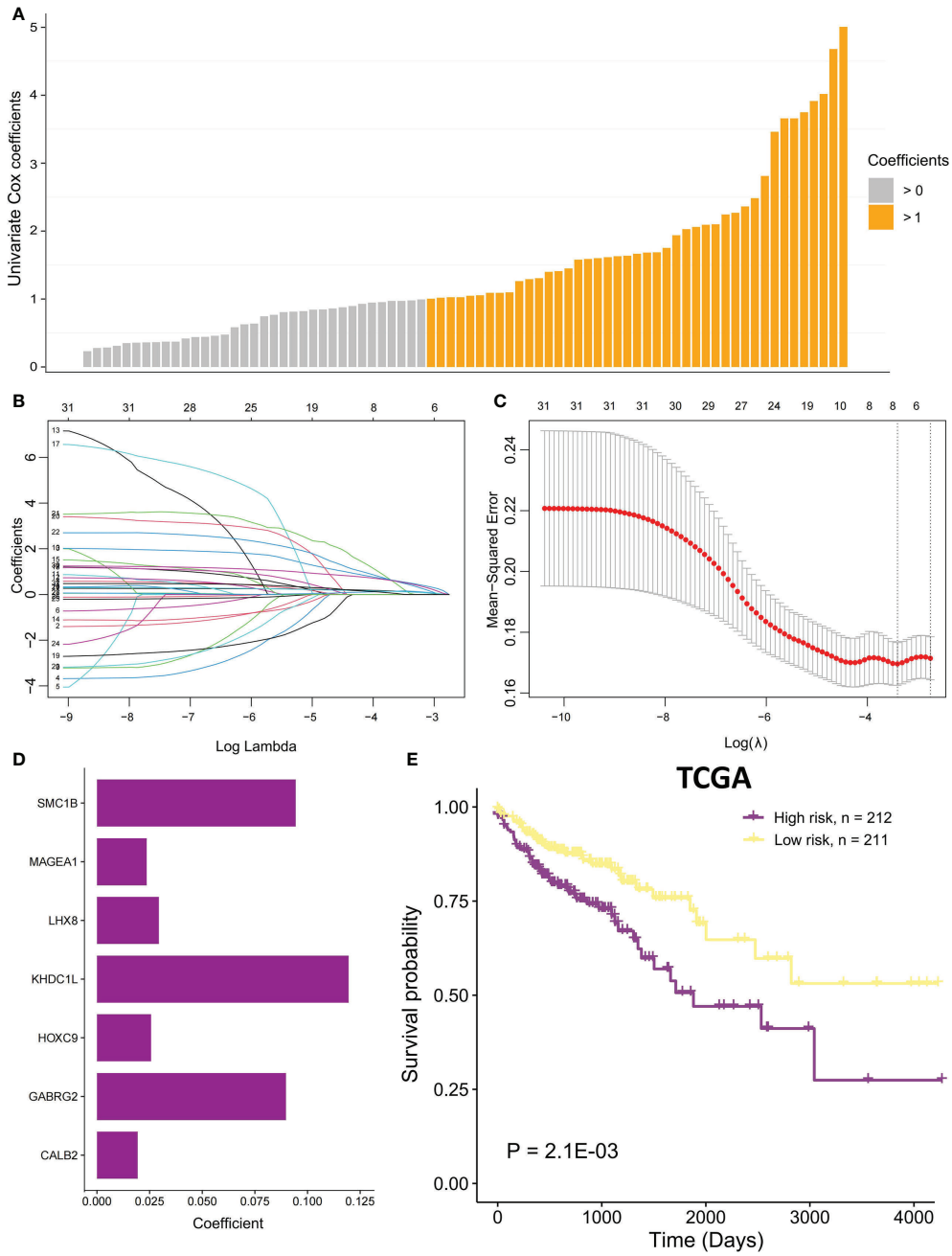


FIGURE 5
The MSI-related prognostic model construction. **(A)** Univariate Cox proportional regression analysis was conducted to identify significant MSI-related co-expression prognostic genes with P < 0.01. The bars mean coefficients of univariate Cox proportional regression analysis. **(B)** LASSO coefficient profiles of 78 MSI-related co-expression prognostic genes. **(C)** Cross-validation for tuning parameter selection in LASSO model. **(D)** The coefficients of seven MSI-related co-expression prognostic genes in the predictive model were caluclated. **(E)** In TCGA COAD cohort, OS difference among the two risk samples was evaluated using a log-rank test. Samples of high risk group: 212. Samples of low risk group: 211.

responses were more than low-risk patients (Figures 10C, F). Besides, the mutation and neoantigen burdens in high-risk patients were higher (Figures 10D, E; P < 0.05). Taken together, such findings indicate that high-risk patients showed better immunotherapy response in IMvigor210 cohort. Then, we investigated the expression of the immune checkpoints among the two risk groups, with the high-risk group in TCGA COAD cohort having considerably higher PD-1, PD-L1, BTLA, and CTLA4 (Figures 10G-J; P < 0.05). In IMvigor210 cohort, the high-risk group had higher PD-1 as well as CD47 (Figure S6).

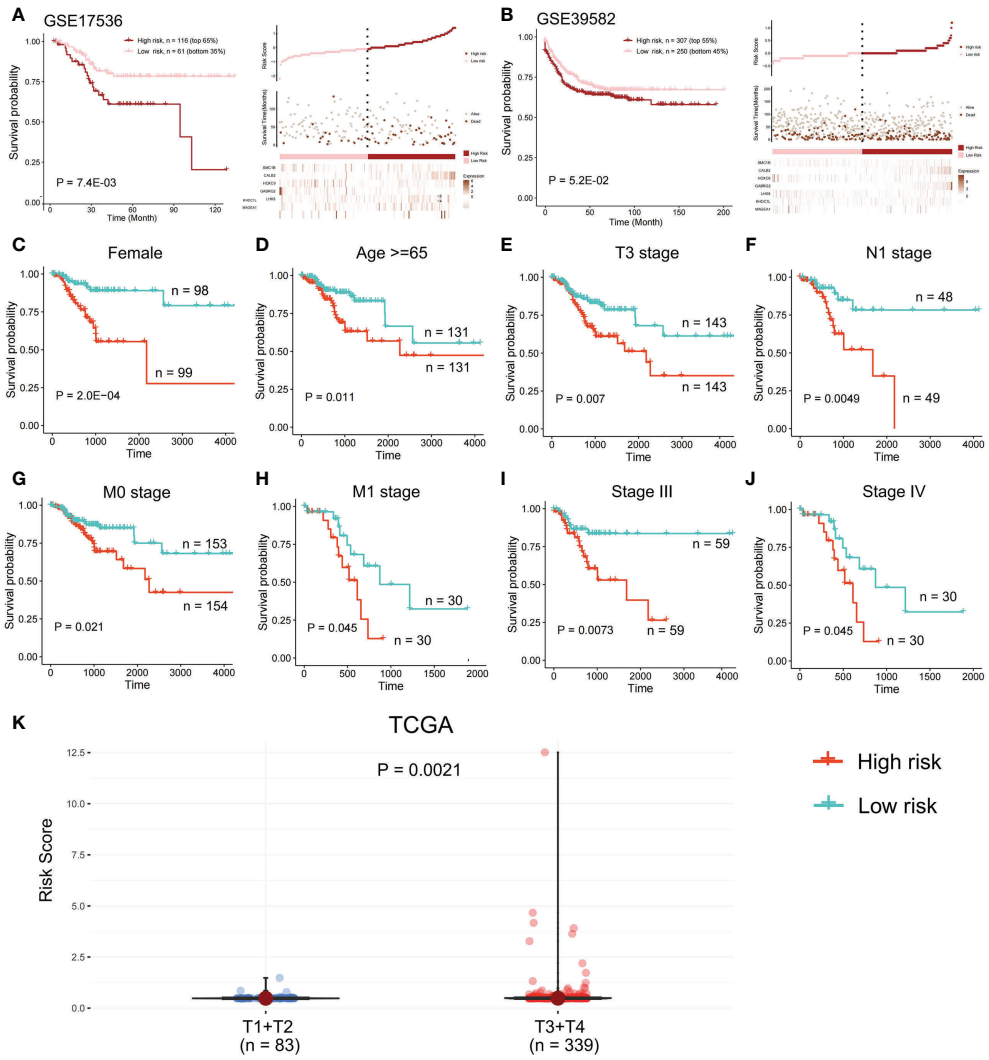


FIGURE 6
Survival analysis in the validation set. OS difference and the survival and risk score distribution among the two risk samples were evaluated using a log-rank test in **(A)** validation set GSE17536 and **(B)** validation set GSE39582. Samples of high risk group in GSE17536: 116. Samples of low risk group in GSE17536: 61. Samples of high risk group in GSE39582: 307. Samples of low risk group in GSE39582: 250. **(C-J)** In TCGA cohort, a log-rank test was employed to evaluate OS difference among two risk samples of females, older patients, T3 stage, N1 stage, M subgroups (M0 and M1), and pathological stage (Stages III and IV). **(K)** Risk score of T1+T2 group and T3+T4 group was compared.

We then examined the linkage between risk score and responsiveness to 20 antitumor agents in GDSC cancer cell lines. Nineteen drugs with a drug response value (LN_IC50) were negatively linked to the risk score, defined as “drug sensitivity”, whereas one drug was positively linked, defined as “drug resistance”, by Pearson correlation analysis (Figure 10K). The drugs with sensitivity were mostly targeting DNA replication and IGF1R signaling pathways (Figure 10L).

Immunohistochemical pathological analysis

To further validate the prognostic value of identified core genes, immunohistochemical pathological analysis was executed to analyze gene CALB2 protein expression status in MSI and MSS subtypes, demonstrating that the gene had a higher expression level in MSI samples (Figures 11A-C).

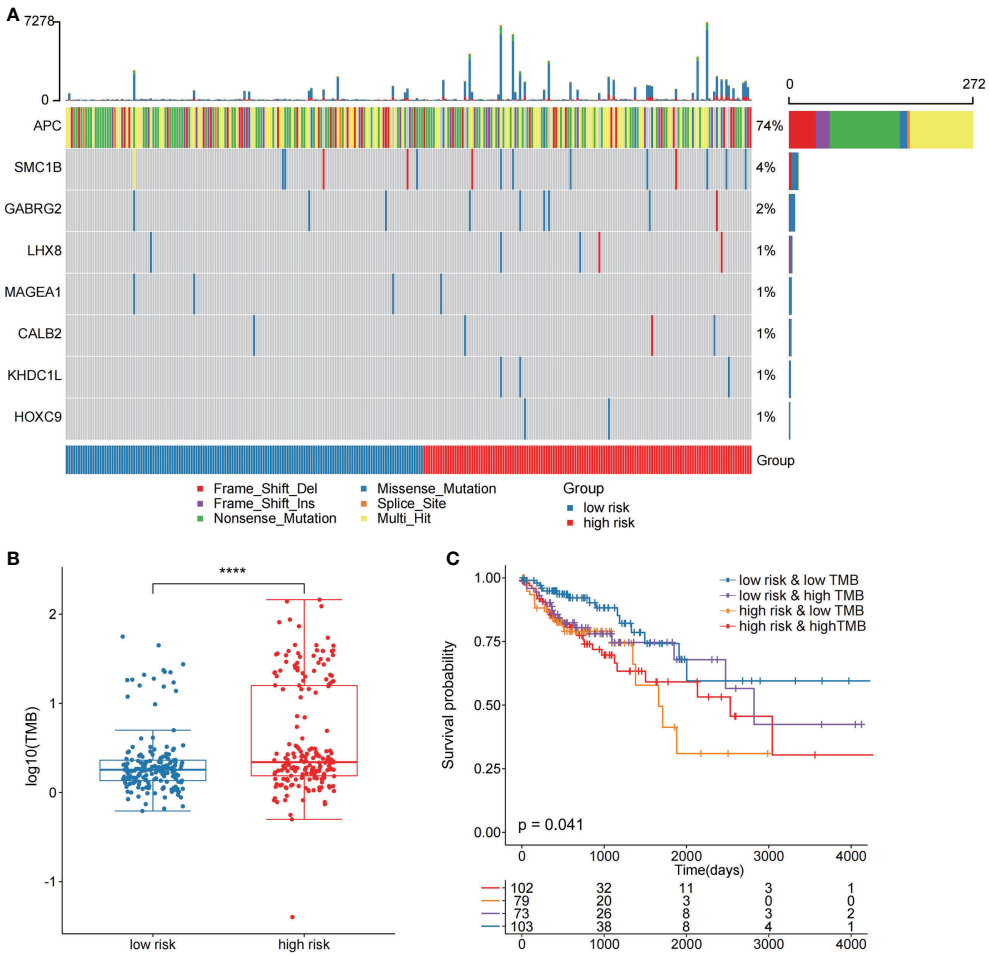


FIGURE 7
The two risk groups have different mutation features. **(A)** Somatic mutation features of the two risk groups. **(B)** TMB was compared between the two risk groups **(C)** Comparison of patient outcome of low-risk and low TMB group, low-risk and high TMB group, high-risk and low TMB group, and high-risk and high TMB group.

Discussion

COAD treatment is challenging because of the advanced stage and poor OS; accordingly, new therapeutic targets are necessary (26). MSI is a high-frequency event in CRC, and recent studies have revealed that MMR deficiency/MSI-H status affects the response to ICI treatment in mCRC patients (27, 28). However, the efficacy of MSI for COAD treatment requires further research. Currently, genes are used to establish a predictive model to evaluate COAD prognosis and responsiveness to therapy, and several gene signatures have been constructed using large-scale publicly available datasets (29, 30). Consequently, the current study established an MSI and immune-related prognostic model comprising seven genes to identify COAD patients who may have better immunotherapy responsiveness. We further validated OS predicting the efficacy of this model in COAD patients *via* a validation analysis for prognostic signatures. The prognostic model can distinguish COAD patients with different responses to ICI treatment.

As an assessment of microsatellite status, we considered the order of "MSS MSI-L MSI-H" as progressive relationships, revealing that the infiltration of "M1 macrophages", "follicular helper T cells", and "neutrophils" in MSS samples was significantly lower compared to MSI samples; a similar pattern was observed in the comparison of MSI-H and MSI-L ($P < 0.05$). M1 macrophages are activated macrophages, defined due to pro-inflammatory cytokine production, mediating pathogens resistance, and exhibiting strong microbicidal characteristics (31). In addition, they are tissue destructive and have anti-tumoral ability (32). Follicular helper T cells are a subpopulation of $CD4^+$ T cells that have a pivotal function in protective immunity because they assist B cells in antibody production versus foreign pathogens (33). Neutrophils have a pivotal function in the host defense

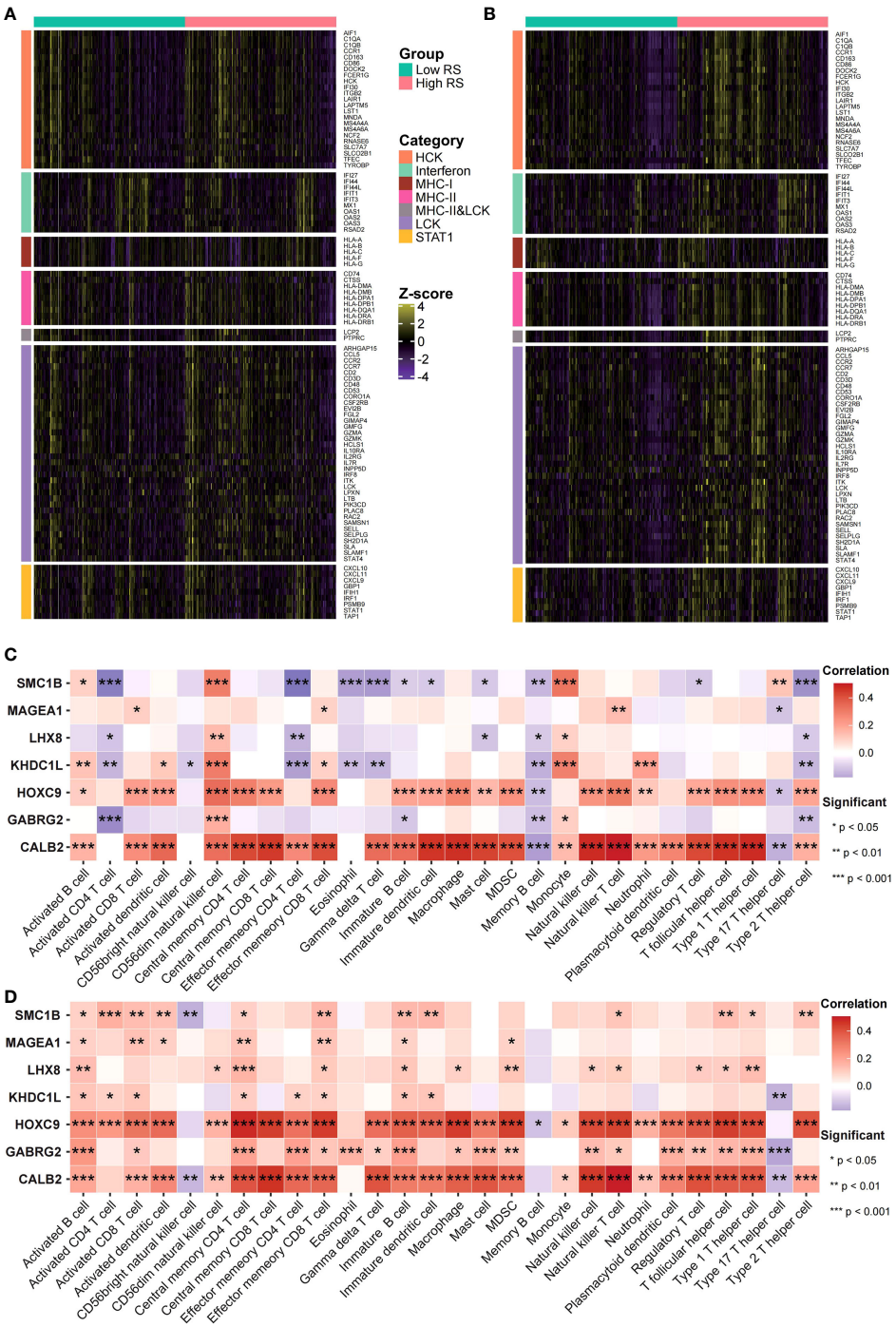


FIGURE 8
Risk scores were related to cancer's immune features. **(A, B)** Expression status of immune-related genes in GEO **(A)** and TCGA **(B)** data sets was analyzed. **(C, D)** The risk score and immune infiltration level correlation in GEO **(C)** as well as TCGA **(D)** data sets were analyzed.

versus infection (34). Meanwhile, high T cell infiltration is linked to a favorable cancer prognosis (8, 35). In addition, the immune score was assessed according to ESTIMATE algorithm, and the five immune checkpoints expression increased as MSI level increased ($P < 0.05$). These findings indicate that MSI is linked to the proportion of immune cell infiltration in COAD; higher microsatellite stability indicates an increase in immune infiltration.

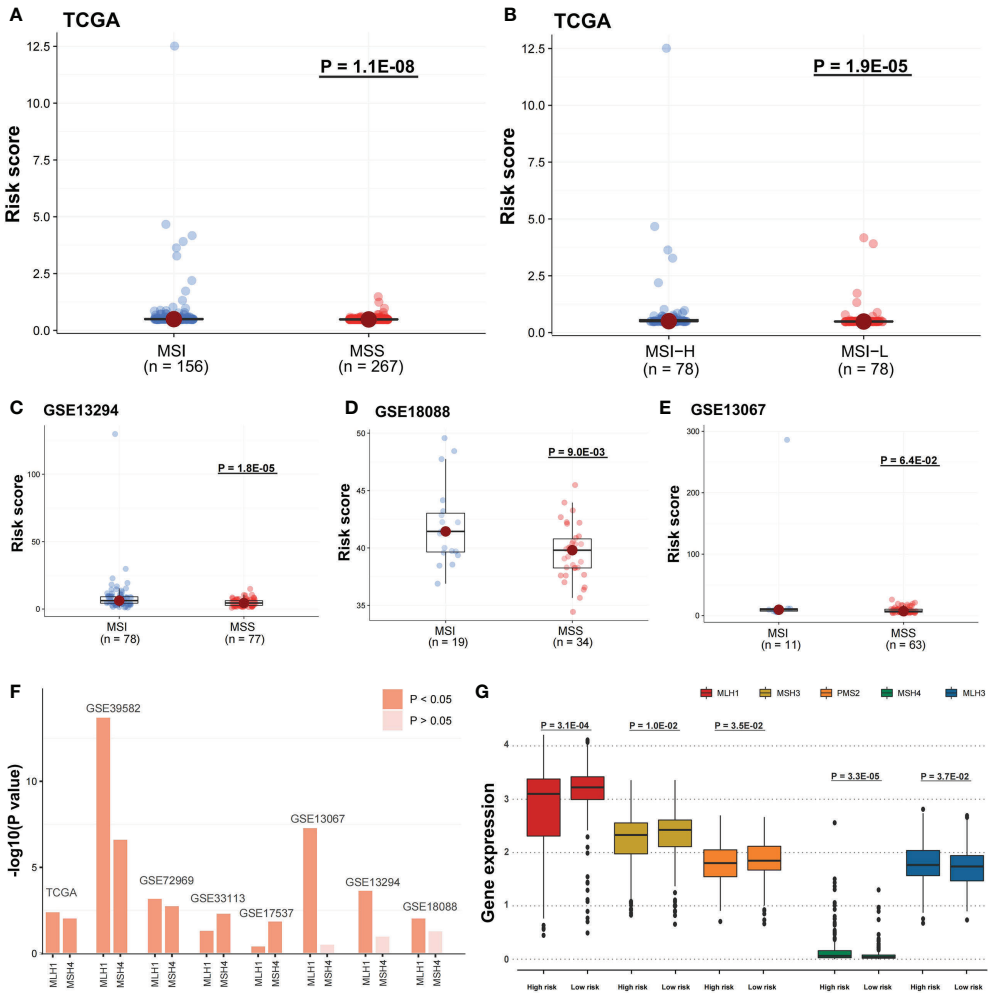


FIGURE 9
Risk scores were related to microsatellite stability status and the MMR gene expression. **(A)** A one-sided Wilcoxon rank-sum test was utilized for evaluating risk score differences in TCGA cohort between MSI and MSS groups and **(B)** MSI-H and MSI-L groups, and **(C-E)** in GSE13294, GSE18088, and GSE13067 cohorts between MSI and MSS groups. **(F)** Pearson correlation analysis was performed to evaluate the correlation between DNA mismatch repair (MMR) gene expression and risk score. The bars mean -log10 (P-value). **(G)** In TCGA cohort, the five MMR gene expression differences among the two risk groups were evaluated by a one-sided Wilcoxon rank-sum test. Statistical significance is determined when P-value < 0.05.

This study identified DEGs related to the microsatellite instability status and functional analysis disclosed that DEGs were significantly enriched in several regulatory pathways. The term "Developmental process" refers to some biological changes linked to growth, information transfer, and differentiation over the organism's life cycle. "Cellular component organization or biogenesis" leads to the constituent parts assembly or a cellular component disassembly. "Negative regulation of biological process" represents any process that reduces, prevents, or stops the biological process rate, frequency, or extent. "Metabolic processes" are chemical reactions and pathways, such as catabolism and anabolism. Our analysis revealed that DEGs related to the microsatellite stability status were closely associated with the growth and activity of cellular components and organisms.

A prognostic model was constructed of seven MSI and immune-related genes in COAD according to WGCNA and other bioinformatics analyses. This approach has demonstrated its effectiveness in cancer research and is commonly utilized (36). The biomarkers identified also have a stable efficacy in COAD prognosis. Apart from T stage, M stage, N stage, Stage and age, gender and sex hormone also contribute to disease prognosis of COAD (37). To further validate our score's robustness, we analyzed the prognosis of high risk patients and low risk patients in subgroups with different clinical pathology features. Results further demonstrated our risk

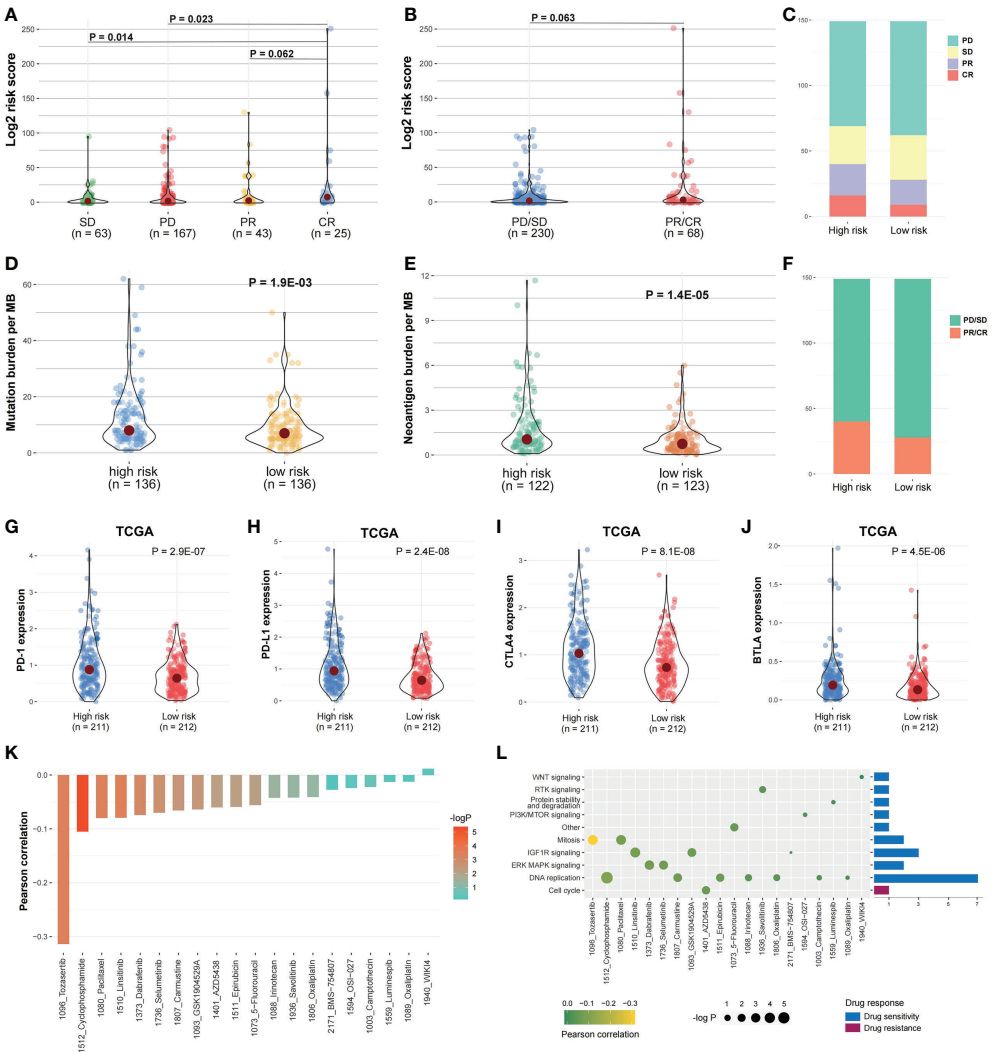


FIGURE 10
Risk score and drug response correlation. **(A, B)** In IMvigor210 cohort, the risk score distribution between the responsive groups and **(C, F)** between the two risk groups. **(D, E)** The mutation and Neoantigen burden distribution in the two risk group patients. **(G–J)** The immune checkpoints expression among the two risk groups. **(K)** The risk score and drug response value correlation using Pearson correlation analysis. Each column refers to a drug. Column brightness represents correlation significance. The column height represents a correlation. **(L)** Signaling pathways targeted by drug resistance to the risk score or sensitivity are presented in red and blue, respectively. Drug names and the signaling pathway targeted by the drug are presented on horizontal and vertical axes, respectively. The number of drugs targeting every signaling pathway is shown on the right of the bar graph. The point size represents the correlation significance.

score's effectiveness. SMC1B associates with cohesin proteins and plays a part in genome stability (38). MAGEA1 codes for an antigen that may cause cancer immune suppression (39). LHX8 is a crucial transcription factor mostly expressed in germ cells (40). HOXC9 controls various cellular processes linked to differentiation *via* activating and repressing the transcription of different gene sets (41). Mutations in GABRG2 have been associated with epilepsy syndromes with varying severities (42). CALB2 is expressed in most poorly differentiated colon carcinomas (43). Although the function of KHDC1L is unclear, we still approve of its effect in the prognostic model. Our prognostic model can differentiate high and low-risk patients, not only in TCGA COAD cohorts (even clinicopathological subgroups) but in the validation sets GSE17536 and GSE39582. Moreover, there were considerable risk scores differences between MSI and MSS samples in TCGA, GSE13294, GSE18088, and GSE13067 datasets, signifying that the prognostic model reflects the microsatellite stability status of patients.

The immunotherapy dataset IMvigor210 was used to validate our prognostic model. Although the cancer type of patients in IMvigor210 is mUCC, IMvigor210 was widely used

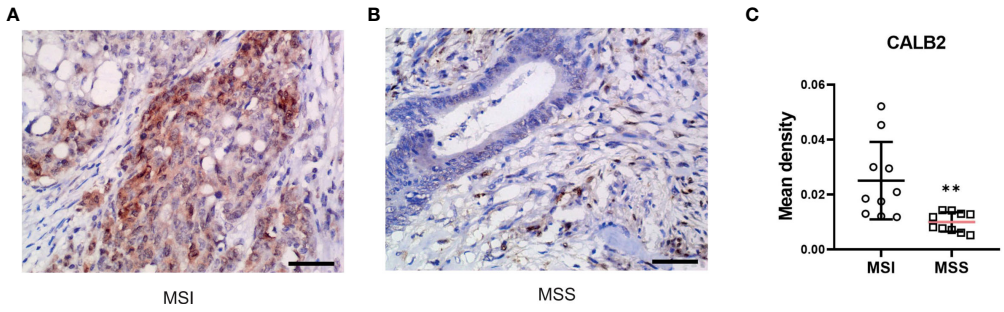


FIGURE 11
CALB2's expression status in clinic acquired tissues. Representative images of IHC-stained MSI COAD tissue **(A)** and MSS COAD tissue **(B)**. **(C)** Comparison of immunohistochemical expression of CALB2 in MSI COAD and MSS COAD tissues. N=20.

as a validation dataset in some studies about other cancers, such as glioblastoma and hepatocellular carcinoma (44, 45). The high-risk patients' score was more likely to profit from ICI (atezolizumab, an anti-PD-L1 antibody) treatment. In addition, we observed a considerable upregulation of immune checkpoints expression in TCGA and IMvigor210, especially PD-L1. The findings revealed that our predictive model might identify groups more susceptible to immunotherapy, and it has potential predictive power for other cancer types (46).

Next, to further validate the potential of our research in clinic applications, we conducted an immunohistochemical pathological analysis. CALB2 (Calbindin 2) encodes an intracellular calcium-binding protein belonging to the troponin C superfamily. This protein plays an important role in message targeting and intracellular calcium buffering and related to cancer progression (47–49). Among DEGs of MSI-H and MSI-L subtypes, CALB2 has the highest fold change. Furthermore, recent research revealed its prognostic value in predicting the outcome and therapy resistance of COAD patients (43, 50). However, the difference in CALB2 expression status in MSI and MSS subtypes remains unclear; consequently, further analysis of clinic-acquired COAD tissues was performed, displaying that CALB2 has a higher expression level in MSI samples which further indicated the therapeutic potential of CALB2.

In conclusion, through a sequence of bioinformatics analyses, a seven-gene predictive model was created to predict COAD patients' outcomes. It could accurately distinguish COAD patients with different prognoses. By categorizing patients and determining a suitable therapy course, our data may help choose the precision medicine in COAD.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The present study was approved by the ethics committee of the Fourth Affiliated Hospital of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

ML, ZS, GL, and DS contribute to conception and design of the study. ZS, GL, and DS wrote or contributed to the writing of the manuscript. ZS, JZ, and LA collected the data. ZS, DS, and JZ conducted the bioinformatic analysis. ML contributed to the material support of the study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 82072673) and the Natural Science Foundation of Heilongjiang Province of China (Grant No. LH2020H066).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.988303/full#supplementary-material

References

1. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol* (2019) 14(2):89–103. doi: 10.5114/pg.2018.81072

2. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* (2019) 16(6):361–75. doi: 10.1038/s41575-019-0126-x

3. Bao X, Zhang H, Wu W, Cheng S, Dai X, Zhu X, et al. Analysis of the molecular nature associated with microsatellite status in colon cancer identifies clinical implications for immunotherapy. *J Immunother Cancer* (2020) 8(2). doi: 10.1136/jitc-2020-001437

4. Lin A, Zhang J, Luo P. Crosstalk between the MSI status and tumor microenvironment in colorectal cancer. *Front Immunol* (2020) 11:2039. doi: 10.3389/fimmu.2020.02039

5. Zhao P, Li L, Jiang X, Li Q. Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. *J Hematol Oncol* (2019) 12(1):54. doi: 10.1186/s13045-019-0738-1

6. Chen T, Zhang C, Liu Y, Zhao Y, Lin D, Hu Y, et al. A gastric cancer LncRNAs model for MSI and survival prediction based on support vector machine. *BMC Genomics* (2019) 20(1):846. doi: 10.1186/s12864-019-6135-x

7. Zhou M, Zhang Z, Zhao H, Bao S, Cheng L, Sun J. An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol Neurobiol* (2018) 55(5):3684–97. doi: 10.1007/s12035-017-0572-9

8. Zhang Z, Chen L, Chen H, Zhao J, Li K, Sun J, et al. Pan-cancer landscape of T-cell exhaustion heterogeneity within the tumor microenvironment revealed a progressive roadmap of hierarchical dysfunction associated with prognosis and therapeutic efficacy. *EBioMedicine* (2022) 83:104207. doi: 10.1016/j.ebiom.2022.104207

9. Angell HK, Bruni D, Barrett JC, Herbst R, Galon J. The immunoscore: Colon cancer and beyond. *Clin Cancer Res* (2020) 26(2):332–9. doi: 10.1158/1078-0432.CCR-18-1851

10. Bao S, Zhao H, Yuan J, Fan D, Zhang Z, Su J, et al. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief Bioinform* (2020) 21(5):1742–55. doi: 10.1093/bib/bbz118

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

12. Bao S, Hu T, Liu J, Su J, Sun J, Ming Y, et al. Genomic instability-derived plasma extracellular vesicle-microRNA signature as a minimally invasive predictor of risk and unfavorable prognosis in breast cancer. *J Nanobiotechnol* (2021) 19(1):22. doi: 10.1186/s12951-020-00767-3

13. Sun J, Yan C, Xu D, Zhang Z, Li K, Li X, et al. Immuno-genomic characterisation of high-grade serous ovarian cancer reveals immune evasion mechanisms and identifies an immunological subtype with a favourable prognosis and improved therapeutic efficacy. *Br J Cancer* (2022) 126(11):1570–80. doi: 10.1038/s41416-021-01692-4

14. Zhang X, Zhao H, Shi X, Jia X, Yang Y. Identification and validation of an immune-related gene signature predictive of overall survival in colon cancer. *Aging (Albany NY)* (2020) 12(24):26095–120. doi: 10.18632/aging.202317

15. Baretti M, Le DT. DNA Mismatch repair in cancer. *Pharmacol Ther* (2018) 189:45–62. doi: 10.1016/j.pharmthera.2018.04.004

16. Zhang Q, Feng Z, Zhang Y, Shi S, Zhang Y, Ren S. Identification and verification of a 17 immune-related gene pair prognostic signature for colon cancer. *BioMed Res Int* (2021) 2021:6057948. doi: 10.1155/2021/6057948

17. Zaravinos A, Roufas C, Nagara M, de Lucas Moreno B, Oblovatskaya M, Efstathiades C, et al. Cytolytic activity correlates with the mutational burden and deregulated expression of immune checkpoints in colorectal cancer. *J Exp Clin Cancer Res* (2019) 38(1):364. doi: 10.1186/s13046-019-1372-z

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

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

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

21. Langfelder P, Horvath S. Fast r functions for robust correlations and hierarchical clustering. *J Stat Softw* (2012) 46(11).

22. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* (2018) 554(7693):544–8. doi: 10.1038/nature25501

23. Wang L, Xu M, Qin J, Lin SC, Lee HJ, Tsai SY, et al. MPC1, a key gene in cancer metabolism, is regulated by COUPTFII in human prostate cancer. *Oncotarget* (2016) 7(12):14673–83. doi: 10.18632/oncotarget.7405

24. Vilar E, Gruber SB. Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol* (2010) 7(3):153–62. doi: 10.1038/nrclinonc.2009.237

25. Clark N, Wu X, Her C. MutS homologues hMSH4 and hMSH5: Genetic variations, functions, and implications in human diseases. *Curr Genomics* (2013) 14(2):81–90. doi: 10.2174/1389202911314020002

26. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct Target Ther* (2020) 5(1):22. doi: 10.1038/s41392-020-0116-z

27. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz HJ, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol* (2017) 18(9):1182–91. doi: 10.1016/S1470-2045(17)30422-9

28. Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-Deficient/Microsatellite instability-high metastatic colorectal cancer. *J Clin Oncol* (2018) 36(8):773–9. doi: 10.1200/JCO.2017.76.9901

29. Chang K, Yuan C, Liu X. A new RBPs-related signature predicts the prognosis of colon adenocarcinoma patients. *Front Oncol* (2021) 11:627504. doi: 10.3389/fonc.2021.627504

30. Condelli V, Calice G, Cassano A, Basso M, Rodriquenz MG, Zupa A, et al. Novel epigenetic eight-gene signature predictive of poor prognosis and MSI-like phenotype in human metastatic colorectal carcinomas. *Cancers (Basel)* (2021) 13 (1). doi: 10.3390/cancers13010158

31. Chavez-Galan L, Olleros ML, Vesin D, Garcia I. Much more than M1 and M2 macrophages, there are also CD169(+) and TCR(+) macrophages. *Front Immunol* (2015) 6:263. doi: 10.3389/fimmu.2015.00263

32. Mosser DM. The many faces of macrophage activation. *J Leukoc Biol* (2003) 73(2):209–12. doi: 10.1189/jlb.0602325

33. Krishnaswamy JK, Alsen S, Yrlid U, Eisenbarth SC, Williams A. Determination of T follicular helper cell fate by dendritic cells. *Front Immunol* (2018) 9:2169. doi: 10.3389/fimmu.2018.02169

34. van Rees DJ, Szilagyi K, Kuijpers TW, Matlung HL, van den Berg TK. Immunoreceptors on neutrophils. *Semin Immunol* (2016) 28(2):94–108. doi: 10.1016/j.smim.2016.02.004

35. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* (2017) 541(7637):321–30. doi: 10.1038/nature21349

36. Luo Z, Wang W, Li F, Songyang Z, Feng X, Xin C, et al. Pan-cancer analysis identifies telomerase-associated signatures and cancer subtypes. *Mol Cancer* (2019) 18(1):106. doi: 10.1186/s12943-019-1035-x

37. Hang D, Shen H. Sex hormone and colorectal cancer: The knowns and unknowns. *Cancer Epidemiol Biomarkers Prev* (2021) 30(7):1302–4. doi: 10.1158/1055-9965.EPI-21-0472

38. Mannini L, Cucco F, Quarantotti V, Amato C, Tinti M, Tana L, et al. SMC1B is present in mammalian somatic cells and interacts with mitotic cohesin proteins. *Sci Rep* (2015) 5:18472. doi: 10.1038/srep18472

39. Zhao J, Wang Y, Mu C, Xu Y, Sang J. MAGEA1 interacts with FBXW7 and regulates ubiquitin ligase-mediated turnover of NICD1 in breast and ovarian cancer cells. *Oncogene* (2017) 36(35):5023–34. doi: 10.1038/onc.2017.131

40. Fu L, Koganti PP, Wang J, Wang L, Wang CL, Yao J. Lhx8 interacts with a novel germ cell-specific nuclear factor containing an Nbl1 domain in rainbow trout (Oncorhynchus mykiss). *PloS One* (2017) 12(2):e0170760. doi: 10.1371/journal.pone.0170760

41. Wang X, Choi JH, Ding J, Yang L, Ngoka LC, Lee EJ, et al. HOXC9 directly regulates distinct sets of genes to coordinate diverse cellular processes during neuronal differentiation. *BMC Genomics* (2013) 14:830. doi: 10.1186/1471-2164-14-830

42. Shen D, Chen J, Liu D, Shen M, Wang X, Wu Y, et al. The GABRG2 F343L allele causes spontaneous seizures in a novel transgenic zebrafish model that can be treated with suberanilohydroxamic acid (SAHA). *Ann Transl Med* (2020) 8 (23):1560. doi: 10.21037/atm-20-3745

43. Haner K, Henzi T, Pfefferli M, Kunzli E, Salicio V, Schwaller B. A bipartite butyrate-responsive element in the human calretinin (CALB2) promoter acts as a repressor in colon carcinoma cells but not in mesothelioma cells. *J Cell Biochem* (2010) 109(3):519–31. doi: 10.1002/jcb.22429

44. Gu X, Guan J, Xu J, Zheng Q, Chen C, Yang Q, et al. Model based on five tumour immune microenvironment-related genes for predicting hepatocellular carcinoma immunotherapy outcomes. *J Transl Med* (2021) 19(1):26. doi: 10.1186/s12967-020-02691-4

45. Zhao B, Wang Y, Wang Y, Chen W, Liu PH, Kong Z, et al. Systematic identification, development, and validation of prognostic biomarkers involving the tumor-immune microenvironment for glioblastoma. *J Cell Physiol* (2021) 236 (1):507–22. doi: 10.1002/jcp.29878

46. Sun J, Zhang Z, Bao S, Yan C, Hou P, Wu N, et al. Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *J Immunother Cancer* (2020) 8(1). doi: 10.1136/jitc-2019-000110

47. Bronner F. Extracellular and intracellular regulation of calcium homeostasis. *ScientificWorldJournal* (2001) 1:919–25. doi: 10.1100/tsw.2001.489

48. DeVilliers P, Liu H, Suggs C, Simmons D, Daly B, Zhang S, et al. Calretinin expression in the differential diagnosis of human ameloblastoma and keratocystic odontogenic tumor. *Am J Surg Pathol* (2008) 32(2):256–60. doi: 10.1097/PAS.0b013e3181452176

49. Sharma A, Ramena GT, Elble RC. Advances in intracellular calcium signaling reveal untapped targets for cancer therapy. *Biomedicines* (2021) 9(9). doi: 10.3390/biomedicines9091077

50. Boyer J, Allen WL, McLean EG, Wilson PM, McCulla A, Moore S, et al. Pharmacogenomic identification of novel determinants of response to chemotherapy in colon cancer. *Cancer Res* (2006) 66(5):2765–77. doi: 10.1158/0008-5472.CAN-05-2693

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 19 October 2022
DOI 10.3389/fimmu.2022.986214

OPEN ACCESS

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

REVIEWED BY
Shaocong Mo,
Fudan University, China
Han Shen,
Guangdong Pharmaceutical University, China
Rameez Raja,
Cleveland Clinic, United States

*CORRESPONDENCE
Fang Liang
liangfang924@163.com
Yan-Wei Luo
royalway@csu.edu.cn

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

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

RECEIVED 04 July 2022
ACCEPTED 28 September 2022
PUBLISHED 19 October 2022

CITATION
Liu J-Y, Liu L-P, Li Z, Luo Y-W and Liang F (2022) The role of cuproptosis-related gene in the classification and prognosis of melanoma.
Front. Immunol. 13:986214.
doi: 10.3389/fimmu.2022.986214

The role of cuproptosis-related gene in the classification and prognosis of melanoma

Jin-Ya Liu[1†], Le-Ping Liu[2,3†], Ze Li[4], Yan-Wei Luo[2*] and Fang Liang[4*]

[1]Department of Plastic Surgery, The Third Xiangya Hospital of Central South University, Changsha, China, [2]Department of Blood Transfusion, The Third Xiangya Hospital of Central South University, Changsha, China, [3]Department of Pediatrics, The Third Xiangya Hospital, Central South University, Changsha, China, [4]Department of Hematology and Critical Care Medicine, The Third Xiangya Hospital, Central South University, Changsha, China

Background: Melanoma, as one of the most aggressive and malignant cancers, ranks first in the lethality rate of skin cancers. Cuproptosis has been shown to paly a role in tumorigenesis, However, the role of cuproptosis in melanoma metastasis are not clear. Studying the correlation beteen the molecular subtypes of cuproptosis-related genes (CRGs) and metastasis of melanoma may provide some guidance for the prognosis of melanoma.

Methods: We collected 1085 melanoma samples in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus(GEO) databases, constructed CRGs molecular subtypes and gene subtypes according to clinical characteristics, and investigated the role of CRGs in melanoma metastasis. We randomly divide the samples into train set and validation set according to the ratio of 1:1. A prognostic model was constructed using data from the train set and then validated on the validation set. We performed tumor microenvironment analysis and drug sensitivity analyses for high and low risk groups based on the outcome of the prognostic model risk score. Finally, we established a metastatic model of melanoma.

Results: According to the expression levels of 12 cuproptosis-related genes, we obtained three subtypes of A_1, B_1, and C_1. Among them, C_1 subtype had the best survival outcome. Based on the differentially expressed genes shared by A_1, B_1, and C_1 genotypes, we obtained the results of three gene subtypes of A_2, B_2, and C_2. Among them, the B_2 group had the best survival outcome. Then, we constructed a prognostic model consisting of 6 key variable genes, which could more accurately predict the 1-, 3-, and 5-year overall survival rates of melanoma patients. Besides, 98 drugs were screened out. Finally, we explored the role of cuproptosis-related genes in melanoma metastasis and established a metastasis model using seven key genes.

Conclusions: In conclusion, CRGs play a role in the metastasis and prognosis of melanoma, and also provide new insights into the underlying pathogenesis of melanoma.

KEYWORDS

melanoma, subtype, machine learning, prognostic model, metastasis model, cuproptosis

Introduction

Melanoma is a malignant tumor caused by aberrant melanocyte proliferation. It has a high fatality rate and is prone to metastasis. According to the 2020 global cancer statistics, skin melanoma ranks 19th among the most common cancers in the world (1), with the number of new cases rising to 324,635 and the number of deaths rising to 57,043. Melanoma is one of the malignant tumors with an extremely high metastasis rate. Its metastasis is characterized by local metastasis through lymphatics first, and then systemic metastasis through blood. Local surgery is the main treatment for early melanoma, while palliative remission is the main treatment for aggressive metastatic melanoma due to poor treatment effects (2). Second, as the most heterogeneous tumor, melanoma is prone to misdiagnosis and treatment failure (3). Melanoma can be classified into nine types according to epidemiology, clinical and histologic morphology, and genomic characteristics, namely low-cumulative solar damage (CSD) melanoma, high-CSD melanoma, Desmoplastic melanoma, Spitz melanomas, Acral melanoma, Mucosal melanomas, Melanomas arising in congenital nevi, Melanomas arising in blue nevi, Uveal melanoma (4). Characteristics of precursor lesions of different subtypes play a certain role in the prevention and early treatment of melanoma. Ultraviolet radiation is one of the main risk factors for the formation of melanoma, and sun exposure is also an important criterion for classifying melanoma types (5). However, little research has been done on melanoma subtypes. Due to the high mortality rate of melanoma, subtyping studies are also extremely important for the individualized treatment of patients.

Cuproptosis is a novel form of cell death induced by copper ionophores (6, 7). Under normal circumstances, cells maintain a relatively low level of intracellular copper through homeostatic mechanisms to prevent excessive copper accumulation leading to cellular damage. Copper ions in the body combine with enzymes and play a major role in blood coagulation, hormone maturation, and energy metabolism (8–11). Within tumor tissue, unbalanced copper levels can cause irreversible damage to tumor tissue. It induces various forms of tumor cell death including apoptosis and autophagy through mechanisms such as reactive oxygen species accumulation, proteasome inhibition, and anti-angiogenesis (11, 12). Studies have shown that copper chelate, taken orally with food, has antitumor and antimetastatic benefits in animals and humans (13). Recent studies have identified specific roles of copper in oncogenic signaling pathways and antitumor drug resistance (14).

In recent years, machine learning has been applied more and more deeply in the field of life sciences, and more and more studies have shown that machine learning plays an important role in medical big data and can effectively mine new information (15–18). With the development of microarray and sequencing technology, the gene expression data of various diseases is also increasing, and machine learning has emerged in the processing of gene expression data of various cancers (19). Machine learning can predict the occurrence and prognosis of cancer, as well as unearth new biomarkers of cancer (20, 21). This study aims to use machine learning combined with bioinformatics to classify melanoma based on cuproptosis-related genes (CRGs) and to establish melanoma prognosis and metastasis models.

In this study, we combined the transcriptional information of melanoma samples from seven GEO datasets and TCGA datasets to screen out a total of 12 CRGs. Then the molecular subtypes and gene subtypes of CRGs were constructed according to clinical characteristics and gene expression. Next, we explored the prognostic role of these CGRs between different subtypes, performed functional analysis of differentially expressed genes between different subtypes, and established a prognostic model. In addition, we performed tumor microenvironment analysis and drug sensitivity analysis. Finally, to further understand the role of CRGs in melanoma development, we established metastasis models based on CRGs using 9 different machine learning algorithms. Figure 1 shows the flow chart of this study.

Methods

Patients and datasets

We screened melanoma datasets in two databases, GEO (https://www.ncbi.nlm.nih.gov/geo/.) and TCGA (http://portal.gdc.cancer.gov/). A total of 7 datasets related to prognosis and metastasis were downloaded from the GEO database [datasets

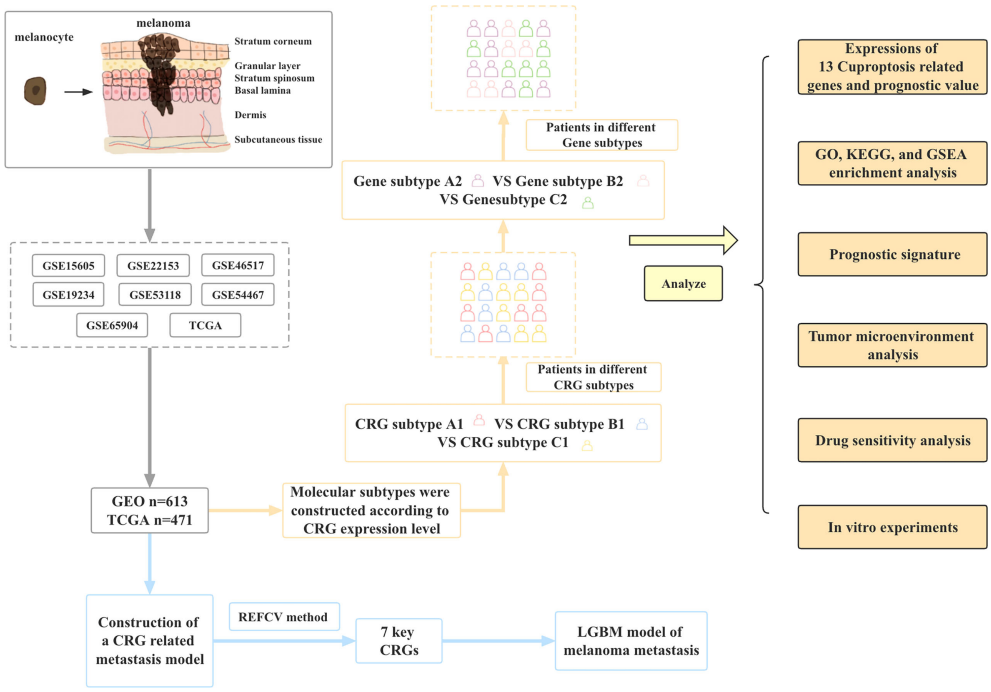


FIGURE 1
Article framework and workflow.

containing prognostic information: GSE19234, GSE22153, GSE54467, GSE69504 (394 melanoma samples)]. Datasets containing metastasis information: GSE15605, GSE22153, GSE46517 (219 samples)). Similarly, we screened melanoma samples in the TCGA database and found 472 samples with prognostic information, of which 471 were melanoma samples. We merged datasets containing prognostic information (GSE19234, GSE22153, GSE54467, GSE69504) with the TCGA dataset. Then, the "perl" language was used to convert the probe matrix into a genes matrix based on the annotation information. Next, we converted the TCGA dataset to TPM format, so that the data form of TCGA was more similar to that of GEO. The "merge" package was used to merge the TCGA dataset with the GEO dataset, and the "sva" package in the R language was used to do a batch correction. Finally, we obtained 862 melanoma samples containing prognostic information and 628 samples containing metastasis information, respectively. In subsequent analyses, we used these combined datasets to build melanoma prognostic models and metastasis models.

Expression of CRGs in melanoma

In the TCGA cohort, "maftools" was used to map the mutation frequencies of CRGs, shown as waterfall plots. Likewise, we analyzed the copy number of CRG in melanoma. The "RCircos" package was used to draw copy number circle diagrams. Next, we constructed a prognostic model based on 12 CRGs (7) in the combined TCGA and GEO cohort. First, we extracted the expression levels of CRGs in datasets with prognostic information and then merged clinical data. The "survival" package was used for survival analysis, cox analysis was used for univariate analysis, and KM analysis was used for survival status analysis.

Construction of molecular subtypes of CRGs

We obtained 13 CRGs (Supplementary Table 1) from previous studies, and after deleting unexpressed CRGs in some samples, we finally selected 12 CRGs for model construction. Consensus Clustering is an unsupervised clustering method and a common research method for cancer subtype classification. It can differentiate samples into different subtypes based on different omics data sets, so as to discover new disease subtypes or perform a comparative analysis of different subtypes. The "ConsensusClusterPlus" R package was used to perform consensus clustering to distinguish different molecular subtypes based on the mRNA expression levels of 12 CRGs. Next, to further analyze the differences between subtypes. We adopted the t-distributed stochastic neighbor embedding (t-

SNE) method to explore the distribution of different subtypes, and the "R t sne" R package was used to estimate the effect of classification. Furthermore, we analyzed the extent of immune cell infiltration between different subtypes. The "heatmap" R package was used to analyze the expression levels of CRGs, tumor grade, gender, and age among different subtypes. Finally, the "GSVA" R package was used to analyze the enriched pathways between different subtypes and displayed as heatmaps.

Survival analysis of gene subtype and differential expression analysis of CRGs

To further understand the correlation between molecular subtypes and differentially expressed genes, we performed gene subtypes. The "limma" R package was used to analyze the differentially expressed genes between different subtypes (logfc > 0.585, p-value < 0.05). After obtaining the differentially expressed genes between each subtype, we took the intersection genes for subsequent analysis. "clusterPrfiler" was used to perform GO enrichment analysis (p-value < 0.05). Similarly, Metascape website (http://metascape.org) (version 2022-04-22) was used to perform enrichment analysis of 71 intergenes.Terms with a P value1.5 are collected and grouped into clusters depended on their membership similarities. The "limma" and "survival" packages were used to analyze the differentially expressed genes associated with prognosis. The Univariate cox regression analyses were used to find intersecting genes associated with prognosis (p-value<0.05). Next, we used the Consensus Clustering method to type the samples according to the expression levels of the intersecting genes. After finding the subtype with the highest internal correlation, survival analysis and clinical trait analysis were performed on different subtypes. We show the above analysis results with KM curve and heat map respectively. Finally, the "limma" package was used to analyze the expression levels of CRGs in different gene types and displayed as boxplots.

Construction of the prognostic model

We divide the samples into training and validation sets in a 1:1 ratio. In the training set, differentially expressed genes associated with prognosis were used to perform Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression analysis through the "glmnet" R package. The risk score was equal to the LASSO regression coefficient for each mRNA multiplied by the sum of the normalized expression levels for each mRNA. Next, we analyzed the AUC of the training set, the validation set, and all samples. Then, based on the samples with survival information, nomogram plots were constructed using the "rms" R package to predict the 1-, 3-, and 5-year survival probabilities of patients. A calibration plot was constructed to assess the agreement of the probabilities predicted by the nomogram with the actual values.

Tumor microenvironment and drug sensitivity analysis

The "CIBERSORT" package was used to perform immune cell infiltration analysis. We analyzed the correlation between 6 key variable genes (AIM2, EDNRB, SLC39A6, TMEM117, PTPRC, and KIF14) and immune cells. At the same time, we also analyzed the correlation between the two prognostic risk groups and the tumor microenvironment. The "estimate" package was used to score the tumor microenvironment in the high-risk and low-risk groups and displayed in a violin plot. Then, we performed a drug sensitivity analysis based on the risk score results. We combined the sample's risk score and drug sensitivity. Then, the high-risk and low-risk groups were analyzed for their sensitivity to the drug, and results with significant differences (p-value > 0.001) were represented by boxplots.

Construction of metastasis model

We integrated all GEO datasets (GSE15605, GSE21153, GSE46517) with melanoma metastasis information. 70% of the samples were set as the training set, and the remaining 30% of the samples were set as the validation set. We used the REFCV method to screen out key metastatic variables by python 3.7. The main idea of recursive feature elimination (REF) is to build the model iteratively and then select the best (or worst) features (selected according to the coefficients). Set the selected features aside and repeat the process on the remaining features until all features are traversed. The order that is eliminated in this process is the ordering of features. REFCV is REF + CV (cross-validation). Its operating mechanism is first to use REF to obtain the ranking of each feature, and then based on the ranking, select [min_features_to_select, len (feature)] feature subsets for model training in turn and cross-validation, and finally select the feature subset with the highest average score. (python 3.7 sklearn 0.22.1 package).

We then use these key variables to build models using 9 different machine learning algorithms (XGBoost', 'Logistic', 'LightGBM', 'RandomForest', 'AdaBoostClassifier', 'GaussianNB', 'ComplementNB', 'SVC' ', 'KNeighbors). Using the cross-validation method, the random seed is set to 1 and the fold is 15. The performance of each model was compared using multi-model forest plots, AUC, accuracy, and F1 values to screen out the best performing models. All Statistical analyses in the process of construction of the metastasis model were performed using python version 3.7 and the Extreme Smart Analysis platform (https://www.xsmartanalysis.com/) (22).

Interpretability of the metastasis model

After filtering out the best performing models, use the "SHAP" package (version 0.39.0, python 3.7) to explain the importance and

contribution of key variables to the model. At the same time, use the force diagram to illustrate 2 samples to show how different variables contribute in different samples ("SHAP" package version 0.39.0, python 3.7). All Statistical analyses in this part were performed using python version 3.7 and Extreme Smart Analysis platform (https://www.xsmartanalysis.com/).

Statistical analysis

The "survival" package was used for survival analysis, cox analysis was used for univariate analysis, and KM analysis was used for survival status analysis. Principal Component Analysis (PCA) was used to demonstrate the differences between CRG subtypes.The "ConsensusClusterPlus" package was used for the subtyping of CRG subtypes and gene subtypes. Lasso regression was used to screen for genes associated with prognosis, and prognostic models were developed using multivariate regression analysis. Wilcoxon rank sum test was used to compare TME scores between the high-risk and low-risk groups. The ROC curve was used to assess the predictive power of the prognostic model. There are several R packages, including "RCircos", "heatmap", and "ggplot" packages for generating graphs. $P < 0.05$ is considered statistically significant.The python software (version 3.7) used in the establishment of the melanoma metastasis model was used for statistical analysis. The REFCV method of the sklearn 0.22.1 package was used to screen key variables in the melanoma metastasis model. In the modeling process of various machine learning algorithms, the xgboost 1.2.1 package was used to perform the XGBoost algorithm, the lightgbm 3.2.1 package was used to perform the LightGBM algorithm, and the sklearn 0.22.1 package was used Used to run other machine learning algorithms. The shap 0.39.0 package was used to demonstrate model interpretability (SHAP graph, feature importance ranking graph, force graph).

Cell lines and constructs for transfection

Human malignant melanoma cell line A375 were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco), supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS, Gibco) at 37°C in a humidified incubator containing 5% CO2. FDX1 siRNAs (1#: 5'-CAUUAACAACCAAAGG AAA-3', 2#: 5'-CAUCUUUGAAGAUCACAUA-3') and control siRNA (5'-UUC UCCGAACGU GUCACGU-3') were obtained from Sangon (Shanghai, China). Transfection of siRNAs was performed with Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher) as recommended.

Western blot analysis

The protein was extracted using RIPA buffer (Beyotime) and the protein concentration was determined using the BCA Protein Assay Kit (Pierced, Grand Island, NY). Protein samples were separated by 12% SDS-PAGE and transferred onto polyvinylidene difluoride membranes ((PVDF, Millipore). To assess the protein expression, the blots were incubated with the primary rabbit antibodies against FDX1 (Abcam) and anti-rabbit secondary antibodies (Cell Signaling Technology) at a dilution of 1:2000 for 1 h at room temperature. β-Actin(Cell Signaling Technology) served as an endogenous control for equal loading.

CCK-8 experiment

The CCK-8 reagent was purchased from GLPBIO (GK10001). Briefly, A375 cells transient transfecting FDX1 siRNA (siFDX1) or the control siRNA (siNC) were seeded at 2x104 cells per well in 96-well plates in quintuplicate, the number of viable cells in each well was measured at 0, 12, 24, and 36 hours according to the manufacturer's instructions.

Wound healing

For wound healing assay, when the cells were grown to 90% confluence after transfection, a straight scratch in the cell monolayer was created by a 10μL pipette tip. A375 cells were incubated with 2% FBS. Images of the scratched area (wound) were taken at the time point of 0h, 24 h, 36 h, and 48 h under a microscope. Wound closure= (original wound area - existing wound area)/original wound area. The area of wound healing was calculated by Fiji (version Fiji for Mac OS X).

Vitro experiment statistical analysis

Statistical analysis was performed using software of Graph Pad Prism 5 (GraphPad, La Jolla, CA). Student's t-tests were used to evaluate significant differences between any two groups of data. All data are represented as means ± SEM. Differences were considered significant if $p < 0.05$.

Results

Article framework and workflow

Flow chart of data collection and data analysis for the article (Figure 1).

Mutation frequency and prognostic value of CRGs in melanoma

Among the 467 patients in the TCGA dataset, 56 patients had CRGs mutations (S1 A). Meanwhile, CRGs chromosome positions are shown as copy number variant plots (S1 B). Besides, the frequency of CRGs copy number variation in the samples is

shown graphically (S1 C), with red representing an increase in copy number and green representing a decrease in mutation. The graphs show a significantly reduced number of mutations in DBT, FDX1, and DLA. Next, we analyzed the association of CRGs with prognosis after combining the TCGA and GEO datasets. 9 of the 13 CRGs were associated with prognosis (S2 A-I). Moreover, Kaplan–Meier analysis results revealed that a higher expression of LIPT1, FDX1, LIAS, and DBT was associated with a better OS ($P < 0.05$), and a lower expression of ATP7B, SLC31A1, PDHA1, DLD, and DLST was associated with a better OS ($P < 0.05$).

Construction of CRGs molecular subtypes of melanoma

To obtain the melanoma subtypes of CRGs, we performed a consensus clustering analysis on the expression level of CRGs on the combined GEO and TCGA datasets. In the cluster analysis of 862 samples, K = 3 was the optimal number of clusters. When K=3, the difference between groups was the smallest, and the difference outside the group was the largest. Therefore, we accurately divided melanoma patients into 3 subtypes, namely A_1, B_1, and C1 (Figure 2A). When dividing melanoma patients into 3 subtypes, the relative change in the area under the CDF curve indicated that the stable distribution of melanoma patients was close (Figure 2B, C). In the Kaplan Meier analysis of A_1, B_1, and C_1 subtypes, the survival outcome of the C_1 subtype was the best, followed by the B_1 subtype, and the worst survival outcome of the A_1 subtype (Figure 2D).

Comparative analysis between three CRGs molecular subtypes

We present the expression level of CRGs and clinical traits, such as Stage, Gender, and Age of the A_1, B_1, and C_1 subtypes in a heat map. CRGs were expressed at the highest level in the B_1 subtype, followed by the C_1 subtype, and lowest in the A_1 subtype. Then, GSVA enrichment pathway analysis was performed on three different subtypes (Figures 2F–H). Comparing the A_1 subtype and the B_1 subtype, it was found that the B_1 subtype was significantly more enriched than the A_1 subtype in cell cycle, non-homologous end linkage, and ubiquitination-mediated hydrolytic protein action. A1 subtype showed significantly higher levels of enrichment in pathways such as neuroactive ligand receptor interactions, cytochrome p450 effects on foreign biometabolism, and drug metabolism of cytochrome p450 than B_1. Comparing the A_1 and C_1 subtypes, the A_1 subtype showed significantly higher levels of enrichment in the drug metabolism cytochrome p450, glycerolipid, and tyramine metabolism pathways than the C_1 subtype. The C_1 subtype was slightly more enriched than the A_1 subtype in pathways such as trap interactions in vesicle transport, ubiquitin-mediated protein hydrolysis, and protein efflux. Comparing the B_1 subtype and C_1 subtype, the enrichment level of the C_1 subtype is higher than that of the B_1 subtype in pathways such as neuroactive ligand receptor interaction, complement system, and leukocyte endothelial migration. The B_1 subtype was significantly more enriched in ubiquitin-mediated protein hydrolysis, aminyl biosynthesis, and citric acid cycle TCA cycle pathways than the C_1 subtype.

Further, we analyzed the level of immune cell infiltration between three CRGs subtypes. Among the 23 immune cells, most of them differed in their degree of infiltration in the A_1, B_1, and C_1 subtypes. Among them, Myeloid-derived suppressor cells (MDSC), Immature B cells, and active B cells had the highest difference in the degree of infiltration, and only Eosinophilna cells had no difference in the degree of infiltration. Overall, the highest level of immune cell infiltration was found in the C_1 subtype and the lowest in the B_1 subtype.

Enrichment analysis of genes with intersections of CRGs subtypes

t-distributed stochastic neighbor embedding(tSNE) analysis showed that the A_1, B_1, and C_1 subtypes are distinguishable from each other. This indicates that our subtype analysis based on CRGs has better typing ability (Figure 3A). Next, we analyzed the differentially expressed genes between A_1, B_1, and C_1 subtypes. There were 1090 differentially expressed genes between A_1 and B1 subtypes, 117 differentially expressed genes between A_1 and C_1 subtypes, and there are 219 Differentially expressed genes between the B_1 and C_1 subtypes. We intersected the differentially expressed genes of the three subtypes and obtained 71 differentially expressed genes that were co-expressed in the three subtypes (Figure 3B). Enrichment analysis in Metascape showed that differentially expressed genes were mainly associated with Signaling by Rho GTPases, Miro GTPases and RHOBTB3, MHC class II antigen presentation, and Platinum drug resistance (Figure 3C). GO (Gene ontology) enrichment analysis indicates the results of intersecting genes in BP (Biological Process), CC (Cellular Component), MF (Molecular Function) respectively (Figure 3D). BP is primarily associated with the establishment of organelle localization, mitotic cell cycle phase transitions, and cytoskeletal-intracellular transport dependence. CC is associated with cell cortex, cell division sites, and membrane microstructure domains. MF is mainly associated with the guanosine triphosphatase binding region, ATP hydrolysis activity, and microtubule binding proteins.

Construction of gene subtypes

To further understand the correlation between CRGs subtypes and differentially expressed genes, we constructed

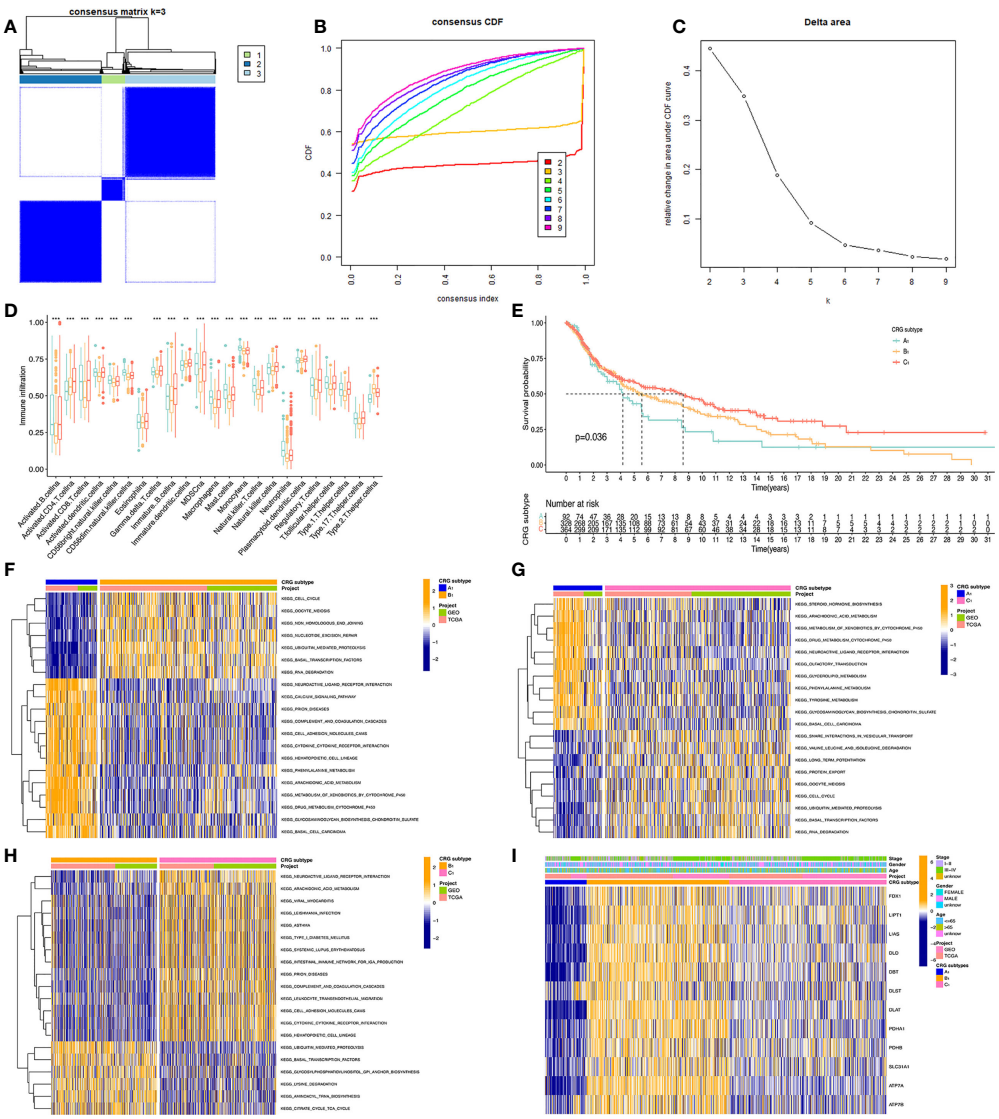


FIGURE 2
Classification of melanoma based on CRGs. **(A)** Molecular subtypes based on CRGs obtained under unsupervised consensus clustering. **(B)** The empirical cumulative distribution function (CDF) plot depicts the consistent distribution of different K values. **(C)** Relative increase in cluster stability by delta area fraction. **(D)** Comparison of the degree of immune cell infiltration of the three molecular subtypes*, P<0.05; **, P<0.01; ***, P<0.001. **(E)** Kaplan Meier analysis results of three molecular subtypes based on 12 CRGs. **(F, G, H)** pictures show the enriched pathways of differentially expressed genes obtained by comparing A_1, B_1, and C_1 molecular subtypes with each other using the GSVA method. **(I)** Heatmap of clinical information and gene expression profiles of the three molecular subtypes based on 12 CRGs.

gene subtypes. We performed univariate regression analysis on 71 differentially expressed genes co-expressed in the three CRGs subtypes, and obtained 16 differentially expressed genes associated with survival. In the cluster analysis, when K=3, we can see that the difference between groups is small, and the difference outside the group is large (S3 A). The comprehensive analysis of the consistent cumulative distribution function (S3 B) and Delt area(S3 C) also shows that K=3 is more suitable. Kaplan Meier analysis was performed on the three gene subtypes, with B_2 having the best survival outcome, A_2 having the second worst survival outcome, and C_2 having the worst survival outcome (S3 D) (P-value<0.001). Then, we illustrate the clinical traits (stage, gender, age)of both gene subtypes and CRGs molecular subtypes in a heat map (S3 E). Besides, we explored the differences in the expression levels of CRGs among the A_2, B_2, and C_2 subtypes. We found that the expression of CRGs was different in A_2, B_2, and C_2 subtypes (p<0.001) (S3 F).

Construction of the prognostic model

A Sankey diagram was used to show our flow chart for two types of melanoma (Figure 4A). AIM2, EDNRB, SLC39A6, TMEM117, PTPRC, and KIF14 were screened out by the LASSO regression algorithm to construct a prognostic model (Figures 4B, C). In the training set, there was a significant difference in prognostic value between the high-risk and low-risk groups (Figure 4D). Survival time was significantly lower in the high-risk group than in the low-risk group. The areas under the time-dependent ROC of the train set are 0.670, 0.662 and 0.683 for 1-, 3-, and 5-year survival. (Figure 4G). Next, the prognostic model was applied to the validation set and to the total sample. In the validation set and in the total sample, the prognostic value of the high-risk group was significantly lower than that of the low-risk group (Figures 4E, F). The areas under the time-dependent ROC of the validation set are 0.587, 0.620, and 0.601 for 1-, 3-, and 5-year survival (Figure 4H). In the total sample, The areas under the time-dependent ROC are 0.626, 0.640 and 0.643 (Figure 4I). The ROC of each group shows that our model has better prediction accuracy. Finally, we used nomograms to predict patient survival (Figure 4J). Calibration curves showed that our model had high accuracy in predicting patient survival at 1, 3, and 5 years (Figure 4K).

Risk curve and tumor microenvironment

We arranged the training set, validation set, and all samples according to the prognostic risk model from low to high risk scores, and obtained the risk curve (Figures 5A–C).Similarly, we obtained the survival status map between risk scores and death samples (Figures 5D–F), and finally, we used heatmaps to show the expression of the model's key variable genes (AIM2, EDNRB, KIF14, PTPRC, SLC39A6, and TMEM117) in the training set,

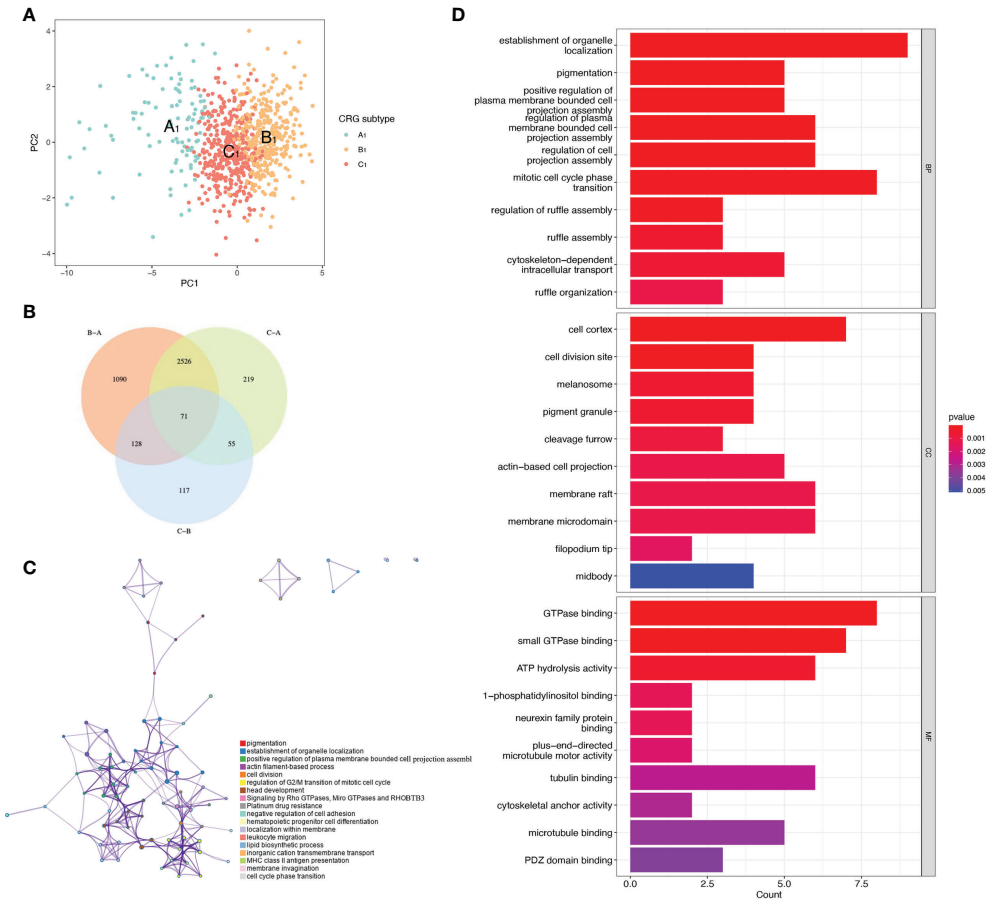


FIGURE 3
Differentially expressed genes of three CRGs molecular subtypes. **(A)** VENN plot showing 71 intersecting differentially expressed genes across three molecular subtypes. **(B)** t-distributed Stochastic Neighbor Embedding (tSNE) analysis of three CRGs molecular subtypes. **(C)** Metascape enrichment analysis of DEGs with intersections of the three molecular subtypes. **(D)** GO enrichment analysis of DEGs with intersections of the three molecular subtypes.

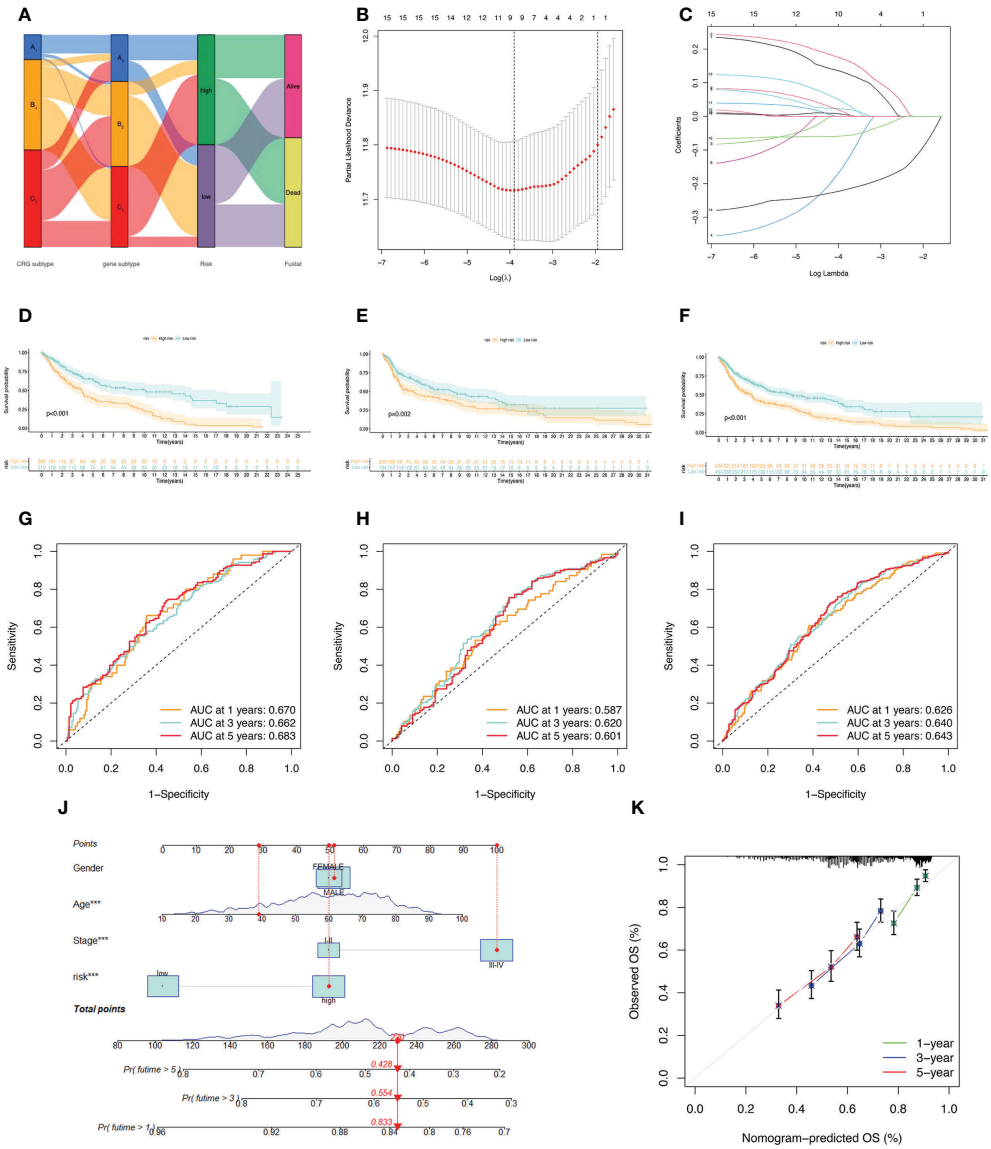


FIGURE 4
Construction of the prognostic model. **(A)** Sankey diagram to describe the process of constructing a prognostic model based on CRGs-subtypes and gene subtypes. **(B, C)** Prognostic genes were screened using LASSO regression. **(D, G)** Kaplan Meier analysis of OS in melanoma patients in the training set; ROC curves for 6 key variable genes. **(E, H)** OS of melanoma patients in Kaplan Meier analysis validation set; ROC curves of 6 key variable genes. **(F, I)** Kaplan Meier analysis of OS in all melanoma patients; ROC curves of 6 key variable genes. **(J)** Nomograms predicting 1-, 3-, and 5-year OS probabilities in melanoma patients. **(K)** Calibration plots of the nomograms.

validation set and all sample (Figures 5G–I). Next, we performed tumor microenvironment analysis on 6 key variable genes (Figure 5J).

The key variable genes were mainly associated with the degree of infiltration of M1 macrophage, M0 macrophage, and memory B cells. KIF14, SLC39A6, TMEM117, and EDNRB, as high-risk genes, were negatively correlated with the degree of infiltration of memory B cells and regulatory T cells, and positively correlated with the degree of infiltration of M1 macrophage, T follicular helper. AIM2 and PTPRC, as low-risk genes, showed a significant positive correlation with the degree of infiltration of memory B cells, activated memory CD4(+) T cells, and CD8(+) T cells, and a significant negative correlation with the degree of infiltration of M0 macrophage. The stromalscore, immunescore, and ESTIMATEscore scores in the high-risk group were significantly lower than those in the low-risk group (Figure 5K).

Comparison of drug sensitivity, subtypes and expression levels of CRGs between high and low risk groups

We divided the high-risk group and the low-risk group according to the model. Screening of sensitive drugs was carried out according to the difference in IC50 concentration between the two groups. A total of 98 drugs (S2) were screened, and we selected 3 high-sensitivity drugs in the low-risk group (Figures 6A–C) and 3 high-sensitivity drugs in the high-risk group (Figures 6D–F). Among the CRG subtypes, B_1 has the lowest risk score and subtype C_1 has the highest risk score (Figure 6G). Among the genesubtypes, C_2 had the highest risk score and B_2 had the lowest risk score (Figure 6H). Among the CRGs genes with differential expression in the high-risk and low-risk groups, only FDX1 expression was decreased in the high-risk group (Figure 6I).

Construction of metastasis model

We used the REFCV method to filter out key metastatic variables: 'FDX1', 'LIPT1', 'LIAS', 'DLD', 'DBT', 'DLAT',

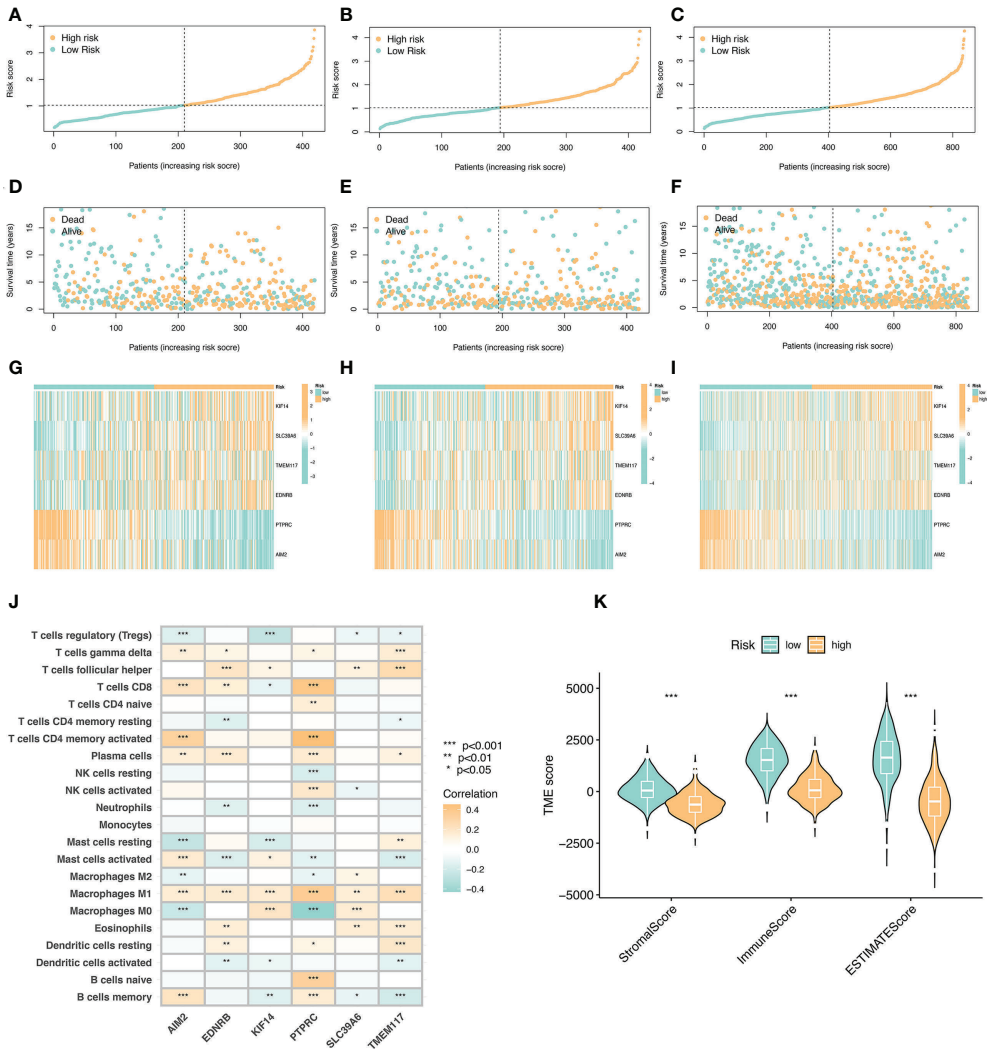


FIGURE 5
Risk curve and immune microenvironment analysis between high and low immune groups. **(A)** Risk curve in the training set. Take the median of the risk scores and use the median to divide the samples into high-risk and low-risk groups. **(B)** Risk curve in the validation set. **(C)** Risk curves of all samples. **(D)** Survival state diagram of the training set, red for dead and blue for survival. **(E)** The living state diagram of the validation set. **(F)** Survival state diagram of all samples. **(G–I)** Heat map showing the expression of 6 key variable genes in training set, validation set and all samples. **(J)** Correlation of 6 key variable genes with immune cells, red represents positive correlation and blue represents negative correlation. **(K)** Correlation of stroma score, immune score, and ESTIMATE with immune microenvironment.

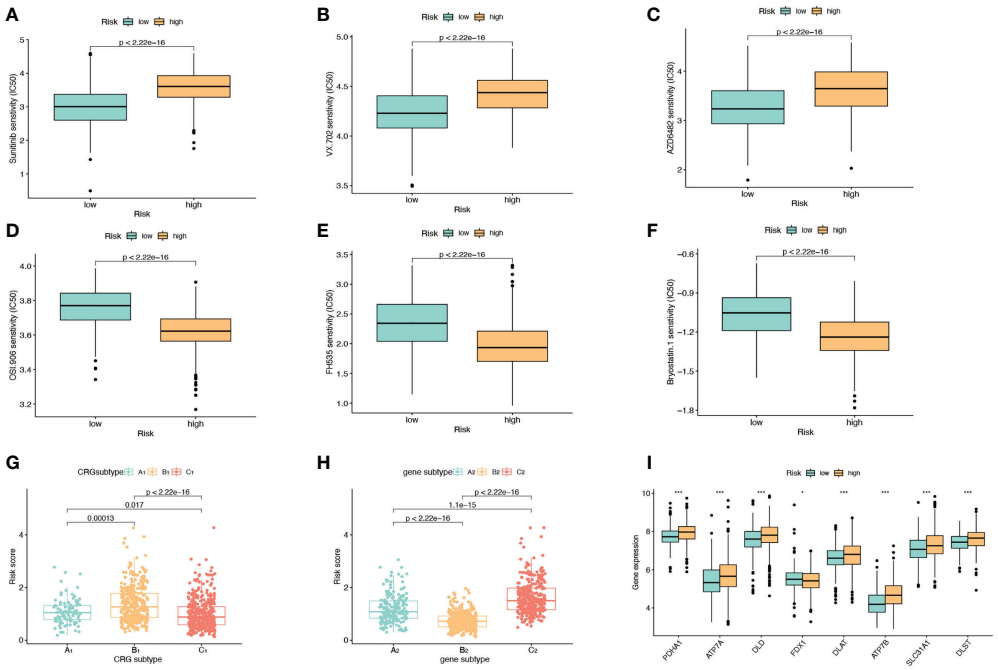


FIGURE 6
Drug Sensitivity Analysis. **(A–C)** The sensitivity of the low-risk group to Sunitinib, VX.702, AZD6482 was higher than that of the high-risk group. The abscissa is the low-risk group and the high-risk group, and the ordinate is the value of the drug IC50. **(D–F)** The high-risk group had higher sensitivity to OSI.906, FH535, and Bryostatin.1 than the low-risk group. **(G)** Risk scores for A_1, B_1, and C_1 subtypes in CRGs molecular subtypes. **(H)** Risk scores for A_2, B_2, C_2 subtypes in genotyping. **(I)** Expression levels of CRGs in high and low risk groups.

'PDHB'. From Figures 7A, B we can see that LightGBM has the highest AUC in both training and validation sets, 1 and 0.750, respectively. The values of LightGBM and XGBoost in the multi-model forest graph in Figure 7C are also the highest at 0.748. Tables 1 and 2 show that the AUC, cutoff, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1 score, Kappa value of LightGBM are 1.000, 0.637, 0.995, 1.000, 1.000, 1.000, 0.986, 1.000, 0.989. In conclusion, LightGBM is the best performing model, and we choose this model to establish a melanoma metastasis model.

Interpretability of the metastasis model

After filtering out the best performing LightGBM model, we used the "SHAP" package to explain the importance of key

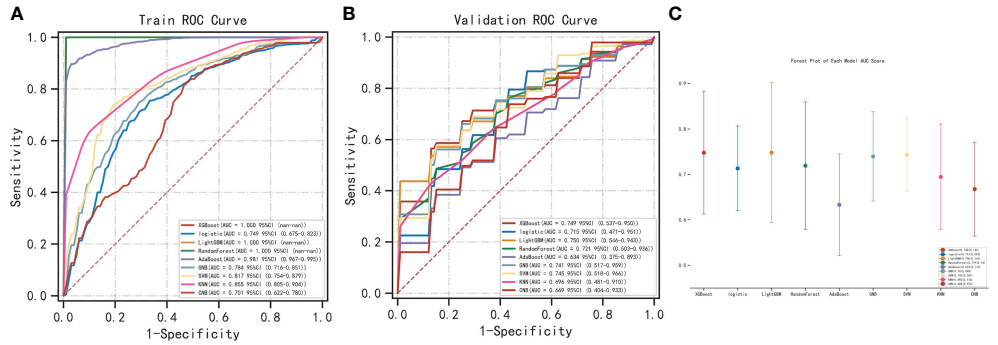


FIGURE 7
Construction of metastasis model. **(A)** REFCV method to filter out key metastatic variables in train set. **(B)** REFCV method to filter out key metastatic variables in validation set. **(C)** Multi-model forest graph.

10.3389/fimmu.2022.986214

TABLE 1 Multi-model comparison, training set results.

Model	AUC (SD)	cutoff (SD)	accuracy (SD)	sensitivity (SD)	specificity (SD)	positive predictive value (SD)	negative predictive value (SD)	F1 score (SD)	Kappa (SD)
XGBoost	1.000 (0.000)	0.863 (0.025)	0.995(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	0.986(0.000)	1.000(0.000)	0.989 (0.000)
logistic	0.749 (0.009)	0.656 (0.018)	0.712(0.018)	0.709(0.045)	0.730(0.042)	0.828(0.015)	0.572(0.024)	0.763(0.023)	0.406 (0.026)
LightGBM	1.000 (0.000)	0.637 (0.019)	0.995(0.000)	1.000(0.000)	1.000(0.000)	1.000(0.000)	0.986(0.000)	1.000(0.000)	0.989 (0.000)
RandomForest	1.000 (0.000)	0.623 (0.048)	0.988(0.006)	0.999(0.003)	0.999(0.004)	0.999(0.002)	0.969(0.017)	0.999(0.002)	0.974 (0.014)
AdaBoost	0.980 (0.004)	0.504 (0.001)	0.918(0.014)	0.894(0.025)	0.976(0.017)	0.986(0.010)	0.825(0.031)	0.938(0.012)	0.828 (0.028)
GNB	0.783 (0.008)	0.629 (0.032)	0.744(0.014)	0.772(0.033)	0.706(0.029)	0.828(0.010)	0.620(0.025)	0.798(0.015)	0.456 (0.022)
CNB	0.700 (0.008)	0.495 (0.001)	0.728(0.013)	0.849(0.023)	0.519(0.026)	0.763(0.008)	0.640(0.030)	0.803(0.011)	0.376 (0.027)
SVM	0.815 (0.006)	0.686 (0.014)	0.764(0.010)	0.744(0.024)	0.814(0.025)	0.880(0.011)	0.627(0.016)	0.806(0.012)	0.515 (0.016)
KNN	0.856 (0.014)	0.760 (0.080)	0.610(0.052)	0.688(0.104)	0.853(0.115)	0.982(0.037)	0.478(0.040)	0.802(0.052)	0.315 (0.066)

variables to the model. As shown in Figure 8A, the importance of 7 variables from high to low is: 'FDX1', ' DBT', 'LIPT1', 'PDHB', 'DLD', 'DLAT', 'LIAS'. Figure 8B shows the contribution of each variable to the model. The red dots indicate positive contributions, and the blue dots indicate negative contributions. A point closer to the left indicates a smaller value and a point closer to the right indicates a larger value. For example, the higher the FDX1 value, the higher the probability of death from heart failure; the lower the FDX1 value, the lower the probability of heart failure death. At the same time, we use the force diagram to illustrate 2 samples to show how different variables contribute to different samples. Figures 8C, D show the model predicts that these two samples are likely to metastasize and not metastasize, respectively, and show the contribution of each gene's expression to the sample prediction. Red indicates a positive contribution. Blue represents a negative contribution. If f(x) is greater than the cut-off value, the tumor sample is more likely to metastasize; if f (x) is less than the cut-off value, the tumor sample is less likely to metastasize.

TABLE 2 Multi-model comparison, validation set results.

Model	AUC (SD)	cutoff (SD)	accuracy (SD)	sensitivity (SD)	specificity (SD)	positive predictive value (SD)	negative predictive value (SD)	F1 score (SD)	Kappa (SD)
XGBoost	0.743 (0.130)	0.863 (0.025)	0.647(0.150)	0.706(0.192)	0.853(0.225)	0.801(0.159)	0.511(0.135)	0.731(0.139)	0.309 (0.275)
logistic	0.726 (0.130)	0.656 (0.018)	0.680(0.094)	0.838(0.143)	0.720(0.138)	0.800(0.089)	0.554(0.120)	0.813(0.099)	0.343 (0.180)
LightGBM	0.746 (0.126)	0.637 (0.019)	0.703(0.101)	0.720(0.178)	0.844(0.181)	0.809(0.118)	0.561(0.126)	0.753(0.142)	0.367 (0.227)
RandomForest	0.763 (0.115)	0.623 (0.048)	0.699(0.097)	0.707(0.150)	0.842(0.149)	0.844(0.075)	0.575(0.121)	0.760(0.101)	0.403 (0.166)
AdaBoost	0.701 (0.116)	0.504 (0.001)	0.668(0.125)	0.712(0.170)	0.804(0.216)	0.766(0.115)	0.537(0.158)	0.727(0.120)	0.296 (0.255)
GNB	0.737 (0.112)	0.629 (0.032)	0.703(0.092)	0.782(0.167)	0.782(0.128)	0.794(0.087)	0.612(0.186)	0.778(0.121)	0.362 (0.196)
CNB	0.676 (0.140)	0.495 (0.001)	0.693(0.144)	0.736(0.253)	0.798(0.184)	0.742(0.101)	0.597(0.348)	0.725(0.178)	0.291 (0.346)
SVM	0.737 (0.115)	0.686 (0.014)	0.689(0.085)	0.768(0.165)	0.789(0.136)	0.843(0.097)	0.551(0.108)	0.793(0.109)	0.375 (0.165)
KNN	0.686 (0.186)	0.760 (0.080)	0.561(0.110)	0.540(0.254)	0.882(0.175)	0.896(0.172)	0.446(0.075)	0.629(0.239)	0.232 (0.170)

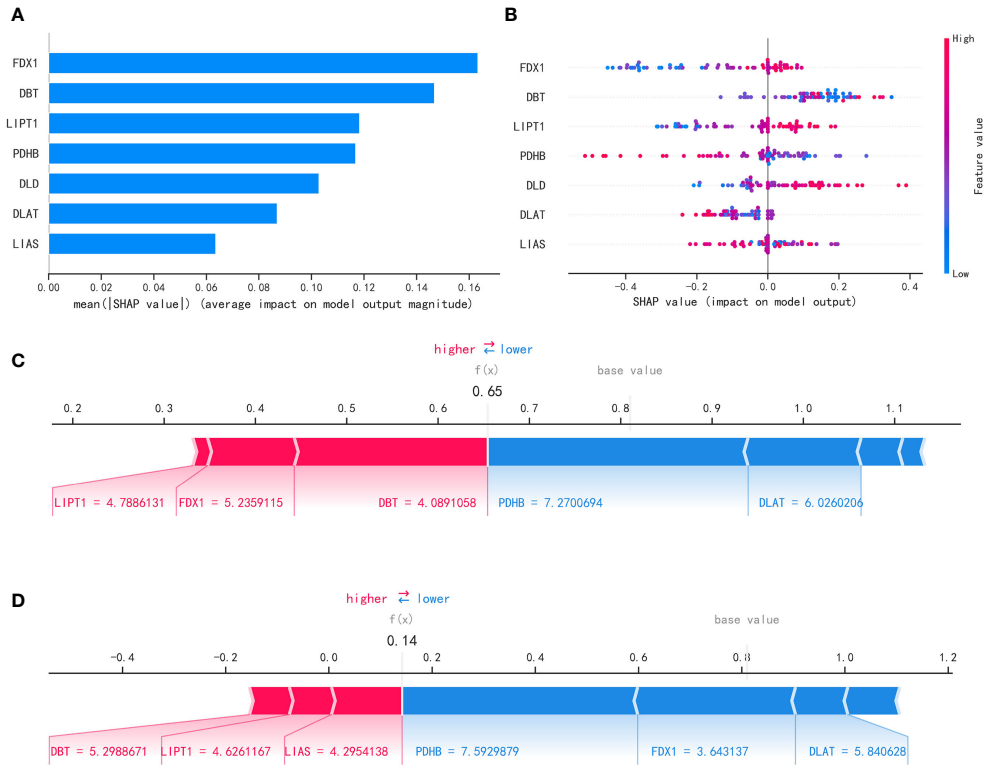


FIGURE 8
Interpretability of the metastasis model. **(A)** "SHAP" package to explain the importance of key variables to the model. **(B)** Contribution of each variable to the model. **(C, D)** Prediction of model.

Knockdown of fdx1 inhibits the proliferation of melanoma cells

We used specific FDX1-targeting siRNAs to knockdown the expression levels of FDX1 in the A375 cells (Figure 9A). siNC was used as a control group for subsequent comparative analysis. CCK-8 assay results showed that the proliferation of FDX1 knockdown cells at 12h, 24h, and 36h was significantly higher than that of the control group (Figure 9B). Wound healing assay results showed that FDX1 knockdown inhibited wound healing (Figure 9C). siNC group healed slightly faster than the siFDX1 group. However, this result is not statistically significant.

Discussion

As one of the deadliest tumors in skin cancer, melanoma is characterized by high invasiveness and high mortality (1). Therefore, a large body of literature has explored the prognosis and metastasis of melanoma. At present, the literature has predicted the prognosis of melanoma patients based on the expression levels of pyroptotic genes, tumor microenvironment status, or m6a-regulated methylation patterns (23–30). Although many bioinformatics studies are predictingognosis of melanoma, the existing cuproptosis-related melanoma research is not abundant.

Recently, Tsvetkov et al. discovered a novel apoptosis-independent cell death pathway, copper-dependent cell death (termed cuproptosis) (7). They proved that copper ions bind directly to the lipoylated components of the tricarboxylic acid cycle. Then, proteotoxic stress and unique cell death were induced. At the same time, the role performed by cuproptosis in tumours is gradually being understood. Zhong Hao et al. discovered that 6 CRGs had good diagnostic efficacy in kidney renal clear cell carcinoma (31). Besides, Liyang et al. developed a safe, mitochondria-targeted, copper-depleted nanoparticle (CDN) and tested its efficacy against triple-negative breast cancer (TNBC) (32). Injection CDN into mice with triple negative breast cancer resulted in a significant reduction in tumour growth and a significantly longer survival time for the mice. Zhang Zheng et al. constructed a prognostic model of HCC using the expression levels of ferredoxin 1 (FDX1) in hepatocellular carcinoma (HCC). They found that the expression level of FDX1 was significantly lower in HCC patients than in the non-HCC population (33). At the same

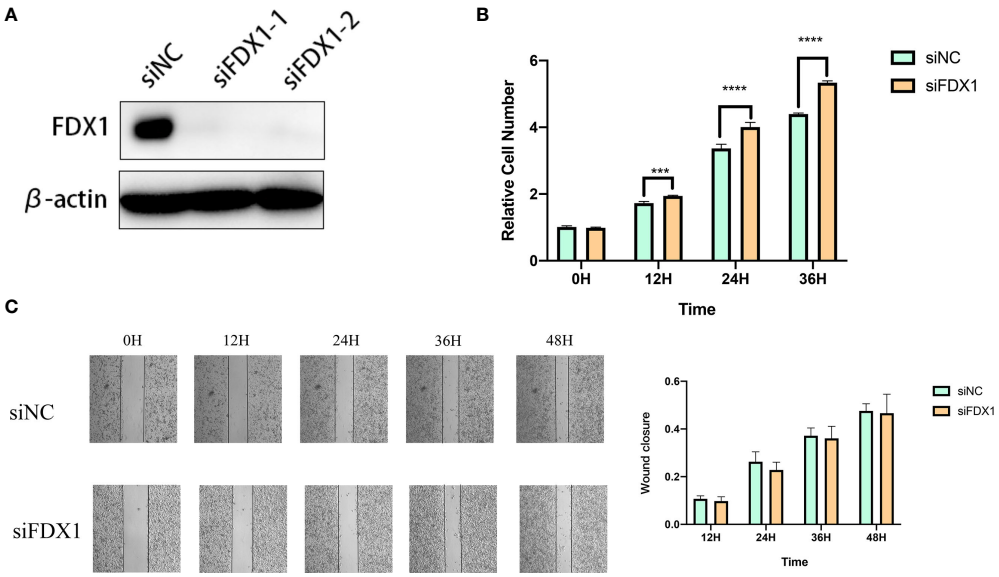


FIGURE 9
Knockdown of fdx1 inhibits the proliferation of melanoma cells. **(A)** FDX1-targeting siRNAs to knockdown the expression levels of FDX1. **(B)** The proliferation of FDX1 knockdown cells at 12 h, 24 h, and 36 h. **(C)** Wound healing assay at 12 h, 24 h, 36 h, and 48 h.

time, survival time was significantly higher in patients with high expression of HCC than in those with low expression of HCC.These studies suggest that cuproptosis has implications for the clinical diagnosis and treatment of tumours.

Therefore, CRGs were used to construct molecular subtypes of melanoma and to construct metastasis models in this research. The molecular subtypes of melanoma based on CRGs can give us a more comprehensive understanding of melanoma. At the same time, the metastasis model established based on CRGs can also fill the gap in melanoma bioinformatics research.

In this study, we explored the effects of CRGs on both survival and metastasis in melanoma patients. We analyzed the expression of 12 CRGs in TCGA and GEO cohorts. First, through bioinformatics analysis, we constructed molecular subtypes of 3 CRGs (A_1, B_1, C_1) based on 12 CRGs. Among the three molecular subtypes of CRGs, the C_1 subtype had the best survival outcome, and the A_1 subtype had the worst survival outcome. Next, we obtained 71 Differentially expressed genes that were co-expressed by all three subtypes. Based on 71 Differentially expressed genes, we genotyped melanoma and obtained 3 gene subtypes (A_2, B_2, C_2). Among them, the B_2 type had the best survival outcome, and the C_2 type had the worst survival outcome. Then, we screened prognosis-related genes from 71 co-expressed Differentially expressed genes. After obtaining 16 prognosis-related genes, the LASSO algorithm was used to screen out 6 key variable genes(AIM2, EDNRB, SLC39A6, TMEM117, PTPRC, and KIF14) for model construction and validation. Ultimately, our risk score model can distinguish between high-risk and low-risk groups. And KM analysis, AUC analysis, nomogram, and calibration curve indicated that our model could predict the prognosis of melanoma patients more accurately. Finally, in the analysis of metastasis, we used the LightGBM machine learning algorithm to screen out 7 CRGs to establish the metastasis model of melanoma.

In the TCGA cohort, we found that 9 out of 13 CRGs had an impact on the prognosis of melanoma patients. Therefore, this sparked our interest in investigating the role of CRGs in melanoma prognosis and metastasis. The degree of immune cell infiltration was also significantly different among the three molecular subtypes of CRGs. We selected 23 immune cells for analysis, except for Eosinophilna cells, the other 22 immune cells had significantly different infiltration degrees in the three subtypes. This suggests that immune cells play different roles in different subtypes. In many tumors, the immune microenvironment plays an important role in tumor angiogenesis, tumor invasion, and metastasis. Patients with high expression of CXCL9, CXCL10, CXCL13, CCL4, and CCL5 in SKCM (Skin cutaneous melanoma) had better overall survival (27). Some studies have constructed risk models based on immune-related genes and found that immune cell infiltration is different between patients with high and low immune scores, and the survival time of patients with high immune scores is significantly lower than that of patients with low immune scores. Other studies have shown that in melanoma patients, IL27 is closely related to CD8+ cells, and is related to the treatment effect and prognosis of patients (34–36).

Enrichment analysis of Metascape shows 71 intergenes were mainly enriched in MHC class II antigen presentation and platinum resistance pathways. Melanoma-specific MHC-II expression predicted anti-pd-1/PD-L1 treatment efficacy (37). Overexpression of BCL2L10 in melanoma has also been shown to promote cisplatin and ABT-737 resistance (38). In a case report, a patient with metastatic melanoma was also associated with hyperprolactinemia (39). Next, we used cox regression analysis and the LASSO algorithm to screen out 6 key variable genes(AIM2, EDNRB, SLC39A6, TMEM117, PTPRC, and KIF14) to construct a risk model. Absent in melanoma 2 (AIM2) is a cytoplasmic sensor that recognizes double-stranded DNA derived from viruses, bacteria, or the host itself, and is a member of the interferon inducible p200-protein (IFI-P200) family of immune-related proteins. AIM2 plays a significant role in autoimmune diseases (40) and the activation of inflammasome (41–44). In the melanoma-related literature, patients with melanoma whose dendritic cells express AIM2 have a significantly lower prognosis than patients with melanoma whose dendritic cells do not express AIM2 (45). In breast cancer treatment, Dihydroartemisinin induces pyroptosis in breast cancer cells by promoting the AIM2/caspase-3/DFNA5 (gasdermin E) axis (46). Endothelin Receptor type B (EDNRB) is widely expressed in vascular endothelial cells of the cardiovascular system, gastrointestinal tract, lung, kidney, adrenal gland, uterus, prostate, and brain. In melanoma-related studies, the prognostic value of patients with high CD8 (+) T cell subpopulations expressing EDNRB was significantly reduced (47). This suggests that EDNRB could be a potential therapeutic target for melanoma. Solute carrier family 39 member 6(SLC39A6) is also known as LIV-1, ZIP-6, and Zinc transporter ZIP6. May act as a zinc-influx transporter. Solute Carrier Family 39 Member 6 (SLC39A6) is also known as LIV-1, ZIP-6, and Zinc transporter ZIP6. May act as a zinc-influx transporter. In studies of esophageal cancer, SLC39A6 increases the invasiveness of esophageal cancer cells and reduces patient prognosis by increasing the level of Zinc expression in esophageal cancer cells (48). SLC39A6 can also be used as an indicator for early diagnosis of esophageal cancer (49). However, in luminal breast cancer, the Oestrogen-regulated protein SLC39A6 acts as a benign prognostic indicator (50). One study reported that transmembrane protein 117 (TMEM117) was associated with endoplasmic reticulum stress-mediated mitochondrial-mediated cell death (51). Studies have shown that in primary liver cancer, miR-631 can target the receptor protein tyrosine phosphatase gene (PTPRE) to inhibit the intrahepatic metastasis of liver cancer (52). In kras mutant lung adenocarcinoma, the PTPRE is highly expressed, which can be used as a novel therapeutic target in kras mutant lung adenocarcinoma (53). The kinesin family member 14 (KIF14), is a novel oncogene located on chromosome 1q. When it malfunctions, it can affect the development of the brain and kidneys, and it can lead to many types of cancer (54, 55). In breast cancer, high expression of KIF14 can promote breast cancer metastasis and is associated with poor prognosis of breast cancer patients (56, 57). Similarly, studies have shown that in gastric cancer, when KIF14 mRNA is highly expressed, the prognosis is significantly lower than that with low KIF14 mRNA expression (58). However, the above two genes have not been deeply studied in melanoma research, and the specific functions of PTPRE and KIF14 in melanoma need to be further explored.

Among these CRGs, we also screened out 7 key genes (FDX1, DBT, LIPT1, PDHB, DLD, DLAT, LIAS) as variables in the metastasis model. In other tumor metastasis models, the roles of some of these genes in tumor metastasis have also been found. Chen found that LIPT1 may be a prognostic-related gene for bladder cancer, and then found that this gene has a certain degree of inhibitory effect on the migration ability of bladder cancer cells by transwell method (59). Zhao found that PDHB is associated with ovarian cancer growth and metastasis, and miR-203 can target the 3'-UTR of PDHB to promote glycolysis. Meanwhile, overexpression of PDHB could abolish the promoting effect of miR-203 on ovarian cancer cell growth (60). Regarding the role of these genes in tumor metastasis, we still need further functional tests to verify.

During the occurrence and development of tumor tissue, there are a large number of gene mutations. Mutated genes can provide tumor antigens that can be recognized by the immune system as non-self tissues, inducing immune cells to respond (61). Immunotherapy takes advantage of the fact that immune cells can recognize and eliminate tumor cells, which plays a great role in the treatment of tumors (62, 63). However, tumors effectively suppress immune responses (immune escape) by activating negative regulatory pathways associated with immune homeostasis (checkpoints) or by adopting features that allow them to actively evade detection (64, 65). Effective immunotherapy drugs have been approved in preclinical and clinical phase I-III trials for highly aggressive, highly refractory, and advanced and metastatic melanoma (66). For example, the anti-PD-1 monoclonal antibodies nivolumab and pembrolizumab and the anti-CTLA-4 antibody ipilimumab are being tested in clinical trials to treat melanoma (67). Studies have shown that commonly used immune checkpoint inhibitors (ICIs) can improve progression-free survival and overall survival in melanoma patients (68, 69). In our study, the risk score model showed that the degree of immune cell infiltration in the high-risk group was significantly lower than that in the low-risk group. Interestingly, the survival time of the high-risk group was significantly lower than that of the low-risk group. In other melanoma studies, the survival time of the high immune score group was significantly higher than that of the low immune score group (70, 71). We also propose a hypothesis here, in melanoma, is the degree of immune cell infiltration positively correlated with the survival time of patients? This problem also needs more clinical data or experiments to confirm. It is worth

mentioning that in this study, we also analyzed the drug sensitivity between high and low risk groups. We screened 98 drugs (Supplementary File 1) with significant differences in IC50 concentrations between high and low risk groups. Among them, worthy of our attention are Sunitinib, VX-702, and Bryostatin.1. Sunitinib is a new class of drugs that can selectively target multiple receptor tyrosine kinases, and is now being used alone or in combination with other antitumor drugs to treat many solid tumors, including liver cancer, renal cancer, and gastric cancer (72–74). VX-702 is a highly selective p38α MAPK inhibitor targeting nimokinase for the treatment of primary and acquired endocrine-resistant breast cancer (75). Bryostatin-1 is a protein kinase C (PKC) inhibitor that inhibits cell entry into mitosis, lowers pH and energy metabolism, and reduces tumor blood flow, thereby inhibiting tumor cell growth (76, 77). We screened 98 drugs to guide the development of melanoma drugs.

Furthermore, our *in vitro* experiments showed that FDX1 promoted the growth, and migration of melanoma cells. Therefore, we speculate that FDX1, as a CRG, is a marker of melanoma. People with high expression of this gene need to be more alert to the occurrence of melanoma. At the same time, it is also a prognostic marker for melanoma patients. The prognosis of cancer patients may be better than that of other patients. Taken together, our results suggest that FDX1 is aberrantly expressed in melanoma and may be associated with patient prognosis. In the future, we need to conduct more in-depth functional experiments to explore how this gene acts on the occurrence and development of melanoma.

This study established prognostic and metastatic models of CRGs in melanoma. But there are still some limitations. Although the sample size of our sequencing data is relatively large, it is mainly based on the data of the network database, and we also need our own sequencing data to verify. In our *in vitro* experiments, knockdown of FDX1 reduced the ability of cells to migrate, but there was no difference compared to the control group. However, this gene was selected in our model, which may be due to the joint effect of multiple genes in the establishment of the metastasis model. In the future, we will conduct more in-depth experiments to explore its transfer mechanism. We analyzed the enriched pathways and functions of these key genes, and functional assays are needed to verify them. Finally, the drugs we screened also need to be verified by drug resistance experiments.

Conclusion

In this study, melanoma was classified based on 12 CRGs and clinical features, and three subtypes, A_1, B_1, and C_1, were established. Among them, the C_1 subtype had the best survival outcome and the highest immune cell infiltration. Then, A_2, B_2, and C_2 subtypes were established based on genotyping, with the B_2 subtype having the best survival outcome. We performed functional analysis on the intergenes between different types, and the results showed that these intergenes were mainly enriched in cell cycle and drug metabolism pathways. We also established a prognostic model using 6 key variable genes and analyzed the tumor microenvironment according to the high and low risk scores of prognosis. In addition, we screened drugs for high and low risk groups and found that 98 drugs had significant differences in IC50 concentrations in high and low risk groups. Finally, we used the LightGBM algorithm to screen out 7 CRGs to build the transfer model of melanoma. These results help us to understand the role of CRGs in the occurrence and development of melanoma, and provide us with new therapeutic ideas and potential treatment methods.

Data availability statement

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

Author contributions

JL was responsible for the study concept and design. Y-WL and FL revised the manuscript and made final approval of the version. JL, ZL, and LL analyzed the data. ZL, JL, and LL helped with the experimental section. LL helped to write the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Key R&D Program of China (No.2020YFC2005000), the National Natural Science Foundation of China (No. 81802668, 82172832), the Wisdom Accumulation and Talent Cultivation Project of the Third Xiangya Hospital of Central South University (YX202108), and the Postgraduate Research and Innovation Project of Central South University (No.2020zzts892).

Acknowledgments

The authors would like to acknowledge the GEO databases (https://www.ncbi.nlm.nih.gov/gds/) and TCGA database (https://tcga-data.nci.nih.gov/tcga/) for providing their platforms and those contributors for uploading their valuable datasets. And we also thanks for Extreme Smart Analysis

(https://www.xsmartanalysis.com/) for providing the platforms to analyse our metastasis model.

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/fimmu.2022.986214/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Mutations and copy number variations of CRGs in the TCGA cohort. **(A)** The proportion of mutation frequency of CRGs in melanoma. **(B)** The chromosome where the mutated CRGs are located. Red represents an increase in copy number and green represents a decrease in copy number. **(C)** CRGs copy number variation graph. The ordinate of the red circle is the number of samples with increased copy number, and the ordinate of the green circle is the number of samples with reduced copy number.

SUPPLEMENTARY FIGURE 2
9 CRGs associated with prognosis. **(A–I)** Comparison of the overall survival time of samples with high expression of CRGs genes (indicated in red) and samples with low expression of CRGs (indicated in blue).

SUPPLEMENTARY FIGURE 3
Melanoma gene subtypes constructed based on prognostic-related intersection differentially expressed genes. **(A)** Three gene subtypes obtained by unsupervised consensus clustering method. **(B)** Consistent distribution of different K values described by a consistent cumulative distribution function (CDF) plot. **(C)** The delta area score displayed the relative growth in cluster stability. **(D)** Kaplan Meier analysis results of three gene subtypes. **(E)** Comparison of CRGs expression levels among the three gene subtypes. **(F)** Heatmap showing clinical information and gene expression profiles for the three gene subtypes.

SUPPLEMENTARY TABLE 1
Names of 13 cuproptosis-related genes

SUPPLEMENTARY FILE 1
98 drugs were with significant differences in IC50 concentrations between high and low risk groups.

References

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

2. Pasquali S, Hadjinicolaou AV, Chiarion Sileni V, Rossi CR, Mocellin S. Systemic treatments for metastatic cutaneous melanoma. *Cochrane Database Syst Rev* (2018) 2(2):CD011123. doi: 10.1002/14651858.CD011123.pub2

3. Grzywa TM, Koppolu AA, Paskal W, Klicka K, Rydzanicz M, Wejman J, et al. Higher mutation burden in high proliferation compartments of heterogeneous melanoma tumors. *Int J Mol Sci* (2021) 22:3886. doi: 10.3390/ijms22083886

4. Elder DE, Bastian BC, Cree IA, Massi D, Scolyer RA. The 2018 world health organization classification of cutaneous, mucosal, and uveal melanoma: Detailed analysis of 9 distinct subtypes defined by their evolutionary pathway. *Arch Pathol Lab Med* (2020) 144:500–22. doi: 10.5858/arpa.2019-0561-RA

5. Strashilov S, Yordanov A. Aetiology and pathogenesis of cutaneous melanoma: Current concepts and advances. *Int J Mol Sci* (2021) 22:6395. doi: 10.3390/ijms22126395

6. Kahlson MA, Dixon SJ. Copper-induced cell death. *Science* (2022) 375:1231–32. doi: 10.1126/science.abo3959

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

8. Newly found roles for copper. *Nutr Rev* (1985) 43:117–19. doi: 10.1111/j.1753-4887.1985.tb06882.x

9. Huff JD, Keung Y-K, Thakuri M, Beaty MW, Hurd DD, Owen J, et al. Copper deficiency causes reversible myelodysplasia. *Am J Hematol* (2007) 82:625–30. doi: 10.1002/ajh.20864

10. da Silva Fonseca J, de Barros Marangoni LF, Marques JA, Bianchini A. Energy metabolism enzymes inhibition by the combined effects of increasing temperature and copper exposure in the coral mussismilia harttii. *Chemosphere* (2019) 236:124420. doi: 10.1016/j.chemosphere.2019.124420

11. Tisato F, Marzano C, Porchia M, Pellei M, Santini C. Copper in diseases and treatments, and copper-based anticancer strategies. *Med Res Rev* (2010) 30:708–49. doi: 10.1002/med.20174

12. Jiang Y, Huo Z, Qi X, Zuo T, Wu Z. Copper-induced tumor cell death mechanisms and antitumor theragnostic applications of copper complexes. *Nanomed (Lond)* (2022) 17:303–24. doi: 10.2217/nnm-2021-0374

13. Shanbhag VC, Gudekar N, Jasmer K, Papageorgiou C, Singh K, Petris MJ. Copper metabolism as a unique vulnerability in cancer. *Biochim Biophys Acta Mol Cell Res* (2021) 1868:118893. doi: 10.1016/j.bbamcr.2020.118893

14. Arnesano F, Natile G. Interference between copper transport systems and platinum drugs. *Semin Cancer Biol* (2021) 76:173–88. doi: 10.1016/j.semcancer.2021.05.023

15. Panja S, Rahem S, Chu CJ, Mitrofanova A. Big data to knowledge: Application of machine learning to predictive modeling of therapeutic response in cancer. *Curr Genomics* (2021) 22:244–66. doi: 10.2174/1389202921999201224110101

16. Nayarisseri A, Khandelwal R, Tanwar P, Madhavi M, Sharma D, Thakur G, et al. Artificial intelligence, big data and machine learning approaches in precision medicine & drug discovery. *Curr Drug Targets* (2021) 22:631–55. doi: 10.2174/1389450122999210104205732

17. Hong M, Tao S, Zhang L, Diao L-T, Huang X, Huang S, et al. RNA Sequencing: new technologies and applications in cancer research. *J Hematol Oncol* (2020) 13:166. doi: 10.1186/s13045-020-01005-x

18. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers* (2021) 25:1315–60. doi: 10.1007/s11030-021-10217-3

19. Garbulowski M, Smolinska K, Çabuk U, Yones SA, Celli L, Yaz EN, et al. Machine learning-based analysis of glioma grades reveals Co-enrichment. *Cancers* (2022) 14:1014. doi: 10.3390/cancers14041014

20. Chen D-L, Cai J-H, Wang CCN. Identification of key prognostic genes of triple negative breast cancer by LASSO-based machine learning and bioinformatics analysis. *Genes (Basel)* (2022) 13:902. doi: 10.3390/genes13050902

21. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. *Physiol Genomics* (2020) 52:200–02. doi: 10.1152/physiolgenomics.00029.2020

22. Fu Q, Hu L, Xu Y, Yi Y, Jiang L. High lipoprotein(a) concentrations are associated with lower type 2 diabetes risk in the Chinese han population: a large retrospective cohort study. *Lipids Health Dis* (2021) 20:76. doi: 10.1186/s12944-021-01504-x

23. Huang R, Mao M, Lu Y, Yu Q, Liao L. A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment. *Aging (Albany NY)* (2020) 12:6966–80. doi: 10.18632/aging.103054

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

25. Liu D, Yang X, Wu X. Tumor immune microenvironment characterization identifies prognosis and immunotherapy-related gene signatures in melanoma. *Front Immunol* (2021) 12:663495. doi: 10.3389/fimmu.2021.663495

26. Ju A, Tang J, Chen S, Fu Y, Luo Y. Pyroptosis-related gene signatures can robustly diagnose skin cutaneous melanoma and predict the prognosis. *Front Oncol* (2021) 11:709077. doi: 10.3389/fonc.2021.709077

27. Huang B, Han W, Sheng Z-F, Shen G-L. Identification of immune-related biomarkers associated with tumorigenesis and prognosis in cutaneous melanoma patients. *Cancer Cell Int* (2020) 20:195. doi: 10.1186/s12935-020-01271-2

28. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* (2019) 176:775-789.e18. doi: 10.1016/j.cell.2018.11.043

29. Tang J, Wan Q, Lu J. The prognostic values of m6A RNA methylation regulators in uveal melanoma. *BMC Cancer* (2020) 20:674. doi: 10.1186/s12885-020-07159-8

30. Shi S, Fan Z, Liu Y, Huang C, Zhou J. Integration analysis of m6A related genes in skin cutaneous melanoma and the biological function research of the SPRR1B. *Front Oncol* (2021) 11:729045. doi: 10.3389/fonc.2021.729045

31. Ji Z-H, Ren W-Z, Wang H-Q, Gao W, Yuan B. Molecular subtyping based on cuproptosis-related genes and characterization of tumor microenvironment infiltration in kidney renal clear cell carcinoma. *Front Oncol* (2022) 12:919083. doi: 10.3389/fonc.2022.919083

32. Cui L, Gouw AM, LaGory EL, Guo S, Attarwala N, Tang Y, et al. Mitochondrial copper depletion suppresses triple-negative breast cancer in mice. *Nat Biotechnol* (2021) 39:357–67. doi: 10.1038/s41587-020-0707-9

33. Zhang Z, Zeng X, Wu Y, Liu Y, Zhang X, Song Z. Cuproptosis-related risk score predicts prognosis and characterizes the tumor microenvironment in hepatocellular carcinoma. *Front Immunol* (2022) 13:925618. doi: 10.3389/fimmu.2022.925618

34. Yuan Y, Zhu Z, Lan Y, Duan S, Zhu Z, Zhang X, et al. Development and validation of a CD8+ T cell infiltration-related signature for melanoma patients. *Front Immunol* (2021) 12:659444. doi: 10.3389/fimmu.2021.659444

35. Tan Y, Chen Q, Li X, Zeng Z, Xiong W, Li G, et al. Pyroptosis: a new paradigm of cell death for fighting against cancer. *J Exp Clin Cancer Res* (2021) 40:153. doi: 10.1186/s13046-021-01959-x

36. Dong C, Dang D, Zhao X, Wang Y, Wang Z, Zhang C. Integrative characterization of the role of IL27 in melanoma using bioinformatics analysis. *Front Immunol* (2021) 12:713001. doi: 10.3389/fimmu.2021.713001

37. Johnson DB, Estrada MV, Salgado R, Sanchez V, Doxie DB, Opalenik SR, et al. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat Commun* (2016) 7:10582. doi: 10.1038/ncomms10582

38. Li Y, Zhang J, Liu Y, Zhang B, Zhong F, Wang S, et al. MiR-30a-5p confers cisplatin resistance by regulating IGF1R expression in melanoma cells. *BMC Cancer* (2018) 18:404. doi: 10.1186/s12885-018-4233-9

39. Manning A, Rassie K, Rivalland G. A case of hyperprolactinaemia in a patient with metastatic melanoma. *Melanoma Res* (2021) 31:277–79. doi: 10.1097/CMR.0000000000000738

40. Chou W-C, Guo Z, Guo H, Chen L, Zhang G, Liang K, et al. AIM2 in regulatory T cells restrains autoimmune diseases. *Nature* (2021) 591:300–05. doi: 10.1038/s41586-021-03231-w

41. Fernandes-Alnemri T, Yu J-W, Datta P, Wu J, Alnemri ES. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* (2009) 458:509–13. doi: 10.1038/nature07710

42. Hu B, Jin C, Li H-B, Tong J, Ouyang X, Cetinbas NM, et al. The DNA-sensing AIM2 inflammasome controls radiation-induced cell death and tissue injury. *Science* (2016) 354:765–68. doi: 10.1126/science.aaf7532

43. Onódi Z, Ruppert M, Kucsera D, Sayour AA, Tóth VE, Koncsos G, et al. AIM2-driven inflammasome activation in heart failure. *Cardiovasc Res* (2021) 117:2639–51. doi: 10.1093/cvr/cvab202

44. Youm Y-H, Nguyen KY, Grant RW, Goldberg EL, Bodogai M, Kim D, et al. The ketone metabolite β-hydroxybutyrate blocks NLRP3 inflammasome-mediated inflammatory disease. *Nat Med* (2015) 21:263–9. doi: 10.1038/nm.3804

45. Fukuda K, Okamura K, Riding RL, Fan X, Afshari K, Haddadi N-S, et al. AIM2 regulates anti-tumor immunity and is a viable therapeutic target for melanoma. *J Exp Med* (2021) 218:e20200962. doi: 10.1084/jem.20200962

46. Li Y, Wang W, Li A, Huang W, Chen S, Han F, et al. Dihydroartemisinin induces pyroptosis by promoting the AIM2/caspase-3/DFNA5 axis in breast cancer cells. *Chem Biol Interact* (2021) 340:109434. doi: 10.1016/j.cbi.2021.109434

47. Deng W, Ma Y, Su Z, Liu Y, Liang P, Huang C, et al. Single-cell RNA-sequencing analyses identify heterogeneity of CD8(+) T cell subpopulations and novel therapy targets in melanoma. *Mol Ther Oncolytics* (2021) 20:105–18. doi: 10.1016/j.omto.2020.12.003

48. Cheng X, Wei L, Huang X, Zheng J, Shao M, Feng T, et al. Solute carrier family 39 member 6 gene promotes aggressiveness of esophageal carcinoma cells by increasing intracellular levels of zinc, activating phosphatidylinositol 3-kinase signaling, and up-regulating genes that regulate metastasis. *Gastroenterology* (2017) 152:1985–97.e12. doi: 10.1053/j.gastro.2017.02.006

49. Cui X-B, Shen Y-y, Jin T-t, Li S, Li T-t, Zhang S-m, et al. SLC39A6: a potential target for diagnosis and therapy of esophageal carcinoma. *J Transl Med* (2015) 13:321. doi: 10.1186/s12967-015-0681-z

50. Althobiti M, El-Sharawy KA, Joseph C, Aleskandarany M, Toss MS, Green AR, et al. Oestrogen-regulated protein SLC39A6: a biomarker of good prognosis in luminal breast cancer. *Breast Cancer Res Treat* (2021) 189:621–30. doi: 10.1007/s10549-021-06336-y

51. Tamaki T, Kamatsuka K, Sato T, Morooka S, Otsuka K, Hattori M, et al. A novel transmembrane protein defines the endoplasmic reticulum stress-induced cell death pathway. *Biochem Biophys Res Commun* (2017) 486:149–55. doi: 10.1016/j.bbrc.2017.03.017

52. Chen B, Liao Z, Qi Y, Zhang H, Su C, Liang H, et al. miR-631 inhibits intrahepatic metastasis of hepatocellular carcinoma by targeting PTPRE. *Front Oncol* (2020) 10:565266. doi: 10.3389/fonc.2020.565266

53. Codreanu SG, Hoeksema MD, Slebos RJC, Zimmerman LJ, Rahman SMJ, Li M, et al. Identification of proteomic features to distinguish benign pulmonary nodules from lung adenocarcinoma. *J Proteome Res* (2017) 16:3266–76. doi: 10.1021/acs.jproteome.7b00245

54. Li KK-W, Qi Y, Xia T, Chan AK-Y, Zhang Z-Y, Aibaidula A, et al. The kinesin KIF14 is overexpressed in medulloblastoma and downregulation of KIF14 suppressed tumor proliferation and induced apoptosis. *Lab Invest* (2017) 97:946–61. doi: 10.1038/labinvest.2017.48

55. Reilly ML, Stokman MF, Magry V, Jeanpierre C, Alves M, Paydar M, et al. Loss-of-function mutations in KIF14 cause severe microcephaly and kidney development defects in humans and zebrafish. *Hum Mol Genet* (2019) 28:778–95. doi: 10.1093/hmg/ddy381

56. Singel SM, Cornelius C, Zaganjor E, Batten K, Sarode VR, Buckley DL, et al. KIF14 promotes AKT phosphorylation and contributes to chemoresistance in triple-negative breast cancer. *Neoplasia* (2014) 16:247–56, 56.e2. doi: 10.1016/j.neo.2014.03.008

57. Li T-F, Zeng H-J, Shan Z, Ye R-Y, Cheang T-Y, Zhang Y-J, et al. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. *Cancer Cell Int* (2020) 20:123. doi: 10.1186/s12935-020-01191-1

58. Yang X, Li C, Yan C, Li J, Yan M, Liu B, et al. KIF14 promotes tumor progression and metastasis and is an independent predictor of poor prognosis in human gastric cancer. *Biochim Biophys Acta Mol Basis Dis* (2019) 1865:181–92. doi: 10.1016/j.bbadis.2018.10.039

59. Chen Y, Xu T, Xie F, Wang L, Liang Z, Li D, et al. Evaluating the biological functions of the prognostic genes identified by the pathology atlas in bladder cancer. *Oncol Rep* (2021) 45:191–201. doi: 10.3892/or.2020.7853

60. Xiaohong Z, Lichun F, Na X, Kejian Z, Xiaolan X, Shaosheng W. MiR-203 promotes the growth and migration of ovarian cancer cells by enhancing glycolytic pathway. *Tumour Biol* (2016) 37:14989–97. doi: 10.1007/s13277-016-5415-1

61. Hegde PS, Chen DS. Top 10 challenges in cancer immunotherapy. *Immunity* (2020) 52:17–35. doi: 10.1016/j.immuni.2019.12.011

62. Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol* (2020) 17:807–21. doi: 10.1038/s41423-020-0488-6

63. Magrì A, Germano G, Lorenzato A, Lamba S, Chilà R, Montone M, et al. High-dose vitamin c enhances cancer immunotherapy. *Sci Transl Med* (2020) 12:eaay8707. doi: 10.1126/scitranslmed.aay8707

64. Galassi C, Musella M, Manduca N, Maccafeo E, Sistigu A. The immune privilege of cancer stem cells: A key to understanding tumor immune escape and therapy failure. *Cells* (2021) 10:2361. doi: 10.3390/cells10092361

65. Ubellacker JM, Tasdogan A, Ramesh V, Shen B, Mitchell EC, Martin-Sandoval MS, et al. Lymph protects metastasizing melanoma cells from ferroptosis. *Nature* (2020) 585:113–18. doi: 10.1038/s41586-020-2623-z

66. Marzagalli M, Ebelt ND, Manuel ER. Unraveling the crosstalk between melanoma and immune cells in the tumor microenvironment. *Semin Cancer Biol* (2019) 59:236–50. doi: 10.1016/j.semcancer.2019.08.002

67. Alrabadi NN, Abushukair HM, Ababneh OE, Syaj SS, Al-Horani SS, Qarqash AA, et al. Systematic review and meta-analysis efficacy and safety of immune checkpoint inhibitors in advanced melanoma patients with anti-PD-1 progression: a systematic review and meta-analysis. *Clin Transl Oncol* (2021) 23:1885–904. doi: 10.1007/s12094-021-02598-6

68. Robert C, Ribas A, Schachter J, Arance A, Grob J-J, Mortier L, et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): *post-hoc* 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. *Lancet Oncol* (2019) 20:1239–51. doi: 10.1016/S1470-2045(19)30388-2

69. Hodi FS, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Cowey CL, et al. Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. *Lancet Oncol* (2018) 19:1480–92. doi: 10.1016/S1470-2045(18)30700-9

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

71. Liu Q, Nie R, Li M, Li L, Zhou H, Lu H, et al. Identification of subtypes correlated with tumor immunity and immunotherapy in cutaneous melanoma. *Comput Struct Biotechnol J* (2021) 19:4472–85. doi: 10.1016/j.csbj.2021.08.005

72. Qi X, Yang M, Ma L, Sauer M, Avella D, Kaifi JT, et al. Synergizing sunitinib and radiofrequency ablation to treat hepatocellular cancer by triggering the antitumor immune response. *J Immunother Cancer* (2020) 8:e001038. doi: 10.1136/jitc-2020-001038

73. Hojo Y, Kishi S, Mori S, Fujiwara-Tani R, Sasaki T, Fujii K, et al. Sunitinib and pterostilbene combination treatment exerts antitumor effects in gastric cancer *via* suppression of PDZD8. *Int J Mol Sci* (2022) 23:4002. doi: 10.3390/ijms23074002

74. Choueiri TK, Motzer RJ, Rini BI, Haanen J, Campbell MT, Venugopal B, et al. Updated efficacy results from the JAVELIN renal 101 trial: first-line avelumab plus axitinib versus sunitinib in patients with advanced renal cell carcinoma. *Ann Oncol* (2020) 31:1030–39. doi: 10.1016/j.annonc.2020.04.010

75. Zhou Y, Zhou H, Hua L, Hou C, Jia Q, Chen J, et al. Verification of ferroptosis and pyroptosis and identification of PTGS2 as the hub gene in human coronary artery atherosclerosis. *Free Radic Biol Med* (2021) 171:55–68. doi: 10.1016/j.freeradbiomed.2021.05.009

76. Koutcher JA, Motwani M, Zakian KL, Li XK, Matei C, Dyke JP, et al. The *in vivo* effect of bryostatin-1 on paclitaxel-induced tumor growth, mitotic entry, and blood flow. *Clin Cancer Res* (2000) 6:1498–507.

77. Kennedy MJ, Prestigiacomo LJ, Tyler G, May WS, Davidson NE. Differential effects of bryostatin 1 and phorbol ester on human breast cancer cell lines. *Cancer Res* (1992) 52:1278–83.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 27 October 2022
DOI 10.3389/fimmu.2022.1013049

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Yongjian Li,
Zhejiang University, China
Fangdie Ye,
Fudan University, China

*CORRESPONDENCE
Zhixian He
hezhixian@ntu.edu.cn
Sujie Ni
nisujie@fudan.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 06 August 2022
ACCEPTED 29 September 2022
PUBLISHED 27 October 2022

CITATION
Zhong X, Xu S, Wang Q, Peng L, Wang F, He T, Liu C, Ni S and He Z (2022) CAPN8 involves with exhausted, inflamed, and desert immune microenvironment to influence the metastasis of thyroid cancer.
Front. Immunol. 13:1013049.
doi: 10.3389/fimmu.2022.1013049

CAPN8 involves with exhausted, inflamed, and desert immune microenvironment to influence the metastasis of thyroid cancer

Xiang Zhong[1†], Shu Xu[2†], Quhui Wang[1], Long Peng[3], Feiran Wang[1], Tianyi He[1], Changyue Liu[1], Sujie Ni[2*] and Zhixian He[1*]

[1]Department of Thyroid and Breast Surgery, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong, China, [2]Department of Oncology, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong, China, [3]Department of Neurosurgery, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong, China

Background: Thyroid cancer (THCA) is the most prevalent malignant disease of the endocrine system, in which 5-year survival can attain about 95%, but patients with metastasis have a poor prognosis. Very little is known about the role of CAPN8 in the metastasis of THCA. In particular, the effect of CAPN8 on the tumor immune microenvironment (TIME) and immunotherapy response is unclear.

Material and methods: Multiome datasets and multiple cohorts were acquired for analysis. Firstly, the expression and the prognostic value of CAPN8 were explored in public datasets and *in vitro* tumor tissues. Then, hierarchical clustering analysis was performed to identify the immune subtypes of THCA according to the expression of CAPN8 and the activities of related pathways. Subsequent analyses explored the different patterns of TIME, genetic alteration, DNA replication stress, drug sensitivity, and immunotherapy response among the three immune phenotypes. Finally, five individual cohorts of thyroid cancer were utilized to test the robustness and extrapolation of the three immune clusters.

Results: CAPN8 was found to be a significant risk factor for THCA with a markedly elevated level of mRNA and protein in tumor tissues. This potential oncogene could induce the activation of epithelial–mesenchymal transition and E2F-targeted pathways. Three subtypes were identified for THCA, including immune exhausted, inflamed, and immune desert phenotypes. The exhausted type was characterized by a markedly increased expression of inhibitory receptors and infiltration of immune cells but was much more likely to respond to immunotherapy. The immune desert type was resistant to common chemotherapeutics with extensive genomic mutation and copy number variance.

Conclusion: The present study firstly explored the role of CAPN8 in the metastasis of THCA from the aspects of TIME. Three immune subtypes were identified with quite different patterns of prognosis, immunotherapy response, and drug sensitivity, providing novel insights for the treatment of THCA and helping understand the cross-talk between CAPN8 and tumor immune microenvironment.

KEYWORDS

CAPN8, thyroid cancer, prognosis, immunotherapy, tumor immune microenvironment

Introduction

Thyroid carcinoma (THCA) is the most prevalent malignant disease of the endocrine system, which can be divided into four histological types, including papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), medullary thyroid cancer, and poorly differentiated thyroid cancer (1). The 5-year survival rate for patients with PTC or FTC can attain about 95%, but patients with metastatic THCA have a poor prognosis (2). Calpain calcium kinase (CAPN) is a kind of cysteine protein kinase widely existing in most eukaryotic cells and plays a key role in regulating cell cycle and apoptosis (3). It is already reported that the aberrant expression of CAPN is involved in several types of cancer progression by inducing NF-κB, focal adhesion kinase, and MYC pathways (4–7). However, very little is known about the role of CAPN8 in the genesis and development of THCA up to now. In particular, the effect of CAPN8 on the tumor immune microenvironment (TIME), which is a well-recognized factor in promoting the metastasis of THCA, is unclear (8).

Hereby we hypothesize that CAPN8 might facilitate the metastasis of thyroid cancer cells and lead to poor prognosis by inducing an inhibitory TIME pattern. Firstly, we explored the expression and downstream signaling pathways of CAPN8 in The Cancer Genome Atlas—Thyroid Cancer (TCGA-THCA) cohort and *in vitro* tumor tissues. Next, clustering analysis was performed, and three immune-related clusters (immune exhausted, immune desert, and inflamed) were identified for THCA according to the expression of CAPN8 and related pathways. Subsequent analyses examined the different patterns of genetic alteration, DNA replication stress, TIME, immunotherapy response, drug sensitivity, and prognosis amid the three immune clusters of THCA. Finally, external validation cohorts were utilized to test the robustness and extrapolation of the three immune clusters. Overall, the present study is aimed at elucidating the role of CAPN8 in the metastasis of THCA from the aspects of TIME, DNA replication stress, and genetic variation. These findings will provide novel insights for the treatment of THCA and help understand the cross-talk between CAPN8 and the tumor immune microenvironment.

Materials and methods

Multiome dataset acquisition and processing

Multiome datasets of thyroid cancer were obtained from TCGA-THCA (497 tumor samples and 71 normal samples) (9). RNA-seq data, downloaded in the format of fragments per kilobase million at the UCSC Xena website (9), was transformed into the value of transcripts per kilobase million for further analysis. Information about copy number variance (CNV) was acquired from the FireBrowse (10) data portal. Detailed somatic mutation categories were retrieved from the cBioPortal (11) online platform.

Meanwhile, five datasets of thyroid cancer were also exported from the Gene Expression Omnibus database for external validation, including GSE3467 (n = 9), GSE3678 (n = 7), GSE33630 (n = 49), GSE60542 (n = 33), and GSE27155 (n = 95) cohorts. The batch effect amid different arrays was eliminated by using the ComBat function of R (version 4.1.3) package sva (12). As these data are open access resources from public database where patients' consents were already obtained, extra informed consents are not needed.

Immunoreactive score calculation

Firstly, the number of positive cells and the total cells in each stained section were counted to calculate the positive rate (PP) (PP% = positive cells/total cells). By averaging the PP values of 10 discontinuous fields of the experimental tissue in a microscope with a 200-fold high-power lens, the patients were scored with 0 point for no positive cells and 1, 2, 3, and 4 points

for 0% < PP ≤ 10%, 11% ≤ PP <50%, 50% ≤ PP < 80%, and PP ≥ 80%, respectively.

Then, the staining intensity (SI) of cells in the tissue was estimated based on the shades of cell color. The SI score was marked as 0 when there was no obvious staining and 1, 2, and 3 for light brownish yellow, brownish yellow, and brown staining, respectively.

The final IRS score was calculated by the following formula: IRS = PP × SI. IRS >3 indicates a high expression, while IRS ≤3 represents a low expression.

Detecting the mRNA and protein expression of CAPN8 in THCA

The expression difference of CAPN8 between 33 types of cancer and cancer-adjacent tissues was illustrated in a boxplot by using the UCSC Xena web browser. Meanwhile, immunohistochemistry (IHC) was also performed on isolated thyroid cancer tissues to detect the level of CAPN8 protein. IHC was conducted as described previously (13), and tumor sections were obtained from patients who had received radical surgery for thyroid carcinoma in the Affiliated Hospital of Nantong University. The primary anti-bodies used for IHC were anti-CAPN8 (1:30, biorbyt, orb140072).

Prognostic value and biological function of CAPN8 in THCA

To test the prognostic value of CAPN8, 496 patients in the TCGA-THCA cohort were divided into CAPN8-high and CAPN8-low groups according to the median mRNA value. The Kaplan–Meier (K-m) curve and log-rank test were then utilized to show their difference in progression-free survival (PFS) time.

To explore the biological function of CAPN8, differential expression analysis was carried out between CAPN8-high and CAPN8-low groups by using R package limma (14). $|\text{Log}_2$ fold change (FC)| >1 and false discovery rate (FDR) <0.05 were set as the significant threshold. Subsequently, Gene Set Enrichment Analysis (GSEA) (15) was performed to recognize the differentially expressed pathways (DEPs) between CAPN8-high and CAPN8-low groups by using R package clusterProfiler. In total, 50 well-known cancer hallmarks (16) were set as the background gene sets, and FDR <0.05 was chosen as the significant threshold.

Screening for CAPN8-related cancer hallmarks in THCA

To elucidate the potential regulating mechanism of CAPN8 toward thyroid cancer cells, Least Absolute Selection and Shrinkage Operator (LASSO) penalty and ridge regression were implemented to screen the 22 significant DEPs by using R package glmnet (17). In addition, random survival forest (RSF) algorithm was also employed to compute the significance of each DEP by using the R packages randomForestSRC and randomSurvivalForest (https://CRAN.R-project.org/package=beeswarm) with the minimal depth method to determine the final number of prognostic variables. Pathways with a certain contribution to patients' overall survival were screened in the two models. The importance of each variable was then visualized in a bar plot, and the marginal effect was displayed by the function plot.error of R package randomSurvivalForest.

Identifying the subtypes of THCA by hierarchical clustering analysis

The six DEPs, obtained by adaptive LASSO regression, and eight DEPs, obtained by random survival forest, were taken into intersection with a Venn diagram depicting the common DEPs by using R package VeenDiagram (18). Gene Set Variation Analysis (GSVA) (15) was then performed to quantify the pathway activities of three common DPEs in THCA. Afterwards, survival analysis was conducted to demonstrate the impact of the three DEPs on PFS. The patients were stratified into two groups according to the median activity score of each DEP, with the K-m curve showing their difference in PFS. Of the common DEPs, HALLMARK_E2F_TARGETS was a prominent risk factor for THCA patients. Therefore, the core enrichment genes of HALLMARK_E2F_TARGETS were then submitted to hierarchical clustering analysis to identify the subtypes of THCA, resulting in an E2F-Clust with two sub-clusters. Hierarchical clustering was completed with Ward's Clustering, computing the Euclidean distance among each patient by using R function hclust. Consensus Cumulative Distribution Function (CDF) and Delta area (relative change of area under the CDF curve) were used to select the proper clustering numbers. The two indices were provided in R package "ConsensusClusterPlus".

Characterizing the different TIME patterns between the two E2F-clusters and three ImmClusters of THCA

GSEA analysis was wielded to dissect the biological features of the two E2F-Clusters of THCA, and survival analysis was conducted to explore their difference in PFS. Moreover, the infiltrating proportions of 10 immune cells were calculated for each TCHA patient by using R package MCPcounter (19) to probe the different TIME pattern (tumor immune microenvironment) between the two E2F-Clusters. In addition, the expression profiles of eight well-known immune inhibitory receptors (IRs)—CD274, PDCD1, CD247, PDCD1LG2, CTLA4, TNFRSF9, TNFRSF4, and TLR9—were also explored in the two E2F-Clusters.

With these immune-related information, the two E2F-Clusters were further subdivided into three clusters (ImmCluster) by using the hclust function in R package ComplexHeatmap (20). Subsequently, the immune enrichment score (IES) and the stromal enrichment score (SES) were compared among the three ImmClusters. IES and SES were obtained by applying ESTIMATE algorithm (21) to each TCHA patient, where IES represented the enrichment score of immune cell ingredients, while SES reflected the ratio of stromal components in tumor tissues. Furthermore, the patients' diverse responses to immunotherapy were predicted among the three ImmClusters by using Tumor Immune Dysfunction and Exclusion) algorithm (22) and R package Submap (23, 24).

Resolving the genetic alteration paradigm in two E2F-Clusters and three ImmClusters of THCA

To characterize the different genetic alteration profile among the two E2F-Cluster and three ImmClusters, somatic mutation, CNV, and chromosome instability for each patient were explored by R package MOVICS (Multi-Omics Integration and Visualization in Cancer Subtyping.) (25). CN GISTIC score was also computed for patients in each Immun-Cluster to manifest their different chromosome instability. Specifically, FGA, FGL, and FGG represented the fraction of CN altered genome, fraction of CN-lost genome, and CN-gained genome, respectively.

Different patterns of DNA replication stress and drug resistance amid two E2F-Clusters and three ImmClusters of THCA

As E2F is part of the cell cycle-related pathways, which can lead to DNA replication stress and drug resistance in cancer cells, we further investigated the activities of 21 pathways related to DNA replication stress (26). Pathway activity was estimated by the GSVA strategy as previously described, and a heat map was utilized to demonstrate the difference between two E2F-Clusters and three ImmClusters. Moreover, the IC50 (half-maximal inhibitory concentration) of five typical chemotherapeutics for THCA was computed and compared between each cluster by using R package pRRophetic (27).

External validation for the applicability of E2F-Clust and ImmClust in five individual thyroid cancer cohorts

To extrapolate the E2F-Clust and ImmClust for further application in clinical practice, five external datasets of thyroid cancer cohorts were used for validation. One dataset was GSE27155 (n = 95) which was quantified by Affymetrix Human Genome U133A Array and annotated by GPL96 platform. The other four datasets—GSE3467 (n = 9), GSE3678 (n = 7), GSE33630 (n = 49), and GSE60542 (n = 33) were quantified by Affymetrix Human Genome U133 Plus 2.0 Array and annotated by GPL570 platform. The last four datasets were consolidated by using the ComBat function of R package sva. Principal component analysis was then utilized to visualize the homogeneity of different samples after combination.

Then, based on the core enrichment genes of the E2F pathway, hierarchical clustering analysis was implemented to seek similar subclusters in these datasets. Similarly, patterns of IRs, TIME, DNA replication stress, drug resistance, and immunotherapy responses were explored in different subclusters by using the same analysis strategies as described above.

Statistic and software

Data processing and bioinformatics analyses were accomplished by R (version 4.1.3). Packages like limma, ggplot2, survminer, clusterProfiler, GSVA, glmnet, MCPcounter, SubMap, MOVICS, *etc.*, were employed for analyses with proper citation. Wilcox or Kruskal–Wallis tests were applied for comparisons between two or more groups involved in this study. Pearson and Spearman rank correlation were adopted to estimate the statistical correlation of parametric or non-parametric variables. A log-rank test was utilized for survival analysis. Two-sided P <0.05 was considered the significant threshold for all statistical tests.

Results

Expression profile, prognostic value, biological function, and immune features of CAPN8 in THCA

CAPN8 showed a substantial rise in protein, mRNA, and relative IHC score (D) in thyroid cancer tissues (Figures 1A, B, D). Upon dividing the patients into two groups according to the median value of CAPN8 mRNA, patients in the CAPN8-high group were observed to have a noticeable worse survival outcome, suggesting a potential role of oncogene for CAPN8 (Figure 1C). The differential expression analysis showed that many genes showed a considerable un-regulation in the CAPN8-high group, while the levels of a relatively few genes decreased (Figure 1E). In terms of the 50 cancer hallmarks,

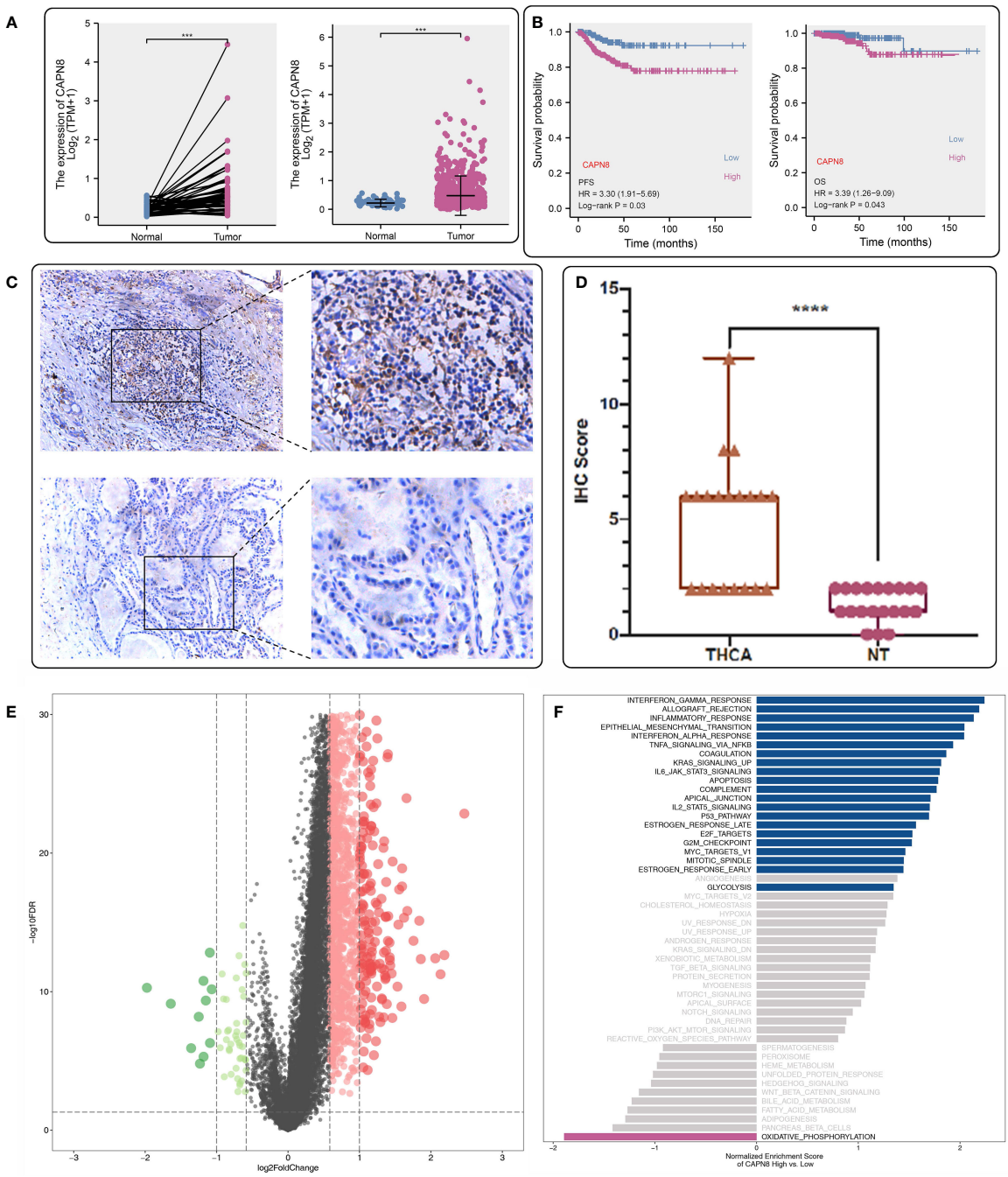


FIGURE 1
Expression profile and prognostic value of CAPN8 in thyroid cancer. **(A, B, D)** There was an obvious increase in the level of CAPN8 protein **(A)**, mRNA **(B)**, and relative immunohistochemistry score **(D)** in thyroid cancer tissues. **(C)** Influence of CAPN8 expression on the progression-free survival and overall survival of thyroid cancer. **(E)** Differentially expressed genes between the CAPN8-high and CAPN8-low groups. **(F)** Enrichment analysis identified many pathways which were activated in the CAPN8-high group. ***$P < 0.001$, ****$P < 0.0001$.

interferon-gamma response, inflammatory response, E2F-targets, *etc.*, were significantly upregulated in the CAPN8-high group, while the oxidative phosphorylation pathway was slightly downregulated (Figure 1F). We also detected that the strong relationship of CAPN8 with immolators were presented in the BEST website (Figure 2A), and CAPN8 could be an immunotherapy predictor for patients who underwent immunotherapy (Figure 2B).

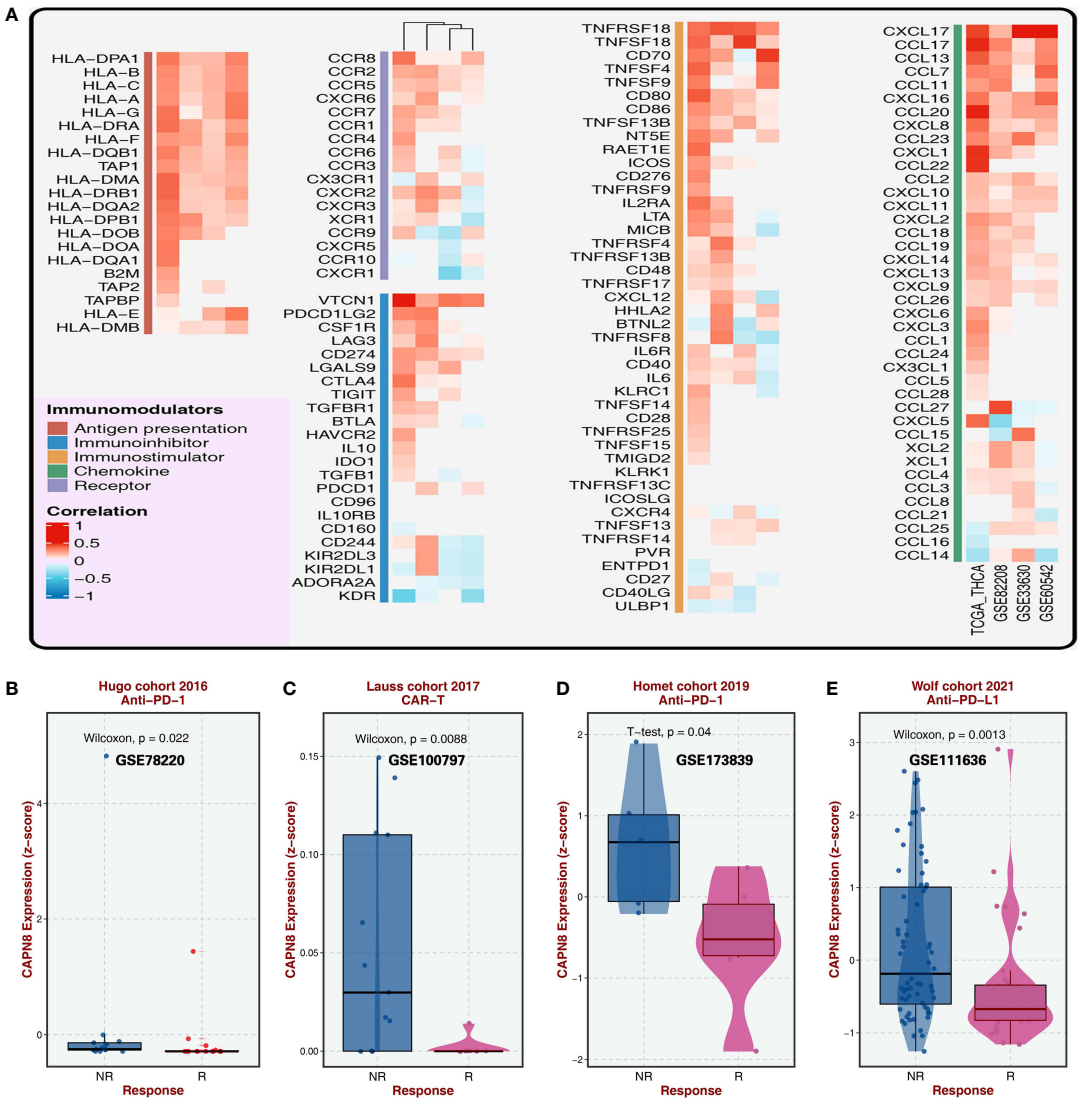


FIGURE 2
Relationship of CAPN8 with immolators and immunotherapy from the BEST website. **(A)** Relationship of CAPN8 with immunomodulators in the four thyroid cancer cohorts. **(B–E)** Different expression of CAPN8 in different status of patients from four cohorts with immunotherapy.

Screening of CAPN8-related signaling pathways and construction of E2F-Clust for THCA

Six and eight CAPN8-related cancer hallmarks were screened out by the adaptive LASSO regression and random survival forest algorithm, respectively (Figures 3A–C). Each pathway was graded in order of their importance, and the E2F-targets pathway showed a dominant impact on survival time in the RSF analysis (Figure 3D). In addition, the marginal effect of RSF was demonstrated in the scatter diagram where the E2F-target and G2M-checkpoint pathways exhibited a mild positive correlation with the mortality of THCA patients (Figure 3E). These findings suggest that the E2F-targets pathway could be the downstream signaling pathway of CAPN8 and plays a key role in THCA progression.

Subsequently, three shared cancer hallmarks were documented after taking the intersection of six pathways from LASSO regression and eight pathways from RSF analysis, including E2F-targets, oxidative phosphorylation, and inflammatory-response pathways (Figure 4A). Of the three pathways, E2F-targets was a risk factor, and patients with a high pathway activity of E2F-targets demonstrated a worse survival outcome (Figure 4B).

A further cluster analysis identified two subtypes of THCA based on the core genes of the E2F-targets pathway (Figure 4C).

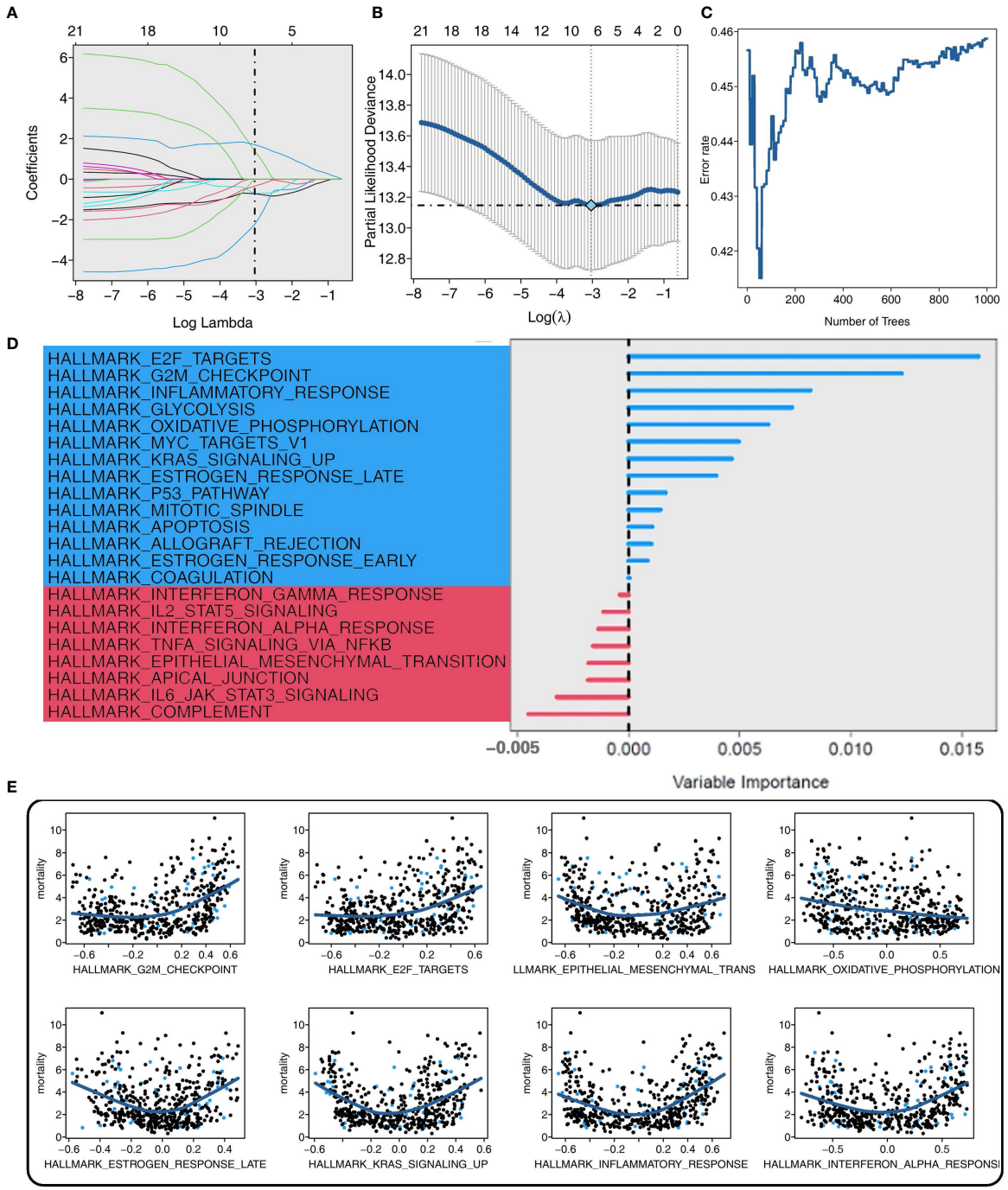


FIGURE 3
Screening of CAPN8-related hub pathways in thyroid cancer. **(A, B)** Coefficient profiles **(A)** and deviance profiles **(B)** of the adaptive Least Absolute Shrinkage and Selection Operator regression model. Six cancer hallmarks were selected. **(C)** Error rate of the random survival forest (RSF) model. Eight cancer hallmarks were determined. **(D, E)** Importance **(D)** and marginal effect **(E)** of each pathway in the RSF model. The blue color represented pathways with a certain influence on overall survival, while the red color represented background noise with no impact on the dependent variable.

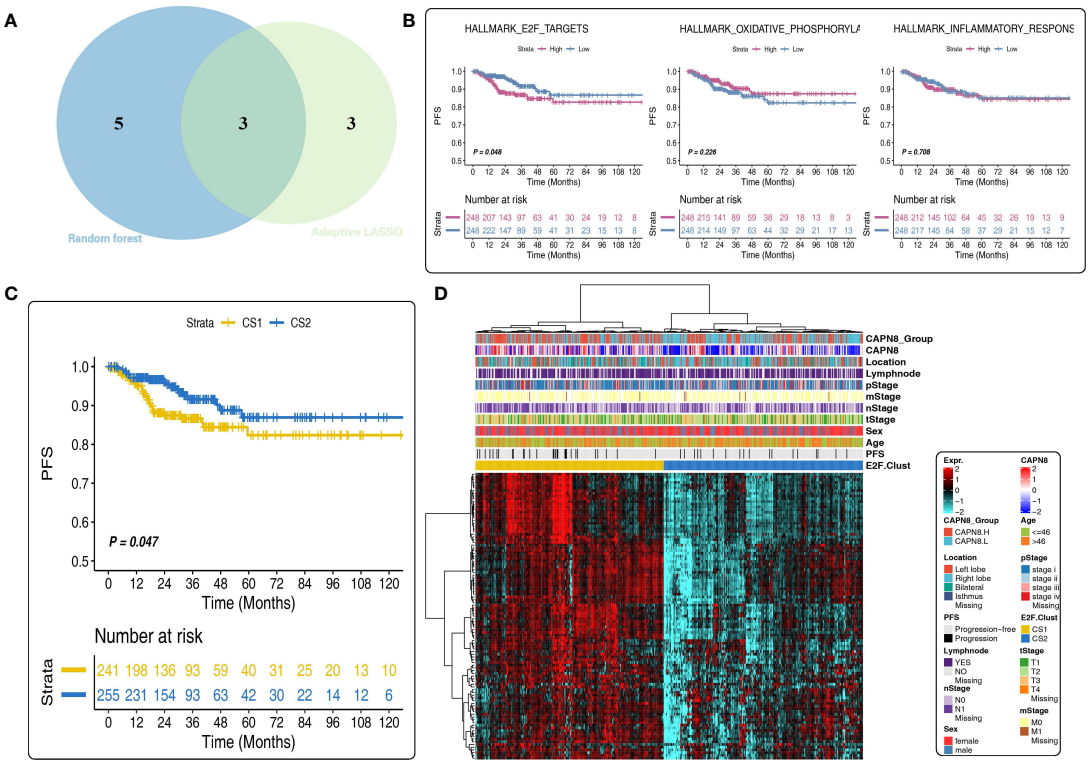


FIGURE 4
Prognostic value of the E2F-Clust and E2F-targets pathway. **(A)** Three mutual pathways were identified by taking an intersection of the results from the Least Absolute Shrinkage and Selection Operator model and the random survival forest model. **(B)** Prognostic value of three hub pathways. Only E2F-targets pathway was a significant risk factor for thyroid cancer. **(C, D)** E2F-Clust stratified the patients into two groups with a distinct difference in survival outcome **(C)** and clinical TNM stages **(D)**.

Of the E2F-Clust, CS1 (cluster 1) was characterized by an evidently high expression of CAPN8 as well as core genes in E2F-targets pathways, showing an unfavorable effect on patients' survival outcome (Figure 4D) compared to the superior influence of CS2 (cluster 2). This E2F-Cluster further supported that CAPN8 may lead to THCA progression by regulating the E2F-targets pathway.

Immune exhausted, immune desert, and inflamed TIME patterns in the two E2F-Clusters and three ImmClusters of THCA

There was a considerable diversity of biological function and TIME pattern between the two E2F-Clusters. CS1 was identified by the elevated pathway activity of E2F-targets, epithelial–mesenchymal transition (EMT), inflammatory response, and interferon-gamma response (Figure 5A).

Compared to the two E2F-Clusters, three subtypes were identified in the ImmClust with a significant difference in survival outcome (Figure 5B). Considering the information of 10 immune cells and eight IRs, CS1 of the ImmuClust was accompanied with an increased expression of IRs and an infiltrating ratio of almost all types of immune cells, underlying a strongly inhibitory TIME pattern in the CS1 group. This result suggested that CAPN8 could induce T cell exhaustion to inhibit immune response and lead to a poor prognosis of THCA (Figure 5C). Moreover, CS1 was characterized by a higher infiltrating proportion of fibroblast than CS2, accounting for its distinctly higher enriched score of SES (Figure 5D).

CS2 of the ImmClust, however, was characterized by an inflammatory TIME pattern with high levels of infiltration of neutrophil and endothelial cells. Distinctively, CS3 of the ImmClust lacked immune infiltration, suggesting a potential phenotype of immune desert for this subtype (Figure 5C). Keeping consistent with the exhausted TIME feature of CS1, immunotherapy seemed to be a feasible strategy for this subtype. Patients in CS1 significantly responded to PD1-R treatment (Figure 5E).

10.3389/fimmu.2022.1013049

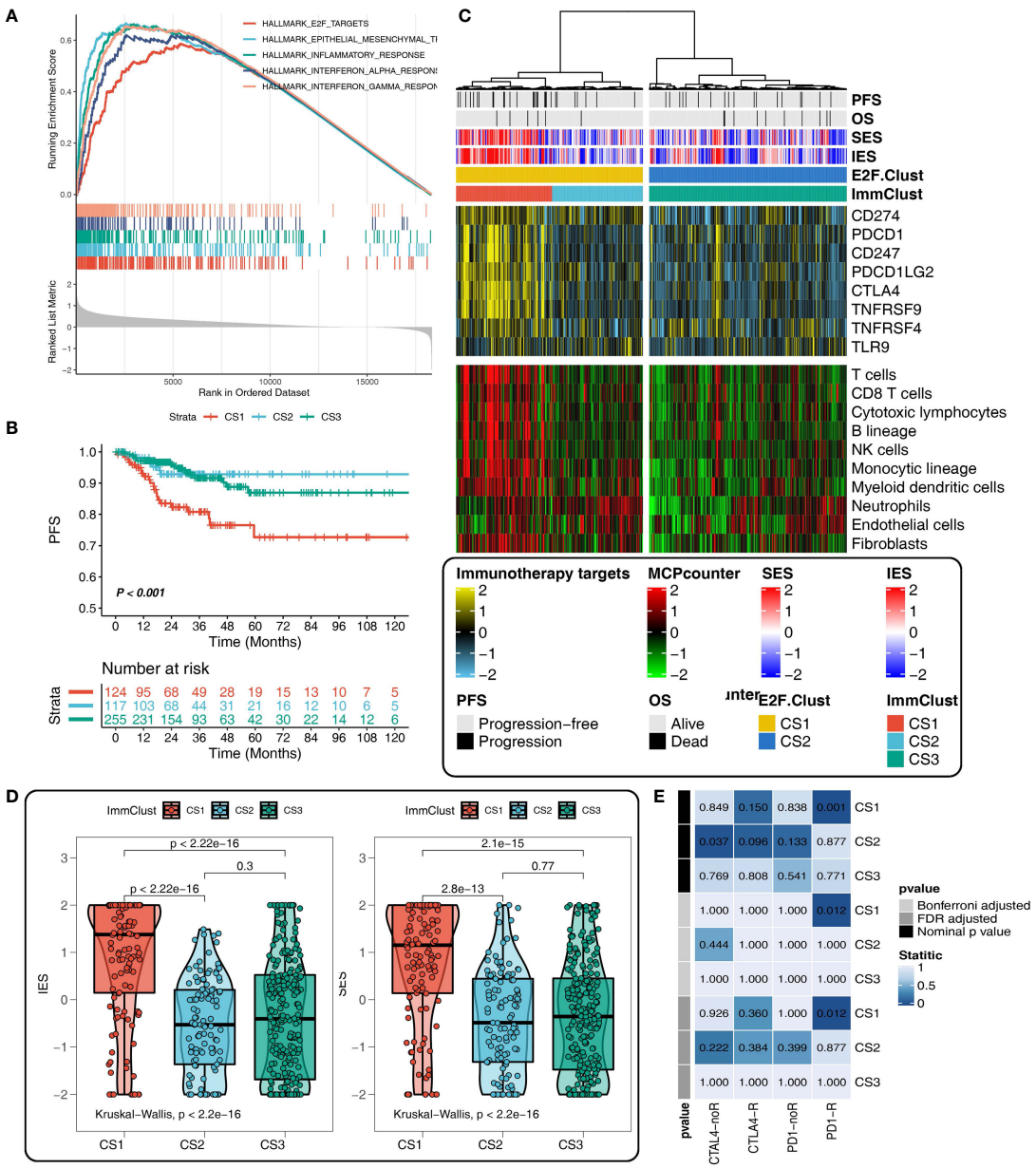


FIGURE 5
Different tumor immune microenvironment (TIME) patterns in the E2F-Clust and ImmClust. **(A)** The enrichment analysis identified five upregulated pathways in CS1 of E2F-Clust. **(B)** There was a significant difference in PFS among the three ImmClusters. **(C)** Different TIME patterns among the three ImmClusters, including the immune exhausted (CS1), inflammatory (CS2), and immune desert (CS3) subtypes. **(D)** The Immune Enrichment Score and the Stromal Enrichment Score differ in the three ImmClusters. **(E)** Different responses to immunotherapy of the three ImmClusters. OS, overall survival; PFS, progression-free survival; FDR, false discovery rate; TIME, tumor immune microenvironment.

Genetic alteration paradigm in the two E2F-Clusters and three ImmClusters of THCA

Oncoplot was illustrated to resolve the mutation profile of E2F-Clusters and ImmClusters. CS2 of the ImmClust seems to have the highest mutation frequency of BRAF and ZFHX3, while CS1 exceeded the other two clusters in the mutation frequency of COL5A3 and AKT1, which is a hub element of the PI3K proliferation pathway (Figure 6A).

By contrast to the mutation frequency, CS2 of the ImmClust was fairly stable in CNV with the lowest FGA, FGL, FGG, and

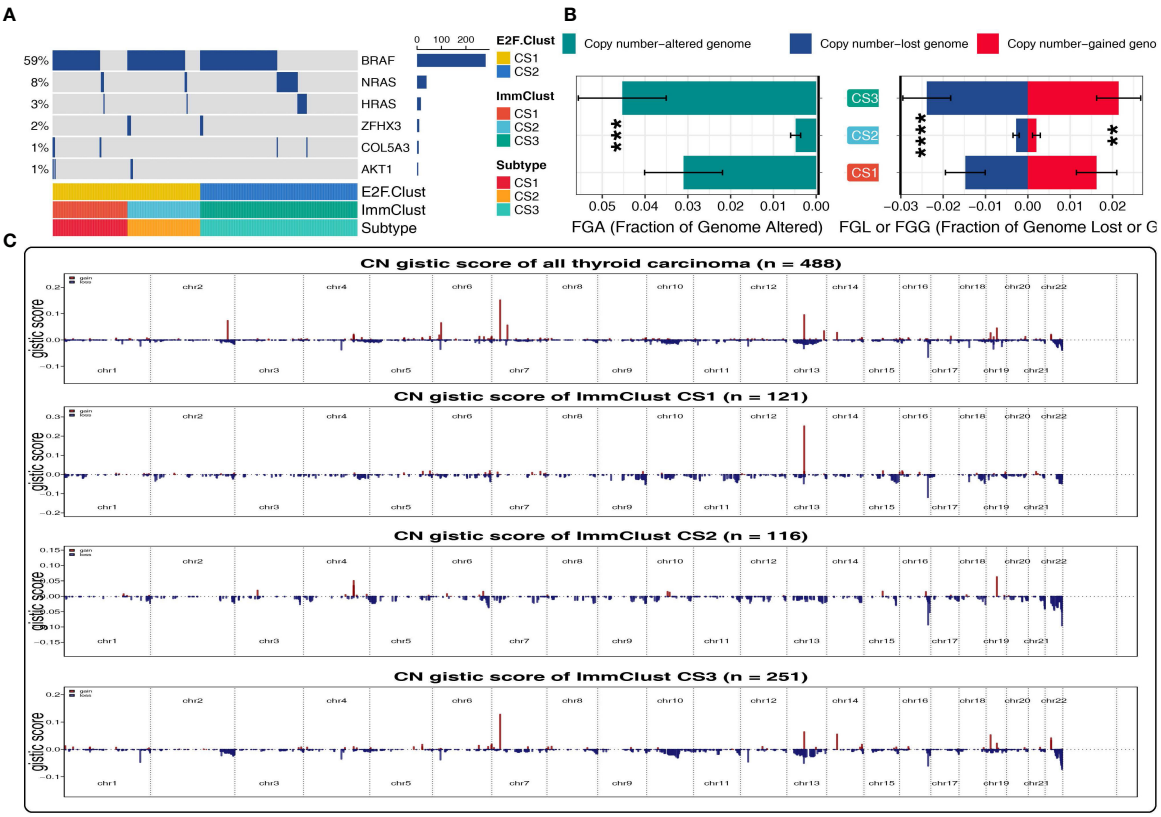


FIGURE 6
Patterns of genetic alteration in the E2F-Clust and ImmClust. **(A)** Oncoplot showing the mutation information of the E2F-Clust and ImmClust. **(B)** The fraction of copy number variance among the three ImmClusters. **, ***, and **** represent $p < 0.01$, $p < 0.001$, and $p < 0.0001$, respectively. **(C)** Copy number GISTIC score of the E2F-Clust and ImmClust.

CN GISTIC score than CS1 and CS3 (Figure 6B), suggesting a rather recent cancer origin and fairly low chromosomal instability for this subtype.

CS3, however, was in the lead in CNV frequency and CN GISTIC score, underlying a quite earlier cancer origin and extremely high chromosomal instability (Figure 6C) for this subtype.

Patterns of DNA replication stress and drug resistance in two E2F-Clusters and three ImmClusters of THCA

With regards to the 21 pathways related to DNA replication stress, CS3, the immune desert subtype, was dramatically downregulated in the pathway activity of cell cycle, G1S-DNA damage checkpoints, G2M-DNA damage checkpoint, and mitotic spindle checkpoint, implying a considerably declined ability to maintain the correct paradigm of DNA replication (Figure 7A).

This is in line with the result mentioned above, namely: there was a highly altered genomic CNV situation and increased chromosomal instability for CS3 cluster (Figure 6C), suggesting that CS3, the immune desert subtype, was prone to being drug resistant by inducing extensive epigenetic variations.

As expected, CS2 of the E2F-Clust, similar to CS3 of the ImmClust, was relatively insensitive to many chemotherapeutics and correlated with poor survival outcome (Figure 7B). Specifically, the estimated IC50 values of VE-822, AZD67738, VE821, and MK-1775 were widely increased (Figure 7C). These drugs are famous inhibitors of ATR and week1, which are famous cell cycle regulatory proteins.

Application of the E2F-Clust and ImmClust in five individual thyroid cancer cohorts

The z-score normalization for GSE3467 (n = 9), GSE3678 (n = 7), GSE33630 (n = 49), and GSE60542 (n = 33) was appropriate as the heterogeneity among the four datasets was eliminated after combination. After applying the E2F-Clust and ImmClust to this combined thyroid cancer cohort (n = 78), the

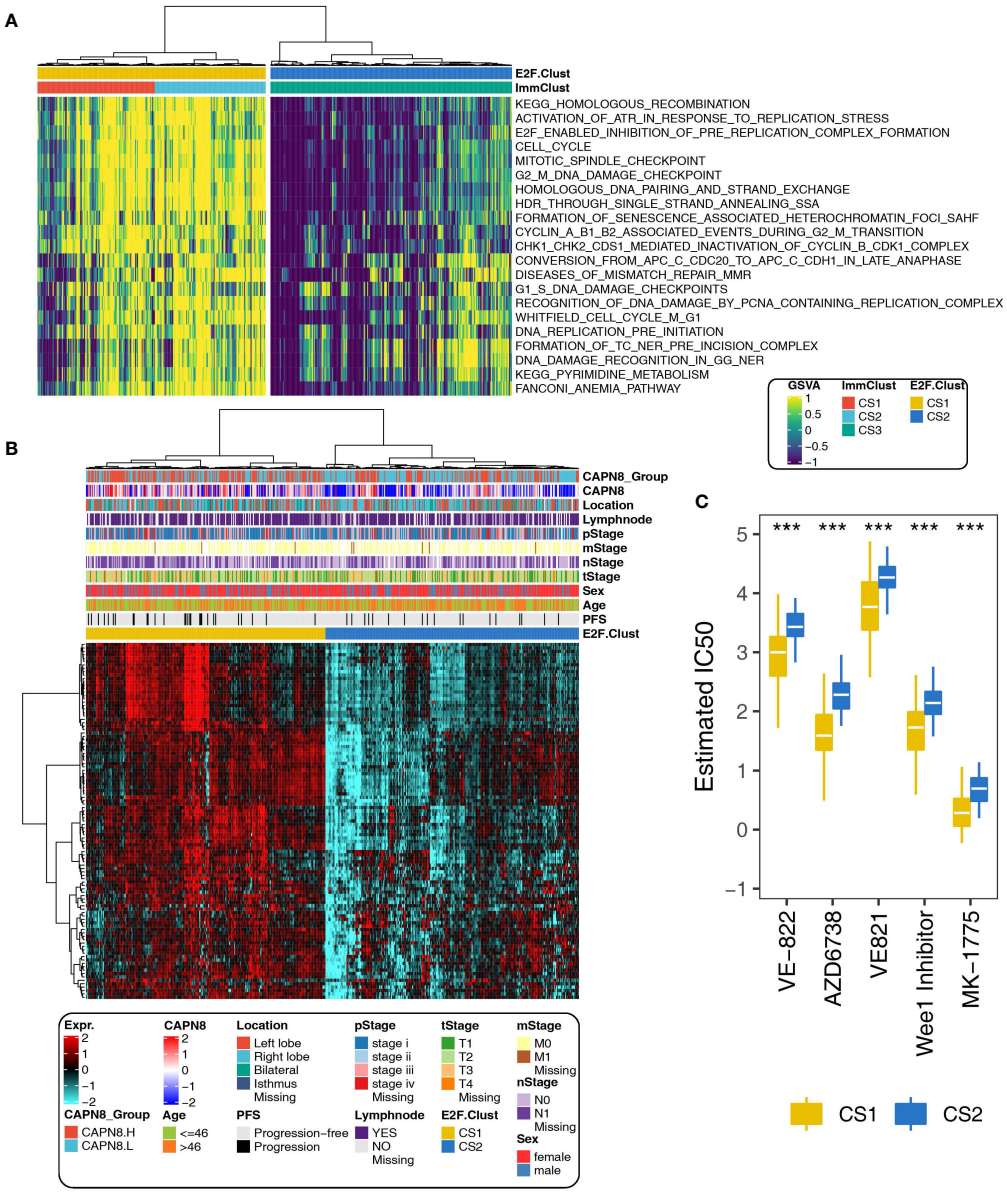


FIGURE 7
Different paradigms of DNA replication stress and drug sensitivity in the E2F-Clust and ImmClust. **(A)** Activities of 21 pathways related to DNA replication stress in the E2F-Clust and ImmClust. There was evident downregulation of DNA replication stress in the CS3 subtype. **(B)** The E2F-Clust stratified patients into two groups with distinct differences in clinical TNM stages and progression-free survival. **(C)** IC50 (half maximal inhibitory concentration) of typical inhibitors targeting ATR and week 1 protein. Patients of CS2 showed a tendency to be drug resistant. ***$P < 0.001$.

same TIME pattern in the TCGA-THCA cohort was still recognizable. CS3 remained to be the phenotype of immune desert with decreased infiltration of immune cells and expression of IRs. CS1 remained as the phenotype of immune exhausted for its increased infiltration of immune cells and IRs expression, while CS2 maintained the phenotype of inflamed TIME with elevated infiltration of neutrophils (Figure 8A).

The pattern of DNA replication stress, however, seems to be indistinguishable in the combined cohort. CS1, CS2, and CS3 were all seemingly accompanied by the increased activity of 21 DNA replication-related pathways (Figure 8A). Despite this uncertain result, CS2 was still much more likely to be drug resistant, with generally rising IC50 for ATR and week1 inhibitors (Figure 8C). Moreover, immunotherapy was plausible for the CS1 subtype in this combined cohort as patients in the CS1 subtype significantly responded to PD1-R treatment (Figure 8E).

Similar patterns of TIME, drug sensitivity, and immunotherapy response were still distinguishable in another

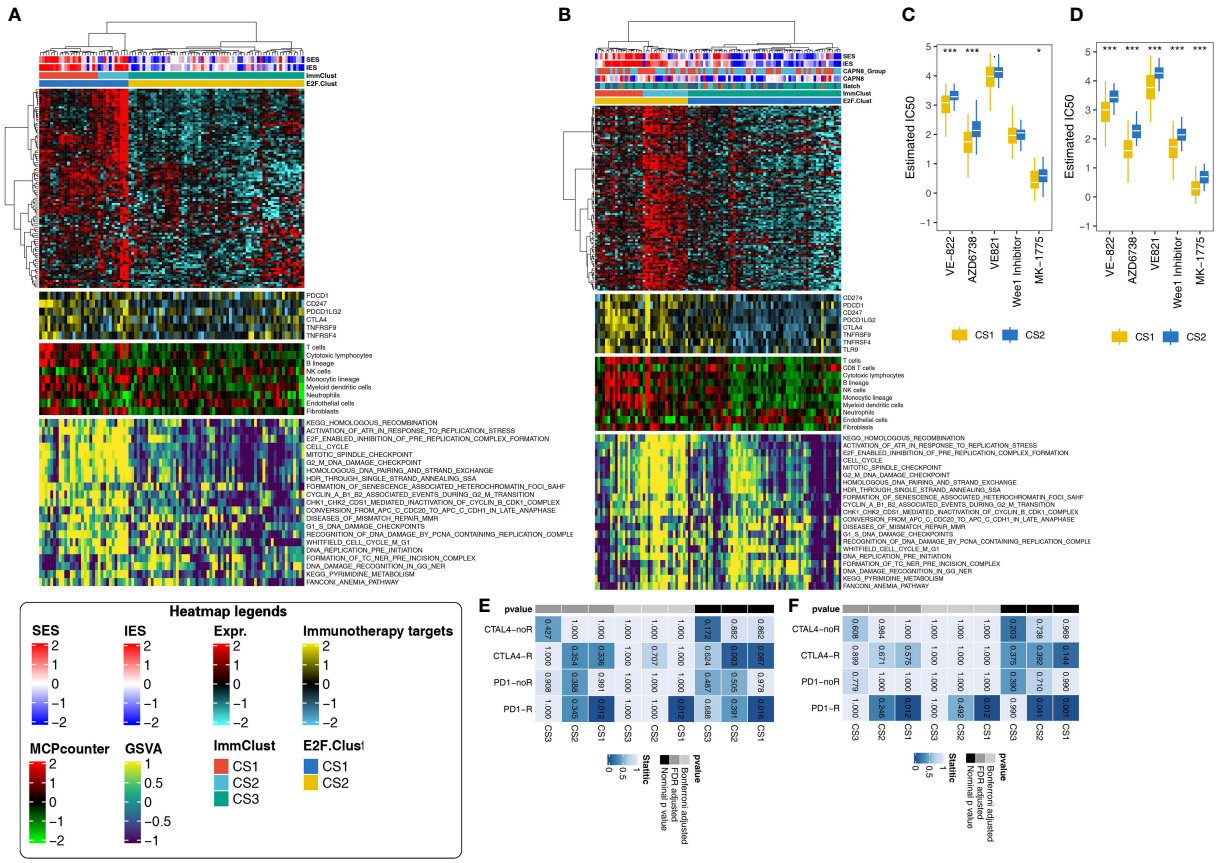


FIGURE 8
Application of the E2F-Clust and ImmClust in two external validation datasets (N = 78, N = 95). **(A)** The exhausted, inflamed, and desert TIME patterns remain recognizable after the application of E2F-Clust and ImmClust in the combined validation dataset (N = 78). **(B)** The exhausted, inflamed and desert TIME patterns are still distinct after applying the E2F-Clust and ImmClust in GSE27155 (N = 95). **(C)** CS2 still showed a tendency to be drug resistant in the combined validation dataset (N = 78). **(D)** Patients of CS2 are potentially resistant to common chemotherapeutics in GSE27155 (N = 95). **(E)** CS1 is prone to response to anti-PD-1 treatment in the combined validation dataset (N = 78). **(F)** Anti-PD-1 treatment is feasible in patients of CS1 subtype in GSE27155 (N = 95). *P < 0.05, ***P < 0.001.

validation cohort: GSE27155 cohort (n = 95), annotated by GPL96 platform. The immune exhausted inflammatory and immune desert phenotypes still corresponded to CS1, CS2, and CS3 subtypes, respectively (Figure 8B). The activities of 21 DNA replication stress-related pathways and drug sensitivity kept decreasing in the CS3 subtype (Figure 8D). The CS1 group demonstrated the same certain probability of benefiting from PD1-R treatment (Figure 8F).

Discussion

The present study identified three immune subtypes of THCA according to the expression of CAPN8 and related pathways, including the immune exhausted (CS1), inflamed (CS2), and immune desert (CS3) phenotypes. Three subclusters for THCA demonstrated quite diverse patterns of TIME, genetic variation, drug sensitivity, immunotherapy response, and patient prognosis—for instance, patients with CS1, with high expression of CAPN8, demonstrated rather detrimental survival outcomes when receiving chemotherapy but were much more likely to respond to anti-PD-1 treatment. These findings will provide novel insights for the treatment of THCA and help to understand the cross-talk between CAPN8 and the tumor immune microenvironment.

Firstly, CAPN8 was found to be a significant risk factor for THCA with a markedly elevated level of mRNA and protein in tumor tissues. This potential oncogene also induced the activation of EMT and E2F-targets, which are well-acknowledged pathways to promote cancer metastasis and proliferation. Consistent with our studies, CAPN8 was claimed to be a potential oncogene in gastric cancer, hepatic carcinoma, and lung cancer, causing the occurrence of precancerous lesions and cancer progression (5, 6, 28). In addition, members of CAPN family have been reported to facilitate the invasion of THCA by inducing MMP2 and MMP9 secretion, which can

contribute to extracellular matrix degradation during cancer cell migration (29).

Furthermore, three immune phenotypes of THCA were identified according to the expression of CAPN8 and related pathways. CS1, the immune exhausted subtype, was accompanied with a distinctly increased expression of IRs and proportion of infiltrating T cells. Accordingly, exhausted T cells lose their killing ability because of the increased expression of IRs but can be restored by immune checkpoint inhibitors (30). This exactly accords with our result: CS1 was found to be positively responsive to anti-PD-1 treatment.

In terms of CS3, this immune desert subtype of THCA demonstrated the absence of anti-tumor immunity and surging level of CNV. This is not surprising as there was a general downregulation of 21 pathways related to DNA replication stress in CS3. It is widely accepted that replication stress plays a key role in initiating anti-tumor immunity by inducing cancer-related neoantigens (31, 32), and its absence is intensively correlated with the immune desert phenotype. Consequently, the low immunity and high level of CNV can jointly contribute to strong cancer stemness (33), accounting for the result that patients of CS3 were tolerant of many ATR and week1 inhibitors in our study.

Overall, three immune subtypes of THCA were identified based on the expression of CAPN8 and related pathways in our study. The three types displayed rather different paradigms of TIME, immune therapy response, drug sensitivity, and genomic variance. Moreover, these three immune subtypes are highly coincident with the results of previous studies on cancer classification (34–38), where the exhausted, inflamed, and desert phenotypes of breast cancer, prostate cancer, and bladder cancer were characterized by using a similar clustering analysis.

Our study has several advantages. This is the first study to elucidate the role of CAPN8 in the metastasis of THCA from the aspects of TIME, DNA replication stress, and genetic alteration. Three immune subtypes identified in our study will provide new insights for the treatment of THCA, as different subtypes showed distinctly different responses to immunotherapy and chemotherapy. Most importantly, external validation in five individual cohorts made the extrapolation and robustness of the classification convincing.

There were also some limitations to the current study. Firstly, further *in vitro* experiments will make it more authentic for the existence of the three subtypes of THCA. Secondly, the prognostic effects of CAPN8 could be validated in actual cohorts of THCA to make it more persuasive. Lastly, analysis on cancer stemness can be added to further explain the relationship between different subtypes and various drug sensitivities.

In conclusion, we highlighted the role of CAPN8 in THCA metastasis and identified three distinct immune subtypes that can be distinguished in terms of prognosis, immunotherapeutic response, and drug sensitivity, which provide new insights for the treatment of THCA and contribute to the understanding of the interaction between CAPN8 and the tumor immune microenvironment.

Data availability statement

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

Author contributions

All authors contributed to the study's conception and design. XZ, SX, and QW performed data collection and analysis. XZ and SX wrote the manuscript. LP, FW, and TH polished and revised the manuscript. ZH and SN provided analytical ideas. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Research Project of Maternal and Child Health of Jiangsu Province (F201953) and the Science and Technology Project of Nantong (JC2020067) to ZH.

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/fimmu.2022.1013049/full#supplementary-material

References

1. Erler P, Keutgen XM, Crowley MJ, Zetoune T, Kundel A, Kleiman D, et al. Dicer expression and microRNA dysregulation associate with aggressive features in thyroid cancer. *Surgery* (2014) 156:1342–50. doi: 10.1016/j.surg.2014.08.007

2. Vasko VV, Saji M. Molecular mechanisms involved in differentiated thyroid cancer invasion and metastasis. *Curr Opin Oncol* (2007) 19:11–7. doi: 10.1097/CCO.0b013e328011ab86

3. Campbell RL, Davies PL. Structure-function relationships in calpains. *Biochem J* (2012) 447:335–51. doi: 10.1042/BJ20120921

4. Yoshikawa Y, Mukai H, Hino F, Asada K, Kato I. Isolation of two novel genes, down-regulated in gastric cancer. *Jpn J Cancer Res* (2000) 91:459–63. doi: 10.1111/j.1349-7006.2000.tb00967.x

5. Nian H, Ma B. Calpain-calpastatin system and cancer progression. *Biol Rev Camb Philos Soc* (2021) 96:961–75. doi: 10.1111/brv.12686

6. Chen B, Tang J, Guo YS, Li Y, Chen ZN, Jiang JL. Calpains are required for invasive and metastatic potentials of human HCC cells. *Cell Biol Int* (2013) 37:643–52. doi: 10.1002/cbin.10062

7. Storr SJ, Carragher NO, Frame MC, Parr T, Martin SG. The calpain system and cancer. *Nat Rev Cancer* (2011) 11:364–74. doi: 10.1038/nrc3050

8. Menicali E, Guzzetti M, Morelli S, Moretti S, Puxeddu E. Puxeddu EJFie, Immune Landscape Thyroid Cancers: New Insights *Front Endocrinol (Lausanne)* (2020) 11:637826. doi: 10.3389/fendo.2020.637826

9. Goldman M, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the xena platform. (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8

10. Deng M, Brägelmann J, Kryukov I, Saraiva-Agostinho N, Perner S. FirebrowseR: an r client to the broad institute's firehose pipeline. *Database J Biol Database Curation* (2017) 2017. doi: 10.1093/database/baw160

11. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095

12. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

13. Liu JZ, Hu YL, Feng Y, Guo YB, Liu YF, Yang JL, et al. Rafoxanide promotes apoptosis and autophagy of gastric cancer cells by suppressing PI3K/Akt/mTOR pathway. *Exp Cell Res* (2019) 385:111691. doi: 10.1016/j.yexcr.2019.111691

14. Robinson M, McCarthy D, Smyth G. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinf (Oxford England)* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

15. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7–. doi: 10.1186/1471-2105-14-7

16. Hanahan D. Hallmarks of cancer: New dimensions. *Cancer Discov* (2022) 12:31–46. doi: 10.1158/2159-8290.CD-21-1059

17. Lu X, Meng J, Zhu J, Zhou Y, Jiang L, Wang Y, et al. Prognosis stratification and personalized treatment in bladder cancer through a robust immune gene pair-based signature. *Clin Transl Med* (2021) 11(6):e453. doi: 10.1002/ctm2.453

18. Chen H, Boutros P. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* (2011) 12:35. doi: 10.1186/1471-2105-12-35

19. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17:218. doi: 10.1186/s13059-016-1070-5

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

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

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

23. McGranahan N, Furness A, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. (2016) 351:1463–9. doi: 10.1126/science.aaf1490

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

25. Lu X, Meng J, Zhou Y, Jiang L, Yan F. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics (Oxford, England)* (2020) btaa1018. doi: 10.1093/bioinformatics/btaa1018

26. Bertoli C, Herlihy A, Pennycook B, Kriston-Vizi J, de Bruin R. Sustained E2F-dependent transcription is a key mechanism to prevent replication-Stress-Induced DNA damage. *Cell Reports* (2016) 15(7):1412–22. doi: 10.1016/j.celrep.2016.04.036

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

28. Zhao Z, Sun Y, Hou N, Teng Y, Wang Y, Yang XJG. Capn8 promoter directs the expression of cre recombinase in gastric pit cells of transgenic mice. *Genesis* (2009) 47:674–9. doi: 10.1002/dvg.20552

29. Kalhori V, Törnquist KJM. MMP2 and MMP9 participate in S1P-induced invasion of follicular ML-1 thyroid cancer cells. *Mol Cell Endocrinol* (2015) 404:113–22. doi: 10.1016/j.mce.2015.01.037

30. Binnewies M, Roberts E, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* (2018) 24(30):541–50. doi: 10.1038/s41591-018-0014-x

31. Lin YL, Pasero P. Replication stress: from chromatin to immunity and beyond. *Curr Opin Genet Dev* (2021) 71:136–42. doi: 10.1016/j.gde.2021.08.004

32. Ye Z, Shi Y, Lees-Miller SP, Tainer TA. Function and molecular mechanism of the DNA damage response in immunity and cancer immunotherapy. *Front Immunol* (2021) 12:797880. doi: 10.3389/fimmu.2021.797880

33. Dai Z, Liu P. High copy number variations, particular transcription factors, and low immunity contribute to the stemness of prostate cancer cells. *J Transl Med* (2021) 19(1):206. doi: 10.1186/s12967-021-02870-x

34. Ren C, Li J, Zhou Y, Zhang S, Wang Q. Typical tumor immune microenvironment status determine prognosis in lung adenocarcinoma. *Transl Oncol* (2022) 18:101367. doi: 10.1016/j.tranon.2022.101367

35. Chen Z, Wang M, De Wilde R, Feng R, Su M, Roche Torres-de LA, et al. A machine learning model to predict the triple negative breast cancer immune subtype *Front Immunol* (2021) 12:749459. doi: 10.3389/fimmu.2021.749459

36. Zhang E, He J, Zhang H, Shan L, Wu H, Zhang M, et al. Immune-related gene-based novel subtypes to establish a model predicting the risk of prostate cancer. *Front Genet* (2020) 11:595657. doi: 10.3389/fgene.2020.595657

37. Shinwari K, Chen Z, Liu G, Chen L, Bolkov MA, Tuzankina IA, et al. Identification of the immune subtype among muscle-invasive bladder cancer patients by multiple datasets. (2022) 54:62–71.

38. Zhou Q, Yan X, Liu W, Yin W, Xu H, Cheng D, et al. Three immune-associated subtypes of diffuse glioma differ in immune infiltration, immune checkpoint molecules, and prognosis. *Front Oncol* (2020) 10:586019. doi: 10.3389/fonc.2020.586019

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fimmu.2022.984480

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Changjing Wang,
Hebei Medical University Third Affiliated Hospital, China
Sun Zhaoyang,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Mengle Peng
pengmengle1990@126.com
Lifeng Li
lilifeng0317@163.com

†These authors have contributed equally to this work

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

RECEIVED 02 July 2022
ACCEPTED 07 October 2022
PUBLISHED 28 October 2022

CITATION
Zheng X, Ma Y, Bai Y, Huang T, Lv X, Deng J, Wang Z, Lian W, Tong Y, Zhang X, Yue M, Zhang Y, Li L and Peng M (2022) Identification and validation of immunotherapy for four novel clusters of colorectal cancer based on the tumor microenvironment.
Front. Immunol. 13:984480.
doi: 10.3389/fimmu.2022.984480

Identification and validation of immunotherapy for four novel clusters of colorectal cancer based on the tumor microenvironment

Xiaoyong Zheng[1†], Yajie Ma[2†], Yan Bai[3†], Tao Huang[4], Xuefeng Lv[5], Jinhai Deng[6], Zhongquan Wang[7], Wenping Lian[7], Yalin Tong[8], Xinyu Zhang[2], Miaomiao Yue[1], Yan Zhang[1], Lifeng Li[4,9,10]* and Mengle Peng[7]*

[1]Department of Digestion, Henan Provincial Third People's Hospital, Zhengzhou, China, [2]Department of Medical Affair, Henan Provincial Third People's Hospital, Zhengzhou, China, [3]Department of Digestion, Zhengzhou First People's Hospital, Zhengzhou, China, [4]Medical School, Huanghe Science and Technology University, Zhengzhou, China, [5]Department of Clinical Laboratory, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [6]Richard Dimbleby Department of Cancer Research, Comprehensive Cancer Centre, Kings College London, London, United Kingdom, [7]Department of Clinical Laboratory, Henan Provincial Third People's Hospital, Zhengzhou, China, [8]Department of Digestion, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [9]Cancer Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [10]Internet Medical and System Applications of National Engineering Laboratory, Zhengzhou, China

The incidence and mortality of colorectal cancer (CRC) are increasing year by year. The accurate classification of CRC can realize the purpose of personalized and precise treatment for patients. The tumor microenvironment (TME) plays an important role in the malignant progression and immunotherapy of CRC. An in-depth understanding of the clusters based on the TME is of great significance for the discovery of new therapeutic targets for CRC. We extracted data on CRC, including gene expression profile, DNA methylation array, somatic mutations, clinicopathological information, and copy number variation (CNV), from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) (four datasets—GSE14333, GSE17538, GSE38832, and GSE39582), cBioPortal, and FireBrowse. The MCPcounter was utilized to quantify the abundance of 10 TME cells for CRC samples. Cluster repetitive analysis was based on the Hcluster function of the Pheatmap package in R. The ESTIMATE package was applied to compute immune and stromal scores for CRC patients. PCA analysis was used to remove batch effects among different datasets and transform genome-wide DNA methylation profiling into methylation of tumor-infiltrating lymphocyte (MeTIL). We evaluated the mutation differences of the clusters using MOVICS, DeconstructSigs, and GISTIC packages. As for therapy, TIDE and SubMap analyses were carried out to forecast the immunotherapy response of the clusters, and chemotherapeutic sensibility was estimated based on the pRRophetic package. All results were verified in the TCGA and GEO data. Four immune

clusters (ImmClust-CS1, ImmClust-CS2, ImmClust-CS3, and ImmClust-CS4) were identified for CRC. The four ImmClusts exhibited distinct TME compositions, cancer-associated fibroblasts (CAFs), functional orientation, and immune checkpoints. The highest immune, stromal, and MeTIL scores were observed in CS2, in contrast to the lowest scores in CS4. CS1 may respond to immunotherapy, while CS2 may respond to immunotherapy after anti-CAFs. Among the four ImmClusts, the top 15 markers with the highest mutation frequency were acquired, and CS1 had significantly lower CNA on the focal level than other subtypes. In addition, CS1 and CS2 patients had more stable chromosomes than CS3 and CS4. The most sensitive chemotherapeutic agents in these four ImmClusts were also found. IHC results revealed that CD29 stained significantly darker in the cancer samples, indicating that their CD29 was highly expressed in colon cancer. This work revealed the novel clusters based on TME for CRC, which would guide in predicting the prognosis, biological features, and appropriate treatment for patients with CRC.

KEYWORDS

colorectal cancer, tumor microenvironment, clusters, immunotherapy, bioinformatics

Introduction

Colorectal cancer (CRC) is a common malignant tumor in the digestive system. In recent years, the incidence of CRC is gradually increasing, and the mortality is also on the rise, ranking at the forefront of all malignant tumors, seriously endangering human health (1). The treatment of CRC is based on radical surgery, supplemented by chemotherapy, but nearly half of the patients are still trapped in tumor recurrence or metastasis without effective treatment (2). The traditional clinical and pathological predictors of CRC mainly include intestinal obstruction, pathological stage, level of cell differentiation, invaded vessels, invaded nerves, microsatellite status, etc. However, the final clinical significance is not very obvious. Recently, in order to more accurately predict the prognosis of patients with CRC, more and more researchers have begun to pay attention not only to tumor cells themselves but also to the tumor microenvironment (TME) of tumor cells.

The occurrence of CRC is a multistage mutation accumulation process involving multiple oncogenes, and the TME also plays an important role in the regulation (3) and drug resistance (4, 5) of CRC. The TME consists of a variety of cell types, including immune cells, inflammatory cells, adipocytes, fibroblasts, and vascular endothelial cells, as well as non-cellular components in and around the tumor (6). TME cells can be induced by tumor cells to produce a large number of cytokines and growth factors, thus forming a microenvironment conducive to the survival and proliferation of tumor cells. The TME can mediate the immune escape of tumor cells with the participation of tumor-associated immunosuppressive molecules (transforming growth factor-β, TGF-β), tumor-associated immunosuppressive cells (tumor-associated macrophages, TAMs), and tumor-associated proinflammatory responses (tumor-associated neutrophils) (7). The cellular components in the TME have become key modulators of tumor progression, organ-specific metastasis, and therapeutic response, among which tumor-infiltrating immune cells are the key to immunotherapy (8). Furthermore, tumor-infiltrating lymphocytes (TILs) can directly affect the prognosis and response to immunotherapy (9).

The heterogeneity of tumors is one of the important characteristics of tumors, which enables tumors to evolve various characteristics to adapt to the environment and even to resist the treatment of tumors (10, 11). Traditionally, tumors have been classified according to the type of cell or tissue they originate

Abbreviations: CRC, colorectal cancer; TME, tumor microenvironment; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; MeTIL, methylation of tumor-infiltrating lymphocyte; ImmClust, immune clusters; CAF, cancer-associated fibroblast; TGF-β, transforming growth factor-β; TAM, tumor-associated macrophage; TIL, tumor-infiltrating lymphocyte; CMS, consensus molecular subtype; MSI, microsatellite instability; FPKM, fragments per kilobase million; CNV, copy number variation; PCA, principal component analysis; ICB, immune checkpoint blockade; SCNA, somatic copy number variation; GDSC, Genomics of Drug Sensitivity in Cancer; EMT, epithelial–mesenchymal transition; IES, immune enrichment score; CRG, cancer-associated fibroblast-related gene; OS, overall survival; dM-MR, deficient mismatch repair; MSI-H, microsatellite instability-high; pMMR, proficiency of mismatch repair; MSS, microsatellite stability.

from and, thus, have a "one-size-fits-all" approach to pathology and treatment. It was not until sequencing became widely available that we realized that there were differences in genomic, transcriptome, and epigenetic characteristics within the same type of tumor (10). For example, the CRC Subtyping Consortium proposed a consensus molecular model, which divided CRC into four consensus molecular subtypes (CMS) according to pathological features (12). CMS1 is the type involved in microsatellite instability (MSI), also known as a high mutation type, which is manifested by mismatched gene repair changes. CMS2 is typical and is associated with abnormal activation of WNT or MYC signaling pathways. CMS3 is a metabolic type, showing a high mutation degree of KRAS and metabolic disorder. CMS4 indicates an abnormal activation of the TGF-β signaling pathway. Even within the same tumor, its genetic characteristics are different between subcellular populations and change dynamically as the tumor develops (11, 13). The understanding of tumor heterogeneity has led to a more detailed classification of tumors, and the development of different treatment regimens based on the molecular characteristics of tumors has improved the therapeutic outcomes of multiple tumors. For example, imatinib is used to treat BCR-ABL tyrosine kinase constitutively activated chronic myeloid leukemia (14), HER2 protein-targeting drugs are used to treat HER2-positive breast cancer (15), and estrogen antagonists are used to treat estrogen receptor-positive breast cancer (16).

In this study, we integrated TME cells of CRC to identify four immune clusters (ImmClust-CS1, ImmClust-CS2, ImmClust-CS3, and ImmClust-CS4), which were validated using data from the Gene Expression Omnibus (GEO) datasets. We described each according to their biological characteristics, including the prognosis, immune status, somatic mutations, copy number variation (CNV), and response to treatment.

Materials and methods

Public data acquisition and preprocessing

The RNA-seq FPKM (fragments per kilobase million) data of TCGA-COAD and TCGA-READ were downloaded from the UCSC Xena platform (https://xenabrowser.net/) (17). After primary tumor selection, a total of 390 COAD and 154 READ samples were included in our study. The FPKM style of RNA-seq data was normalized into TPM value (18). The 450K DNA methylation array was also extracted from the UCSC Xena platform (19). The somatic mutation data and the clinicopathological information of patients with COAD or READ were obtained from the cBioPortal platform (http://www.cbioportal.org/datasets) (20). The data on CNV were acquired from FireBrowse (http://firebrowse.org/) (21). Four external independent datasets, namely, GSE14333, GSE17538, GSE38832, and GSE39582, were downloaded from the GEO database and quantitated by Affymetrix Human Genome U133 Plus 2.0 Array (22–25). TCGA and GEO data were combined to remove batch effects by ComBat in R package SVA (26), and the removal of batch effects was tested by principal component analysis (PCA) (27).

TME abundance quantification and immune cluster establishment

MCPcounter is an R package that quantifies the absolute abundance of eight immune cells (B-cell lineage, $CD8^+$ T cells, cytotoxic lymphocytes, monocytic lineage, myeloid dendritic cells, natural killer cells, neutrophils, and T cells) and two stromal cells (fibroblasts and endothelial cells) using transcriptome data (28). We utilized MCPcounter to quantify the abundance of the 10 TME cells for CRC samples. After cluster repetitive analysis based on Hcluster function of Pheatmap package in R (29), four immune clusters (ImmClust-CS1, ImmClust-CS2, ImmClust-CS3, and ImmClust-CS4) were identified. The ESTIMATE R package was applied to compute immune scores and stromal scores (30), representing the enrichment scores for CRC patients.

Immunotherapy response analysis

Based on the ESTIMATE tool, the present study used gene expression data from CRC to estimate stromal and immune cells in cancer tissue to predict the immune score and stromal score in CRC (30). TILs were associated with the clinical outcomes of CRC (31, 32). To further evaluate the local tumor immune response of the four ImmClusts, genome-wide DNA methylation profiling was applied and transformed into methylation of tumor-infiltrating lymphocyte (MeTIL) using PCA analysis (33). TIDE is a computational method for predicting immune checkpoint blockade (ICB) responses (34). Based on RNA expression profiles, TIDE prediction scores were calculated to forecast the likelihood of CRC patients responding to immunotherapy. A lower TIDE score indicated a lower possibility of immune escape (34). In addition, SubMap analysis was carried out to contrast gene expression similarity between ImmClusts and the responders of anti-PD-1 or anti-CTLA-4 therapy (35–38).

Assessment of cancer-associated fibroblasts

Cancer-associated fibroblasts were reported to play an essential role in the TME of CRC (38). Since cancer-associated fibroblast (CAF) may have modeled different patient subpopulations, CAF-related genes and signatures were

mapped to ImmClusts. Previous studies provided seven CAF-related genes, namely, *ACTA2*, *PDGFRA*, *PDGFRB*, *THY1*, *COL1A1*, *FAP*, and *PDPN* (39–43). We obtained eight CAF-related signatures (ecm-myCAF, detox-iCAF, IL-iCAF, TGFB-myCAF, wound-myCAF, IFNG-iCAF, CAF-S1, and normal fibroblast) from the study based on single-cell analysis (44).

Evaluation of mutation differences

The MOVICS package is designed for multi-omics comprehensive clustering and visualization of cancer clusters, which provides a unified interface and standardizes the output for 10 algorithms (CIMLR, iClusterBayes, MoCluster, COCA, ConsensusClustering, IntNMF, LRAcluster, NEMO, PINSPlus, and SNF) (45). The ImmClusts of the TCGA-CRC cohorts were comprehensively characterized by the MOVICS package, including somatic mutation (46), tumor mutational burden (TMB), and fraction genome altered score. The package DeconstructSigs can put 96 mutation spectrums into 30 corresponding mutation signatures of the COSMIC database (47). Mutations mediated by the apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) family are widespread in human cancers (48). APOBEC has been reported to be associated with immunotherapy response (49). We selected two APOBEC-related signatures and weighted them to obtain the weight of the APOBEC mutant signature. Somatic CNV (SCNA) may affect as many as thousands of genes simultaneously, but the selective advantage that drives variation may be mediated by only one or a few of these genes. Based on the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm (50), we compared the chromosomal instability of subtypes.

Prediction of the sensibility of chemotherapeutics

A wide range of drug screening can be performed through the Genomics of Drug Sensitivity in Cancer (GDSC) website (51). Based on pRRophetic package in R (52, 53), Ridge's regression was used to construct a prediction model between drug sensitivity and the expression profile of cell lines. Whereafter, we applied the aforementioned model to estimate the half-maximal inhibitory concentration (IC_{50}) value of chemotherapeutics for CRC patients.

Human tissue sample collection

Formalin-fixed and paraffin-embedded (PPFE) specimens were collected from cancerous and paracancerous tissues of CRC patients in the First Affiliated Hospital of Zhengzhou University. All the samples were stored at room temperature (20°C–25°C). According to pathological features, at least two pathologists diagnosed all the specimens and reached an agreement. Lastly, this study included 10 cases of CRC cancerous and paracancerous tissues. The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (Ethics No. 2021-KY-0147-002).

Immunohistochemical staining

According to the manufacturers' protocol, immunohistochemistry (IHC) staining of CRC cancerous and paracancerous tissues was performed. First, we dewaxed, hydrated, and blocked the paraffin-embedded sections and then incubated them overnight at 4°C with a CD29 antibody (Affinity, China). The next day, sections were washed three times with PBS and then incubated with secondary antibodies at 37°C. Again, sections were washed with PBS, dropped into a DAB reagent, and restained with hematoxylin for 2 min. Finally, these sections were visualized by light microscopy, and the results of Masson staining and IHC were analyzed using ImageJ software.

Results

Immune-related cluster establishment

If the sample dataset collection time, collection institutions, sequencing platform, and other factors are different, they may automatically form different batches, thus affecting the real data. Therefore, batch effects should be checked and removed before subsequent analysis; otherwise, all subsequent analysis results will be invalid. Figure 1A shows the PCA diagram of the TCGA-COAD and TCGA-READ data before and after batch removal, indicating that batch effect removal was relatively successful.

We developed an immune-related cluster using the Hcluster function of the Pheatmap package, and four immune clusters (ImmClust-CS1, ImmClust-CS2, ImmClust-CS3, and ImmClust-CS4) were identified for CRC (Figure 1B). The heatmap showed the differences in the distribution of the four ImmClusts in clinicopathological features, TME compositions, functional orientation, and immune checkpoints (Figure 1B). The four ImmClusts exhibited distinct TME compositions. ImmClust-CS1 was characterized by a high enrichment of immune cells and low fibroblasts (Figure 1B). ImmClust-CS2 was dominated by immune-cell-related genes, as well as endothelial cells and fibroblasts (Figure 1B). ImmClust-CS3 and ImmClust-CS4 were both characterized by immune low and fibroblast high and fibroblast low, respectively (Figure 1B). As for the functional orientation (immunosuppression, T-cell activation, T-cell survival, regulatory T cells, major histocompatibility complex class I, myeloid cell chemotaxis, and tertiary lymphoid

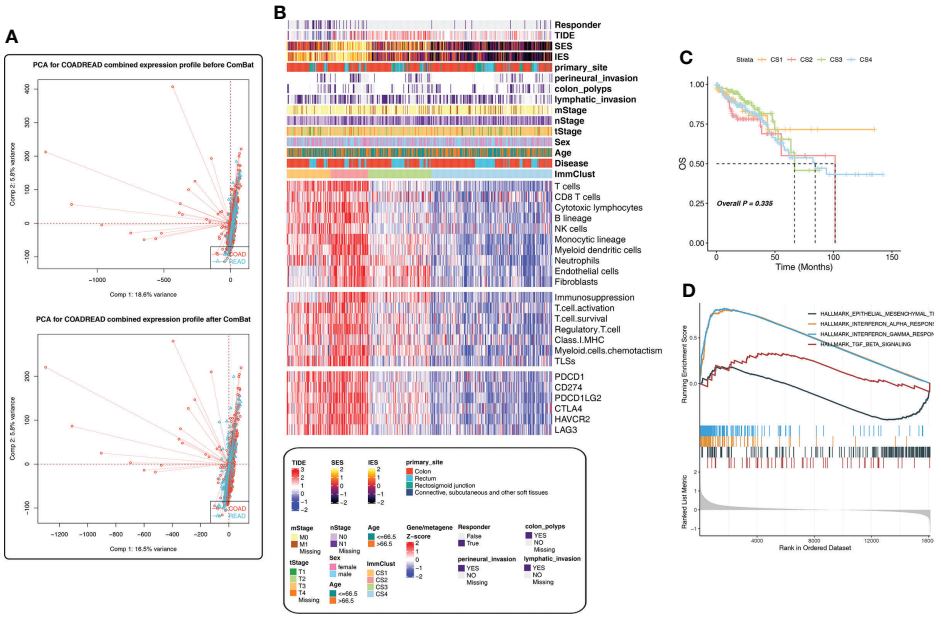


FIGURE 1
Immune-related cluster establishment for The Cancer Genome Atlas (TCGA). **(A)** Batch effect removal of the TCGA-COAD and TCGA-READ cohorts using PCA analysis. **(B)** The heatmap showed the differences in the distribution of the four ImmClusts in clinicopathological features, TME compositions, functional orientation, immune checkpoints, and so on. **(C)** K-M survival curves showed the differences of overall survival and recurrence rate among the four ImmClusts. **(D)** GSEA analysis comparing the differences of enrichment pathways between CS1 and the other ImmClusts.

structures), on the whole, the expression values of related genes were relatively high in ImmClust-CS1 and ImmClust-CS2 and relatively low in ImmClust-CS3 and ImmClust-CS4 (Figure 1B). The expression of immune checkpoint genes was consistent with the above findings (Figure 1B).

Furthermore, we compared the clinical outcomes of patients in the four ImmClusts. Although the survival patterns of the four ImmClusts exhibited were not statistically significant ($P = 0.335$), ImmClust-CS1 did not reach the median survival period in 10 years and had a better prognosis than other ImmClusts (Figure 1C). Hence, we compared the differences in enrichment pathways between CS1 and the other ImmClusts. GSEA analysis showed that compared with the other three ImmClusts, CS1 was enriched in interferon-alpha response, interferon-gamma response, and TGF beta signaling pathways, while the epithelial–mesenchymal transition (EMT) pathway was downregulated, indicating that CS1 was enriched in immune-related pathways but downregulated in the EMT pathway (Figure 1D).

Immunotherapy response analysis

ESTIMATE is an algorithm that uses transcription profiles of cancer samples to estimate the number of tumor cells, as well as the number of infiltrated immune and stromal cells. Among the ESTIMATE scores of the four ImmClusts, the highest immune (Figure 2A) and stromal scores (Figure 2B) were observed in CS2, in contrast to the lowest scores in CS4 (Figures 2A, B). A study has shown that the MeTIL score system may assess immune and immunotherapy responses in CRC (54). Interestingly, our data suggested that CS2 had the highest MeTIL score (Figure 2C), while CS4 had the lowest MeTIL score, which further suggested the differences in immunotherapy responses among the four ImmClusts. A lower TIDE score indicated a lower possibility of immune escape. From Figures 2D, E, we found that patients in CS1 were more likely to respond to immunotherapy (Fisher's exact test, $P < 0.001$). In mapping CAF-related genes and signatures to ImmClusts, we found a highly positive correlation between CS2 and CAFs (Figure 2F). At the same time, the immune gene expression profiles of CS1 and CS2 were found to be similar to those of melanoma patients responding to anti-PD-1 therapy using SubMap analysis (Figure 2G). To sum up, CS1 may respond to immunotherapy, while CS2 may respond to immunotherapy after anti-CAFs.

Verification analysis based on GEO data

To validate the results of the above analysis, four external CRC cohorts from the GEO database were included in the follow-up study. First, the data of the GEO cohorts were

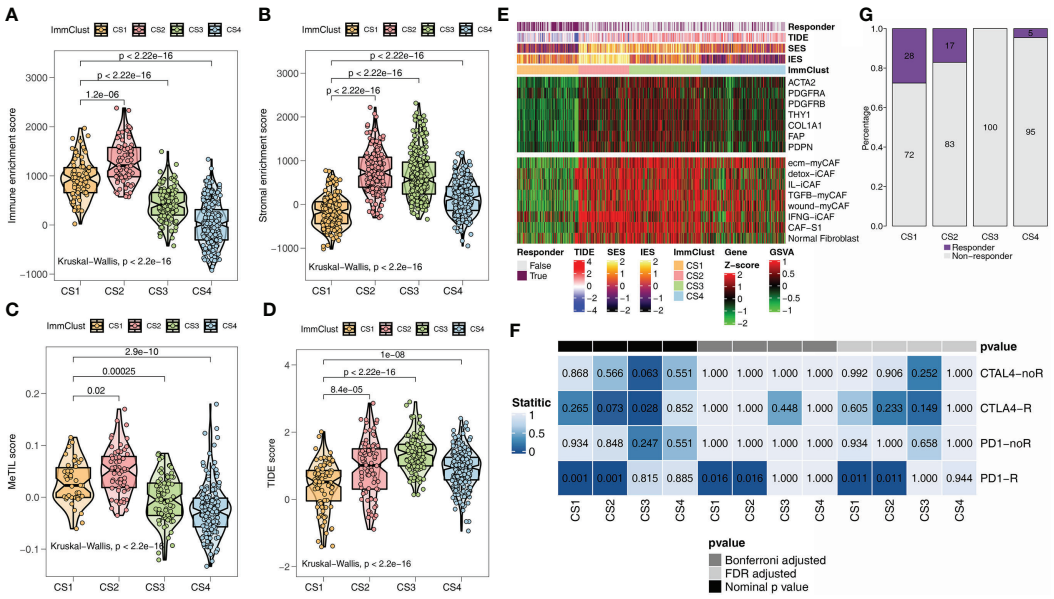


FIGURE 2
Immunotherapy response analysis for the TCGA. **(A)** Immune enrichment score for the four ImmClusts. **(B)** Stromal enrichment score for the four ImmClusts. **(C)** MeTIL score for the four ImmClusts. **(D)** TIDE score for the four ImmClusts. **(E)** Response to immunotherapy of the four ImmClusts. **(F)** The heatmap showed the differences in the distribution of the four ImmClusts in cancer-associated fibroblast (CAF)-related genes and signatures. **(G)** SubMap analysis for the four ImmClusts.

combined and PCA analysis was performed. The batch effect removal was observed successfully (Figure 3A). Subsequently, considering the large cohort size and uneven distribution of tumor purity in GEO, we removed the samples with a tumor purity larger than 0.8 (the higher the tumor purity, the less accurate the TME estimate), leaving 833 CRC samples. Here, four subtypes were obtained by unsupervised clustering, and the distribution of subtypes in clinicopathological features, TME compositions, functional orientation, and immune checkpoints was consistent with the TCGA cohort (Figure 3B). Happily, for survival analysis, we found statistically significant differences in survival curves, and CS1 and CS2 subgroups with high levels of immune cell infiltration fared better, while CS3 with lower immune cell infiltration but higher fibroblast infiltration had a poor prognosis (Figure 3D). Conclusively, GSEA analysis showed that CS1 was associated with interferon-alpha response, interferon-gamma response, TGF beta signaling, and epithelial–mesenchymal transition pathways (Figure 3C), which was consistent with the previous findings.

In addition, the same results as the above findings were confirmed. The immune (Figure 4A) and stromal (Figure 4B) scores were the highest in CS2 and the lowest in CS4. CS1 with the lowest TIDE score (Figure 4C) remained the subgroup most likely to respond to immunotherapy (Figure 4D). The heatmap revealed that CS2 was positively correlated with CAF-related genes and signatures (Figure 4F). SubMap analysis uncovered that CS1 and anti-CAF-CS2 may respond to anti-PD-1 immunotherapy (Figure 4E).

Evaluation of mutation differences

The distribution variations of the somatic mutations among the four ImmClusts were also analyzed based on MOVICS. The top 15 markers with the highest mutation frequency were PIK3CA, FAT4, FAT3, DNAH5, NEB, PCLO, HMCN1, AHNAK2, PCDH15, CACNA1E, DNAH8, ATM, VPS13B, DNAH2, and KMT2B (Figure 5A). TMB and TiTv were calculated by MOVICS, and it was found that CS1 had a higher TMB (Figure 5B). As for APOBEC, mutation weights were significantly different among the four ImmClusts, with CS3 having the highest and CS1 the lowest (Figure 5C). In addition, we found a significant negative correlation between immune enrichment score (IES) and APOBEC mutation weight ($R = -0.12$, $P = 0.012$, Figure 5D), while APOBEC mutation weight was positively correlated with TIDE score ($R = 0.13$, $P = 0.0065$, Figure 5E), suggesting that APOBEC is involved in immunotherapy response.

From the Manhattan plot, we can see CNV at the chromosomal level, which was computed by the GISTIC

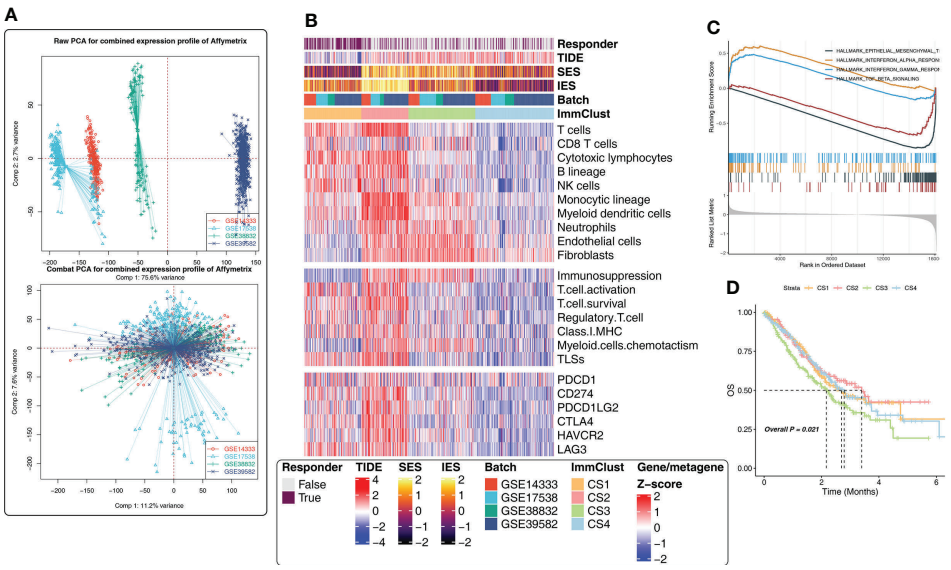


FIGURE 3
Immune-related cluster establishment for the Gene Expression Omnibus (GEO). **(A)** Batch effect removal of four external independent datasets using PCA analysis. **(B)** The heatmap showed the differences in the distribution of the four ImmClusts in clinicopathological features, TME compositions, functional orientation, immune checkpoints, and so on. **(C)** GSEA analysis comparing the differences of enrichment pathways between CS1 and the other ImmClusts. **(D)** K–M survival curves showed the differences of overall survival and recurrence rate among the four ImmClusts.

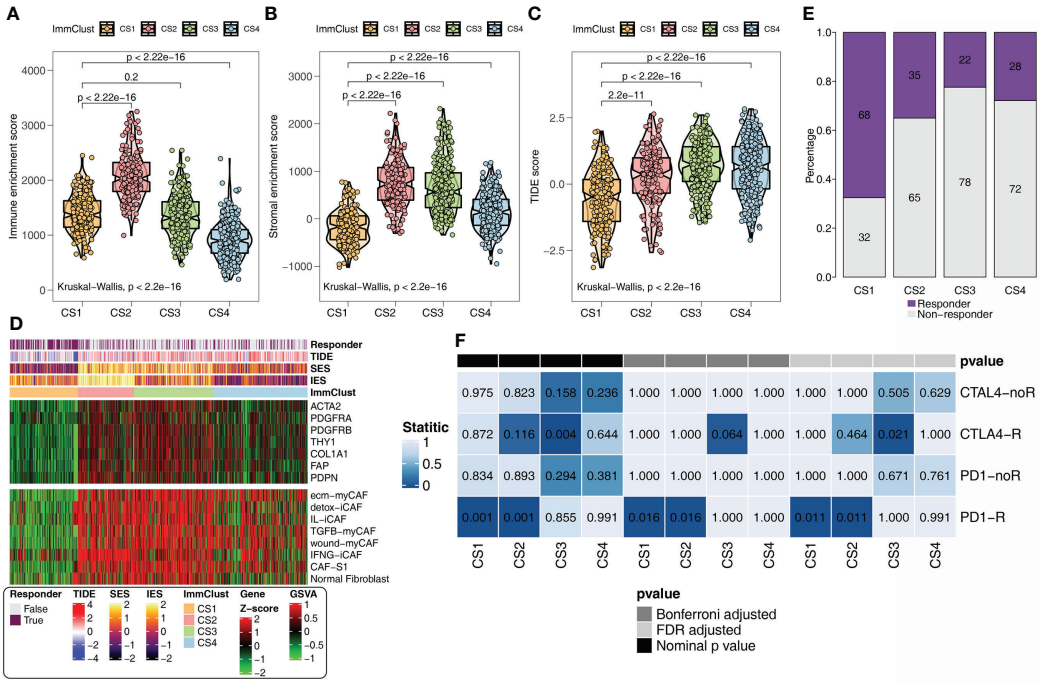


FIGURE 4
Immunotherapy response analysis for the GEO. **(A)** Immune enrichment score for the four ImmClusts. **(B)** Stromal enrichment score for the four ImmClusts. **(C)** TIDE score for the four ImmClusts. **(D)** Response to immunotherapy of the four ImmClusts. **(E)** SubMap analysis for the four ImmClusts. **(F)** The heatmap showed the differences in the distribution of the four ImmClusts in CAF-related genes and signatures.

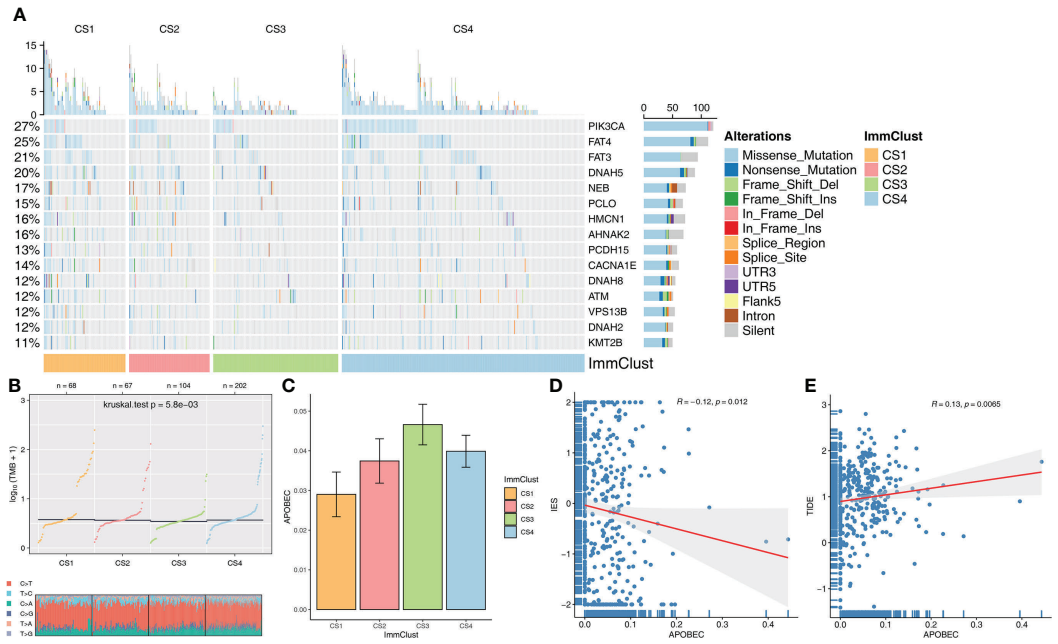


FIGURE 5
Evaluation of mutation differences among the four ImmClusts. **(A)** The waterfall plot of somatic mutation features established with ImmClusts. **(B)** TMB and distribution of TiTv calculated by MOVICS for the four ImmClusts. **(C)** Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) mutation weights of the four ImmClusts. **(D)** Correlation between immune enrichment score (IES) and APOBEC mutation weight. **(E)** Correlation between TIDE score and APOBEC mutation weight.

algorithm (Figure 6A). By counting copy number amplification (Figure 6B) and deletion (Figure 6C), respectively, we found that CS1 had a significantly lower CNA at the focal level than other subtypes. Fraction genome-altered scores (threshold 0.2) were calculated using MOVICS packages to characterize chromosomal instability. The results showed that CS1 and CS2 patients had more stable chromosomes than CS3 and CS4 (Figure 6D).

Univariate Cox regression analysis for CAF-related genes

Since CAF plays an important role in shaping the ImmClusts of CRC, 289 CAF-related genes (CRGs) were extracted from the literature and analyzed by univariate Cox regression analysis. In the TCGA-CRC cohort, 49 risky CRGs and 8 protective CRGs were identified. In the GEO-CRC dataset, 56 risky CRGs and 13 protective CRGs were found. We selected 25 intersection genes to map the forest plot. As can be seen from Figure 7, there were 20 risky CRGs (*SERP2*, *CILP*, *GRP*, *COMP*, *C7*, *SNAI1*, *LAMP5*, *TGFB3*, *OLFM2*, *GAS1*, *IGF1*, *CYP1B1*, *PRICKLE1*, *ZFHX4*, *UST*, *CD36*, *EBF2*, *PCOLCE2*, *PLIN4*, and *STEAP4*) and 5 protective CRGs (*CEBPA*, *PID1*, *CD177*, *DNASE1L3*, *HRCT1*) in the intersection genes (Figure 7).

Prediction of the sensibility of chemotherapeutics

To further investigate the treatment strategies for the four ImmClusts of CRC, we conducted a prediction of the sensibility of chemotherapeutics to evaluate the IC_{50} value using the pRRophetic package. The IC_{50} value can be used to measure the ability of a drug to induce apoptosis, that is, the higher the inducing ability, the lower the value, and of course, it can also be used to reverse the tolerance of a certain cell to the drug. We screened out drugs that showed consistent sensitivity in the TCGA (Figure 8A) and GEO (Figure 8B) databases for display. Compared with the other three ImmClusts, patients in CS1 were most sensitive to metformin, epothilone B, and VX-680; patients in CS2 were most sensitive to DMOG, AICAR, AZD7762, temsirolimus, TW.37, and elesclomol; patients in CS3 were most sensitive to MG.132, A.770041, and cyclopamine; and patients in CS4 were most sensitive to lapatinib.

Validation of the protein expression levels of CD29 using IHC

A comparison of cancer and paracancer IHC results revealed that CD29 stained significantly darker in the cancer samples,

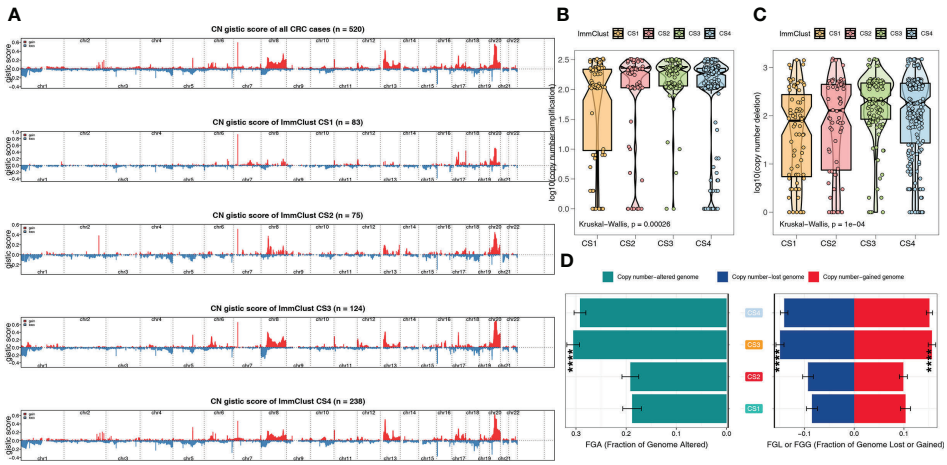


FIGURE 6
The chromosomal instability of the four ImmClusts. **(A)** Manhattan plot showing the CNV at the chromosomal level. **(B)** Copy number amplification of the four ImmClusts. **(C)** Copy number deletion of the four ImmClusts. **(D)** Fraction genome-altered scores of the four ImmClusts. ****p < 0.001.

indicating that CD29 was highly expressed in colon cancer (Figure 9). These results are consistent with our subtype results, indicating the reliability and reproducibility of the classification.

Discussion

CRC is the third most common malignant tumor in the world and the second leading cause of cancer death worldwide, with about 1.8 million new cases and 800,000 deaths every year (55, 56). The tumor microenvironment plays an important role in tumor genesis, development, and tumor immunity. Studies have shown that the infiltration and metastasis of immune cells, memory $CD4^+$ T cells, and $CD8^+$ T cells in the TME can regulate tumor immunity and participate in the three stages of tumor cell clearance, tumor and body balance, and tumor immune escape (57). They not only inhibit tumor growth but also screen tumors for hypoimmunogenicity, leading to tumor escape (58).

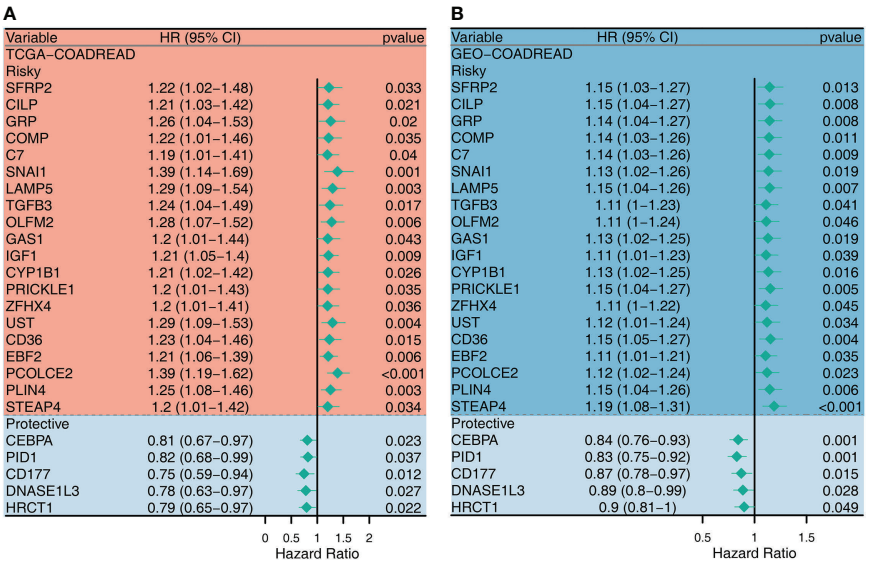


FIGURE 7
Univariate Cox regression analysis for the CAF-related genes in TCGA **(A)** and GEO **(B)** datasets.

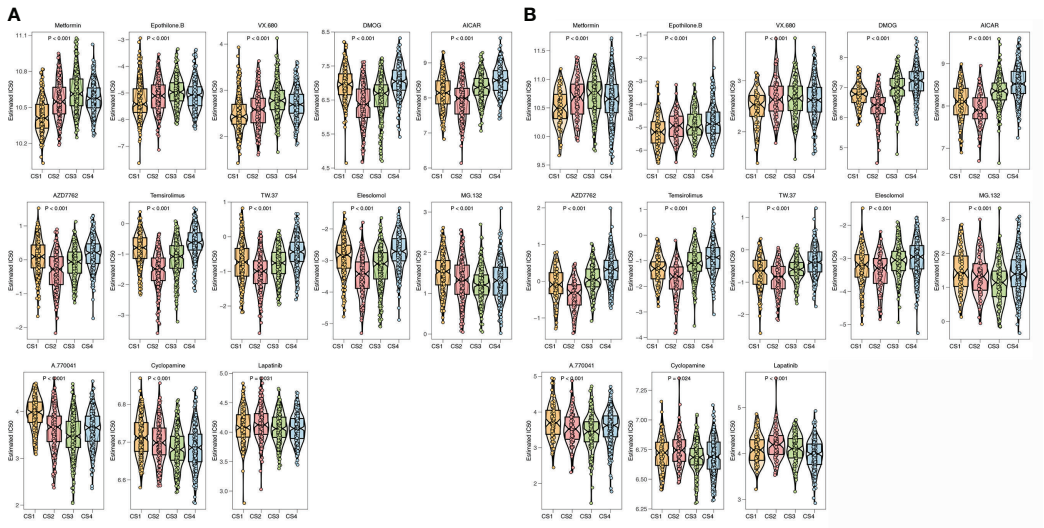


FIGURE 8
Prediction of chemotherapeutic sensibility. **(A)** Prediction of chemotherapeutic sensibility for the TCGA. **(B)** Prediction of chemotherapeutic sensibility for the GEO.

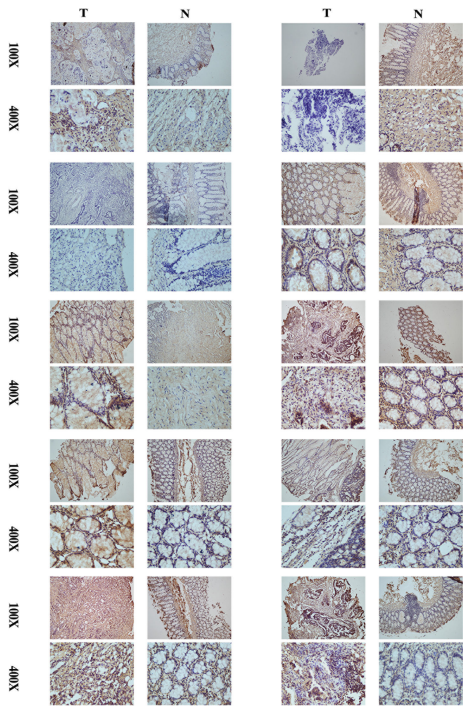


FIGURE 9
Comparison of cancer and paracancer IHC results revealed that CD29 stained significantly darker in the cancer samples, indicating that their CD29 was highly expressed in colon cancer.

Our data revealed that CS1 was enriched in the interferon-alpha (IFN-α) response and interferon-gamma (IFN-γ) response signaling pathways, while it was downregulated in the EMT pathway. EMT is the process by which cells lose epithelial features and gain mesenchymal properties, such as increased motility of tumor cells (59). EMT processes are a series of changes and transformations resulting from environmental stimuli, depending on the organization and signal transduction environment (60). In CRC, EMT is strongly associated with aggressive or metastatic phenotypes (61). In-depth research on the mechanism of action and interaction of the related signaling pathways in CRC EMT will help to develop more new methods for CRC treatment and promote individualized treatment. IFN-α can inhibit the proliferation of tumor cells through the adaptive and innate immune system (62). IFN-α can also activate important components of the immune system such as $CD8^+$ T cells and NK cells and promote the mature differentiation of B cells and DC cells (62). NK cells and $CD8^+$ T cells can inhibit tumor cell metastasis (63); hence, IFN-α can regulate the immune system to play an antitumor role. The secretion of TNF-α, IFN-γ, IL-6, and other cytokines by $CD4^+$ T cells can change the TME (64, 65), induce the local invasion of T lymphocytes into the tumor (66), inhibit the synthesis of DNA and RNA of tumor cells, and thus, induce the apoptosis of tumor cells (65).

In order to further improve the overall survival (OS) rate of CRC, immunotherapy has gradually attracted our attention. For example, the gradual discovery of immune checkpoints such as PD-1, PD-L1, CTLA-4, and OX40 has led to the emergence of immune checkpoint inhibitors for CRC therapy (67). There are

significant differences in treatment outcomes among different subtypes of CRC. For example, immune checkpoint inhibitors represented by the anti-PD-1/PD-L1 pathway have achieved the most outstanding curative effect in the treatment of CRC with deficient mismatch repair (dM-MR) or microsatellite instability-high (MSI-H) and have been used for the second-line treatment (68). However, most metastatic CRC patients represented by the proficiency of mismatch repair (pMMR) or microsatellite stability (MSS) not only do not respond to the above treatments; moreover, it also leads to the progression of the disease, and some patients even have adverse events related to immunotherapy such as colitis, pneumonia, dermatitis, and endocrine diseases after receiving the treatment (69, 70). At the same time, the TME is also closely related to the effect of immunotherapy. A study has shown that the accumulation of memory B cells and T cells in the TME can not only determine the clinical stage of CRC but also indicate the effect of the immune system's antitumor response (71). TIL is characterized by tumor invasion and lymph node metastasis. As a signal of tumor cells attacking the human immune system, TIL reflects the immune response of the host (72, 73). The antitumor effect of TIL can be affected by co-inhibitory immune checkpoints and can be used as a biomarker to evaluate and predict the effect of immune checkpoint inhibitors (74, 75). CAFs that are abundant in CRC and accumulate substantially in the TME are usually associated with poor prognosis (76, 77). CAFs are positively related to PD-L1 expression in CRC tissues, and by secreting CXCL5, CAFs could promote the expression of PD-L1 in cancer cells (78).

In general, immunotherapy has opened a new chapter in cancer treatment and greatly improved the prognosis of CRC, but the therapeutic effect varies greatly among different subtypes of CRC. Therefore, it is very important to distinguish sensitive and insensitive populations by specific biomarkers. In the era of precision cancer therapy, the CRC typing system we identified has great potential to be used to predict and evaluate the effects of immunotherapy on CRC patients. In this study, we integrated the TME cells of CRC to identify four immune clusters (ImmClust-CS1, ImmClust-CS2, ImmClust-CS3, and ImmClust-CS4), which were validated using data from the GEO datasets. The four ImmClusts exhibited distinct TME compositions, CAFs, functional orientation, and immune checkpoints. The highest immune, stromal, and MeTIL scores were observed in CS2, in contrast to the lowest scores in CS4. CS1 may respond to immunotherapy, while CS2 may respond to immunotherapy after anti-CAFs.

Cancer develops through the gradual acquisition of somatic genetic changes, including point mutations, CNV, and fusion events that affect the function of key genes that regulate cell growth and survival. The occurrence of CRC is the cumulative result of a series of gene mutations. CRC not only mutates in well-known tumor-related genes (such as *APC*, *TP53*, *KRAS*, *PIK3CA*, *SMAD4*) but also mutates in other genes, including *SMAD2*, *CTNNB1*, *FAM123B*, *SOX9*, *ARID1A*, etc. (79, 80). The latest research showed that mutations of TP53, APC, KRAS, BRAF, and ATM cover most patients with CRC (81, 82). The top 15 markers with the highest mutation frequency among the four ImmClusts we constructed were PIK3CA, FAT4, FAT3, DNAH5, NEB, PCLO, HMCN1, AHNAK2, PCDH15, CACNA1E, DNAH8, ATM, VPS13B, DNAH2, and KMT2B. These mutation markers may serve as novel molecular targets for the detection or therapy of these ImmClusts. When every cell divides, it acquires random somatic mutations, and only driver mutations lead to malignant development. PIK3CA was previously defined as a driving mutation in CRC (46). Recent studies have found that mutation of PIK3CA can lead to continuous activation of the EGFR signaling pathway, thus affecting the therapeutic efficiency of anti-EGFR drugs (83). Ejima et al. found high-frequency mutations in ATM introns in CRC cell lines (84). Genes with different mutation frequencies are expected to be markers for the detection or treatment of subtypes. Among the four ImmClusts, the top 15 markers with the highest mutation frequency were acquired, and CS1 had significantly lower CNA at the focal level than other subtypes. In addition, CS1 and CS2 patients had more stable chromosomes than CS3 and CS4. The most sensitive chemotherapeutic agents in these four ImmClusts were also found.

To sum up, the high immune infiltration, low fibroblast infiltration, high mutation load, and low chromosomal variation of CS1 are related to the ability of this subtype to respond to immunotherapy.

Our study was the first one to establish the cluster system based on TME for CRC. The combined analysis of data from the TCGA and GEO verified the accuracy of the classification system. Nevertheless, the clustering system constructed by us lacks large prospective studies to verify, and its specificity and sensitivity need to be further determined.

Conclusions

This work revealed the novel clusters based on the TME for CRC, which would guide in predicting the prognosis, biological features, and appropriate treatment for patients with CRC.

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

This study was reviewed and approved by The study was approved by the ethics committee of the First Affiliated Hospital

of Zhengzhou University (Ethics No. 2021-KY-0147-002). The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Collaborative Innovation Major Project of Zhengzhou (Grant No. 20XTZX08017), the Scientific and Technological Project of Henan Province (No. 2018020578), the National Natural Science Foundation of China (Grant No. 82002433), the Science and Technology Project of Henan Provincial Department of Education (Grant No. 21A320036), and the Young and Middle-aged Health Science and Technology Innovation Talents in 2020 (Grant No. YXKC2020049).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.984480/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654

2. Weitz J, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. *Lancet* (2005) 365:153–65. doi: 10.1016/S0140-6736(05)17706-X

3. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct Target Ther* (2020) 5:22. doi: 10.1038/s41392-020-0116-z

4. Woolston A, Khan K, Spain G, Barber LJ, Griffiths B, Gonzalez-Exposito R, et al. Genomic and transcriptomic determinants of therapy resistance and immune landscape evolution during anti-EGFR treatment in colorectal cancer. *Cancer Cell* (2019) 36:35–50. doi: 10.1016/j.ccell.2019.05.013

5. Giordano G, Remo A, Porras A, Pancione M. Immune resistance and EGFR antagonists in colorectal cancer. *Cancers (Basel)* (2019) 11(8):1089. doi: 10.3390/cancers11081089

6. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res* (2019) 79:4557–66. doi: 10.1158/0008-5472.CAN-18-3962

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

8. Schulz M, Salamero-Boix A, Niesel K, Alekseeva T, Sevenich L. Microenvironmental regulation of tumor progression and therapeutic response in brain metastasis. *Front Immunol* (2019) 10:1713. doi: 10.3389/fimmu.2019.01713

9. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* (2018) 564:268–72. doi: 10.1038/s41586-018-0694-x

10. Palucka AK, Coussens LM. The basis of oncoimmunology. *Cell* (2016) 164:1233–47. doi: 10.1016/j.cell.2016.01.049

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

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

13. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LJ, Kinzler KW. Cancer genome landscapes. *Science* (2013) 339:1546–58. doi: 10.1126/science.1235122

14. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med* (2006) 355:2408–17. doi: 10.1056/NEJMoa062867

15. Shepard HM, Lewis GD, Sarup JC, Fendly BM, Maneval D, Mordenti J, et al. Monoclonal antibody therapy of human cancer: taking the HER2 protooncogene to the clinic. *J Clin Immunol* (1991) 11:117–27. doi: 10.1007/BF00918679

16. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S, et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci U.S.A.* (2012) 109:2724–9. doi: 10.1073/pnas.1018854108

17. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data *via* the xena platform. *Nat Biotechnol* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8

18. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* (2020) 26:903–9. doi: 10.1261/rna.074922.120

19. Mallona I, Sierco A, Peinado MA. The pancancer DNA methylation trackhub: A window to the cancer genome atlas epigenomics data. *Methods Mol Biol (Clifton N.J.)* (2018) 1766:123–35. doi: 10.1007/978-1-4939-7768-0_7

20. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* (2013) 6:l1. doi: 10.1126/scisignal.2004088

21. Chabanais J, Labrousse F, Chaunavel A, Germot A, Maftah A. POFUT1 as a promising novel biomarker of colorectal cancer. *Cancers* (2018) 10(11):411. doi: 10.3390/cancers10110411

22. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage b and c colorectal cancer. *Clin Cancer Res* (2009) 15:7642–51. doi: 10.1158/1078-0432.CCR-09-1431

23. Williams CS, Bernard JK, Demory Beckler M, Almohazey D, Washington MK, Smith JJ, et al. ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. *Carcinogenesis* (2015) 36:710–8. doi: 10.1093/carcin/bgv049

24. Tripathi MK, Deane NG, Zhu J, An H, Mima S, Wang X, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res* (2014) 74:6947–57. doi: 10.1158/0008-5472.CAN-14-1592

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

26. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

27. Nyamundanda G, Poudel P, Patil Y, Sadanandam AA. Novel statistical method to diagnose, quantify and correct batch effects in genomic studies. *Sci Rep* (2017) 7:10849. doi: 10.1038/s41598-017-11110-6

28. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17:218. doi: 10.1186/s13059-016-1070-5

29. Diao C, Xi Y, Xiao T. Identification and analysis of key genes in osteosarcoma using bioinformatics. *Oncol Lett* (2018) 15:2789–94. doi: 10.3892/ol.2017.7649

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

31. Li Y, Liang L, Dai W, Cai G, Xu Y, Li X, et al. Prognostic impact of programed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor infiltrating lymphocytes in colorectal cancer. *Mol Cancer* (2016) 15:55. doi: 10.1186/s12943-016-0539-x

32. Bai Z, Zhou Y, Ye Z, Xiong J, Lan H, Wang F. Tumor-infiltrating lymphocytes in colorectal cancer: The fundamental indication and application on immunotherapy. *Front Immunol* (2021) 12:808964. doi: 10.3389/fimmu.2021.808964

33. Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA Methylation-based immune response signature improves patient diagnosis in multiple cancers. *J Clin Invest* (2017) 127:3090–102. doi: 10.1172/JCI91095

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

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

36. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* (2016) 351:1463–9. doi: 10.1126/science.aaf1490

37. Lu X, Jiang L, Zhang L, Zhu Y, Hu W, Wang J, et al. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *Neoplasia* (2019) 21:591–601. doi: 10.1016/j.neo.2019.04.003

38. Kobayashi H, Gieniec KA, Lannagan T, Wang T, Asai N, Mizutani Y, et al. The origin and contribution of cancer-associated fibroblasts in colorectal carcinogenesis. *Gastroenterology* (2022) 162:890–906. doi: 10.1053/j.gastro.2021.11.037

39. Zheng S, Zou Y, Tang Y, Yang A, Liang JY, Wu L, et al. Landscape of cancer-associated fibroblasts identifies the secreted biglycan as a protumor and immunosuppressive factor in triple-negative breast cancer. *Oncoimmunology* (2022) 11:2020984. doi: 10.1080/2162402X.2021.2020984

40. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* (2016) 16:582–98. doi: 10.1038/nrc.2016.73

41. Kawase A, Ishii G, Nagai K, Ito T, Nagano T, Murata Y, et al. Podoplanin expression by cancer associated fibroblasts predicts poor prognosis of lung adenocarcinoma. *Int J Cancer* (2008) 123:1053–9. doi: 10.1002/ijc.23611

42. Kisselbach L, Merges M, Bossie A, Boyd A. CD90 expression on human primary cells and elimination of contaminating fibroblasts from cell cultures. *Cytotechnology* (2009) 59:31–44. doi: 10.1007/s10616-009-9190-3

43. Madsen DH, Leonard D, Masedunskas A, Moyer A, Jurgensen HJ, Peters DE, et al. M2-like macrophages are responsible for collagen degradation through a mannose receptor-mediated pathway. *J Cell Biol* (2013) 202:951–66. doi: 10.1083/jcb.201301081

44. Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discovery* (2020) 10:1330–51. doi: 10.1158/2159-8290.CD-19-1384

45. Lu X, Meng J, Zhou Y, Jiang L, Yan F. MOVICS: An r package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics* (2020) 36(22-23):5539–41. doi: 10.1093/bioinformatics/btaa1018

46. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* (2018) 173:371–85. doi: 10.1016/j.cell.2018.02.060

47. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* (2016) 17:31. doi: 10.1186/s13059-016-0893-4

48. Petljak M, Maciejowski J. Molecular origins of APOBEC-associated mutations in cancer. *DNA Repair (Amst)* (2020) 94:102905. doi: 10.1016/j.dnarep.2020.102905

49. Wang S, Jia M, He Z, Liu XS. APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene* (2018) 37:3924–36. doi: 10.1038/s41388-018-0245-9

50. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 12:R41. doi: 10.1186/gb-2011-12-4-r41

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

52. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol* (2014) 15:R47. doi: 10.1186/gb-2014-15-3-r47

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

54. Zou Q, Wang X, Ren D, Hu B, Tang G, Zhang Y, et al. DNA Methylation-based signature of CD8+ tumor-infiltrating lymphocytes enables evaluation of immune response and prognosis in colorectal cancer. *J Immunother Cancer* (2021) 9:e002671. doi: 10.1136/jitc-2021-002671

55. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* (2019) 16:713–32. doi: 10.1038/s41575-019-0189-8

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

57. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. *Annu Rev Immunol* (2004) 22:329–60. doi: 10.1146/annurev.immunol.22.012703.104803

58. Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, et al. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* (2001) 410:1107–11. doi: 10.1038/35074122

59. Tamaki T, Shimizu T, Niki M, Shimizu M, Nishizawa T, Nomura S. Immunohistochemical analysis of NANOG expression and epithelial-mesenchymal transition in pulmonary sarcomatoid carcinoma. *Oncol Lett* (2017) 13:3695–702. doi: 10.3892/ol.2017.5864

60. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* (2013) 13:97–110. doi: 10.1038/nrc3447

61. Nieto MA. The ins and outs of the epithelial to mesenchymal transition in health and disease. *Annu Rev Cell Dev Biol* (2011) 27:347–76. doi: 10.1146/annurev-cellbio-092910-154036

62. Ferrantini M, Capone I, Belardelli F. Interferon-alpha and cancer: mechanisms of action and new perspectives of clinical use. *Biochimie* (2007) 89:884–93. doi: 10.1016/j.biochi.2007.04.006

63. Farhood B, Najafi M, Mortezaee K. CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: A review. *J Cell Physiol* (2019) 234:8509–21. doi: 10.1002/jcp.27782

64. Yi DH, Stetter N, Jakobsen K, Jonsson R, Appel S. 3-day monocyte-derived dendritic cells stimulated with a combination of OK432, TLR7/8 ligand, and prostaglandin E2 are a promising alternative for cancer immunotherapy. *Cancer Immunol Immunother* (2018) 67:1611–20. doi: 10.1007/s00262-018-2216-y

65. Koya T, Yanagisawa R, Higuchi Y, Sano K, Shimodaira S. Interferon-alpha-inducible dendritic cells matured with OK-432 exhibit TRAIL and fas ligand pathway-mediated killer activity. *Sci Rep* (2017) 7:42145. doi: 10.1038/srep42145

66. Li L, Wang W, Pan H, Ma G, Shi X, Xie H, et al. Microwave ablation combined with OK-432 induces Th1-type response and specific antitumor immunity in a murine model of breast cancer. *J Transl Med* (2017) 15:23. doi: 10.1186/s12967-017-1124-9

67. Payandeh Z, Khalili S, Somi MH, Mard-Soltani M, Baghbanzadeh A, Hajiasgharzadeh K, et al. PD-1/PD-L1-dependent immune response in colorectal cancer. *J Cell Physiol* (2020) 235:5461–75. doi: 10.1002/jcp.29494

68. Kather JN, Halama N, Jaeger D. Genomics and emerging biomarkers for immunotherapy of colorectal cancer. *Semin Cancer Biol* (2018) 52:189–97. doi: 10.1016/j.semcancer.2018.02.010

69. Park JJ, Arafath S, Kumar ST, Sharma R, Dixit D. Managing toxicities associated with immune checkpoint inhibitors. *JAAPA Off J Am Acad Physician Assistants* (2021) 34:32–9. doi: 10.1097/01.JAA.0000735760.65235.3c

70. Winer A, Bodor JN, Borghaei H. Identifying and managing the adverse effects of immune checkpoint blockade. *J Thorac Dis* (2018) 10:S480–9. doi: 10.21037/jtd.2018.01.111

71. Picard E, Verschoor CP, Ma GW, Pawelec G. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal cancer. *Front Immunol* (2020) 11:369. doi: 10.3389/fimmu.2020.00369

72. Lin B, Du L, Li H, Zhu X, Cui L, Li X. Tumor-infiltrating lymphocytes: Warriors fight against tumors powerfully. *BioMed Pharmacother* (2020) 132:110873. doi: 10.1016/j.biopha.2020.110873

73. Maibach F, Sadozai H, Seyed Jafari SM, Hunger RE, Schenk M. Tumor-infiltrating lymphocytes and their prognostic value in cutaneous melanoma. *Front Immunol* (2020) 11:2105. doi: 10.3389/fimmu.2020.02105

74. Maby P, Tougeron D, Hamieh M, Mlecnik B, Kora H, Bindea G, et al. Correlation between density of CD8+ T-cell infiltrate in microsatellite unstable colorectal cancers and frameshift mutations: A rationale for personalized immunotherapy. *Cancer Res* (2015) 75:3446–55. doi: 10.1158/0008-5472.CAN-14-3051

75. Ma Y, Ou J, Lin T, Chen L, Wang J, Qiao D, et al. Phenotypic analysis of tumor-infiltrating lymphocytes from non-small cell lung cancer and their potential application for adoptive cell therapy. *Immunopharmacol Immunotoxicol* (2020) 42:319–29. doi: 10.1080/08923973.2020.1765375

76. Powell DW, Adegboyega PA, Di Mari JF, Mifflin RC. Epithelial cells and their neighbors i. role of intestinal myofibroblasts in development, repair, and cancer. *Am J Physiol Gastrointestinal Liver Physiol* (2005) 289:G2–7. doi: 10.1152/ajpgi.00075.2005

77. Deng L, Jiang N, Zeng J, Wang Y, Cui H. The versatile roles of cancer-associated fibroblasts in colorectal cancer and therapeutic implications. *Front Cell Dev Biol* (2021) 9:733270. doi: 10.3389/fcell.2021.733270

78. Li Z, Zhou J, Zhang J, Li S, Wang H, Du J. Cancer-associated fibroblasts promote PD-L1 expression in mice cancer cells *via* secreting CXCL5. *Int J Cancer* (2019) 145:1946–57. doi: 10.1002/ijc.32278

79. Goto T, Marusawa H, Chiba T. Landscape of genetic aberrations detected in human colorectal cancers. *Gastroenterology* (2013) 145:686–8. doi: 10.1053/j.gastro.2013.07.029

80. Tsilimigras DI, Ntanasis-Stathopoulos I, Bagante F, Moris D, Cloyd J, Spartalis E, et al. Clinical significance and prognostic relevance of KRAS, BRAF, PI3K and TP53 genetic mutation analysis for resectable and unresectable colorectal liver metastases: A systematic review of the current evidence. *Surg Oncol* (2018) 27:280–8. doi: 10.1016/j.suronc.2018.05.012

81. Zhuang Y, Wang H, Jiang D, Li Y, Feng L, Tian C, et al. Multi gene mutation signatures in colorectal cancer patients: predict for the diagnosis, pathological classification, staging and prognosis. *BMC Cancer* (2021) 21:380. doi: 10.1186/s12885-021-08108-9

82. Naccarati A, Polakova V, Pardini B, Vodickova L, Hemminki K, Kumar R, et al. Mutations and polymorphisms in TP53 gene–an overview on the role in colorectal cancer. *Mutagenesis* (2012) 27:211–8. doi: 10.1093/mutage/ger067

83. Hsu HC, Thiam TK, Lu YJ, Yeh CY, Tsai WS, You JF, et al. Mutations of KRAS/NRAS/BRAF predict cetuximab resistance in metastatic colorectal cancer patients. *Oncotarget* (2016) 7:22257–70. doi: 10.18632/oncotarget.8076

84. Ejima Y, Yang L, Sasaki MS. Aberrant splicing of the ATM gene associated with shortening of the intronic mononucleotide tract in human colon tumor cell lines: a novel mutation target of microsatellite instability. *Int J Cancer* (2000) 86:262–8. doi: 10.1002/(sici)1097-0215(20000415)86:2<262::aid-ijc17>3.0.co;2-i

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 09 November 2022
DOI 10.3389/fimmu.2022.1051063

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Yuzhen Gao,
Zhejiang University, China
Qi Zhang,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Minhong Cai
26857276@qq.com
Guanghui Du
690587873@qq.com

†These authors have contributed equally to this work

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

RECEIVED 22 September 2022
ACCEPTED 17 October 2022
PUBLISHED 09 November 2022

CITATION
Wang J, He X, Bai Y, Du G and Cai M (2022) Identification and validation of novel biomarkers affecting bladder cancer immunotherapy *via* machine learning and its association with M2 macrophages.
Front. Immunol. 13:1051063.
doi: 10.3389/fimmu.2022.1051063

Identification and validation of novel biomarkers affecting bladder cancer immunotherapy *via* machine learning and its association with M2 macrophages

Junkang Wang[1†], Xiaojuan He[1†], Yifeng Bai[2], Guanghui Du[1*] and Minhong Cai[3*]

[1]Department of Outpatient, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, Sichuan, China, [2]Department of Cancer Center, Sichuan Academy of Medical Sciences and Sichuan People's Hospital, Chengdu, Sichuan, China, [3]Healthcare-associated Infection Management Office, Sichuan Academy of Medical Sciences and Sichuan People's Hospital, Chengdu, Sichuan, China

Background: Immunotherapy has shown promising results in bladder cancer therapy options.

Methods: Analysis of open-access data was conducted using the R software. Open-access data were obtained from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and IMvigor210 databases. Immunofluorescence and co-culture systems were utilized to validate the effect of PTHLH on M2 macrophage polarization.

Results: Here, through the combined (TCGA, GSE128959, GSE13507, and GSE83586) and IMvigor210 cohorts, we comprehensively investigated the biological and immune microenvironment differences in patients with diverse immunotherapy responses. Meanwhile, we found that M2 macrophage could affect bladder cancer immunotherapy sensibility. Moreover, based on the machine learning algorithm (LASSO logistics regression), PTHLH, BHMT2, and NGFR were identified, which all have good prediction abilities for patient immunotherapy. Then, a logistics regression model was established based on PTHLH, BHMT2, and NGFR, and each patient was assigned a logistics score. Subsequently, we investigated the difference in patients with high low logistics scores, including biological enrichment, immune microenvironment, and genomic characteristics. Meanwhile, data from the Human Protein Atlas database indicated a higher protein level of PTHLH in bladder cancer tissue. Immunofluorescence indicated that the knockdown of PTHLH in bladder cancer cells can significantly inhibit the M2 polarization of co-culture M0 macrophages.

10.3389/fimmu.2022.1051063

Conclusions: Our study investigated the difference between bladder cancer immunotherapy responders and non-responders. Meanwhile, the PTHLH was identified as a novel biomarker for bladder cancer immunotherapy.

KEYWORDS

immunotherapy, PTHLH, machine learning, M2 macrophages, bladder cancer

Introduction

Bladder cancer is the leading malignancy in the urogenital system and is responsible for a serious health concern globally (1). Bladder cancer usually occurs as a result of several factors such as gender, genetic differences, and lifestyle (1). In bladder cancer, there are two clinical phenotypes: non-muscle-invasive bladder cancer (NIMBC) and muscle-invasive bladder cancer (MIBC). Generally, surgery can provide long-term therapeutic benefits for most NIMBC and high-level MIBC patients. Also, MIBC tends to suffer from worse survival and more limited treatment options compared to NIMBC (2). Clinically, bladder cancer patients often benefit from immunotherapy. Nonetheless, immunotherapy is still ineffective for a substantial number of people, leading to poor outcomes (3).

There is a long history of immunotherapy in bladder cancer treatment. In 1976, researchers found that the intravesical instillation of the BCG vaccine can kill bladder cancer cells by inducing a local immune response (4). However, for decades, bladder cancer management has remained relatively unchanged and the high recurrence rate remains a challenge (5). Simultaneously, the microenvironment in bladder cancer is always immunosuppressive. Bladder cancer often has a high infiltration level of Treg cells and is regulated by multiple cytokines (6). Consequently, studies are underway to identify new targets for immunotherapy for bladder cancer. Shi et al. found that the mutagenesis mediated by APOBEC can effectively indicate the survival and immunotherapy of bladder cancer (7). Groeneveld et al. revealed that CXCL13, a marker of tertiary lymphoid structures, can effectively indicate the survival of advanced bladder cancer patients receiving immunotherapy (8). Yi et al. found that IGFBP7 is associated with immunological characteristics and might be an immunotherapy target for bladder cancer (9). Same as other cancers, bladder cancer uses immune checkpoints to regulate immunity, notably PD-1/L1 and CTLA4. Nowadays, some clinical trials that target immune checkpoints have achieved some promising results in bladder cancer (10). Meanwhile, the combination of chemotherapy and immunotherapy can effectively improve the antitumor effect in bladder cancer (11). Therefore, the exploration of the novel biomarkers involved in bladder cancer immunotherapy is meaningful.

Here, based on the open-access data, we comprehensively investigated the biological and immune microenvironment differences in patients with diverse immunotherapy responses. Meanwhile, we found that M2 macrophage could affect the sensitivity of bladder cancer patients to immunotherapy. Moreover, PTHLH, BHMT2, and NGFR were identified, which all have good prediction abilities on patient immunotherapy. Then, a logistics regression model was established based on PTHLH, BHMT2, and NGFR. Next, we explored the difference between patients with high and low logistics scores, including biological enrichment, immune microenvironment, and genomic characteristics. Immunofluorescence showed that the knockdown of PTHLH in bladder cancer cells can significantly inhibit the M2 polarization of co-culture M0 macrophages, making it a potential biomarker for bladder cancer.

Methods

Data collection

Open-access data were obtained from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and IMvigor210 databases. As for the patients in TCGA, the expression profile was in "STAR-Counts" form and clinical data were in "bcr-xml" form. Before analysis, all the data were pre-processed. For the dataset in GEO, the criteria "1. Sample counts > 150; 2. Complete expression profile; 3. Complete survival information" were used for data filtering. Ultimately, GSE128959, GSE13507, and GSE83586 were identified and annotated according to their platforms. The sva package was utilized for data combination. Levels of tumor mutational burden (TMB) and microsatellite instability (MSI) were extracted from the genomic data of the TCGA database. The IMvigor210 dataset was downloaded from http://research-pub.gene.com/IMvigor210CoreBiologies. Baseline information of the included samples is shown in Supplementary Tables S1-S4.

Immunotherapy evaluation

Response of patients on immunotherapy was conducted using the Tumor Immune Dysfunction and Exclusion (TIDE) analysis (12). The input file was the expression profile, and

according to this, each patient owns a TIDE score. The immunotherapy responders were those whose TIDE score > 0; otherwise, non-responders.

Machine learning algorithm

The machine learning algorithm, LASSO logistic regression, was used for optimal variable selection (13). For the identified characteristic genes, the glm function in R software was used for logistics model construction. The "family" was set as "binomial".

Biological enrichment

Biological enrichment difference was identified using the Gene Set Enrichment Analysis (GSEA), whose gene sets for reference were "Hallmark" and "c2.cp.kegg.v2022.1.Hs.symbols" (14). Implementation of gene ontology (GO) analysis was based on the clusterprofiler package (15). The terms of adjusted *p*-value < 0.05 was considered statistically significant.

Immune microenvironment exploration

CIBERSORT was utilized to quantify the infiltration level of 22 immune cells based on the input transcriptional profile data (16). Calculation of the immune score and the stromal score was conducted using the Estimate packages in R software.

Genomic difference

The index reflecting tumor stemness characteristics named mRNAsi was extracted from the previous study (17). The frequency and somatic copy number alteration (SCNA) level difference of human chromosomes was extracted from the TCGA database.

Immunohistochemistry

Evaluation of protein levels of PTHLH, BHMT2, and NGFR was conducted using representative immunohistochemistry images (normal bladder and bladder cancer sections) obtained from the Human Protein Atlas (HPA) database.

Cell culture and transfection

Cell lines T24 and THP-1 were laboratory stock and cultured under routine culture conditions. Lipofectamine 2000 was selected for cell transfection following the standard steps. The following were the sequences of shRNAs used: shRNA1: 5'-CCCTGATTGTGC CATAAAT-3'; shRNA2: 5'-GGCCAGAACAATGAAGAAA-3'; shRNA3: 5'-GGCACTTAGAAGAACCAAT-3'.

Co-culture system and THP-1 differentiation

THP-1 cells were seeded into six-well plates for 24 h (100 ng/ml PMA) to differentiate into adherent M0 macrophages. The transfected T24 cell lines and differentiated M0 macrophages together constituted a co-culture system. The culture supernatant of transfected T24 cells was collected and added to the M0 macrophages. After that, for 48 h, macrophages were collected for further assay.

Immunofluorescence

The collected M0 macrophages were firstly cultured on a glass slide and then fixed with 4% PFA at room temperature. Afterwards, 0.05% Triton X-100 was used for cell permeabilization (2 min) and 5% BSA was used for cell blocking (1 h). Next, cells were incubated with the following specific antibodies overnight at 4°C: anti-F4/80 antibody (1:200) and anti-CD206 (1:200) antibody. DAPI was utilized for cell nuclear staining. Fluorescence microscopy was used to visualize immunofluorescence.

Statistical analysis

Calculation of statistical significance in the different analysis was conducted using the R and GraphPad Prism software. The threshold of statistical significance was 0.05. According to the data distribution, Student's *t*-test and Mann–Whitney *U* test were tested.

Results

Data preparation

The whole flowchart is shown in Figure S1. Firstly, the independent bladder cancer cohorts, TCGA, GSE128959, GSE13507, and GSE83586, were selected for our analysis, which showed a high batch effect (Figure 1A; Comp 1: 84.2% variance, Comp 2: 6.4% variance). Through the sva package, we effectively decreased the batch difference and completed data combination (Figure 1B; Comp 1: 10% variance, Comp 2: 6.6% variance). Furthermore, TIDE analysis was utilized to assess the immunotherapy response (Figure 1C). In the combined cohort, the TIDE score was utilized to divide patients into immunotherapy responder and non-responder groups (Figure 1D). Meanwhile, in the real immunotherapy IMvigor210 cohort, the bladder patients

with SD (stable disease)/PD (progressive disease) and CR (partial response)/PR (complete response) to immunotherapy were enrolled in our analysis (Figure 1E).

Biological difference between the immunotherapy responders and non-responders

GSEA indicated that, in the immunotherapy responders, pathways of spermatogenesis, E2F target, G2/M checkpoint, MYC target, and mitotic spindle were significantly enriched (Figure 2A). Moreover, results of GO analysis revealed that the terms blood microparticle (GO:0076562), positive regulation of vasoconstriction (GO:0045907), zymogen activation (GO:0031638), regulation of vasoconstriction (GO:0019229), positive regulation of blood circulation (GO:1903524), and vasoconstriction (GO:0042310) were remarkably enriched in the immunotherapy responders (Figure 2B). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis indicated that in the immunotherapy responders, the terms antigen processing and presentation, cell cycle, olfactory

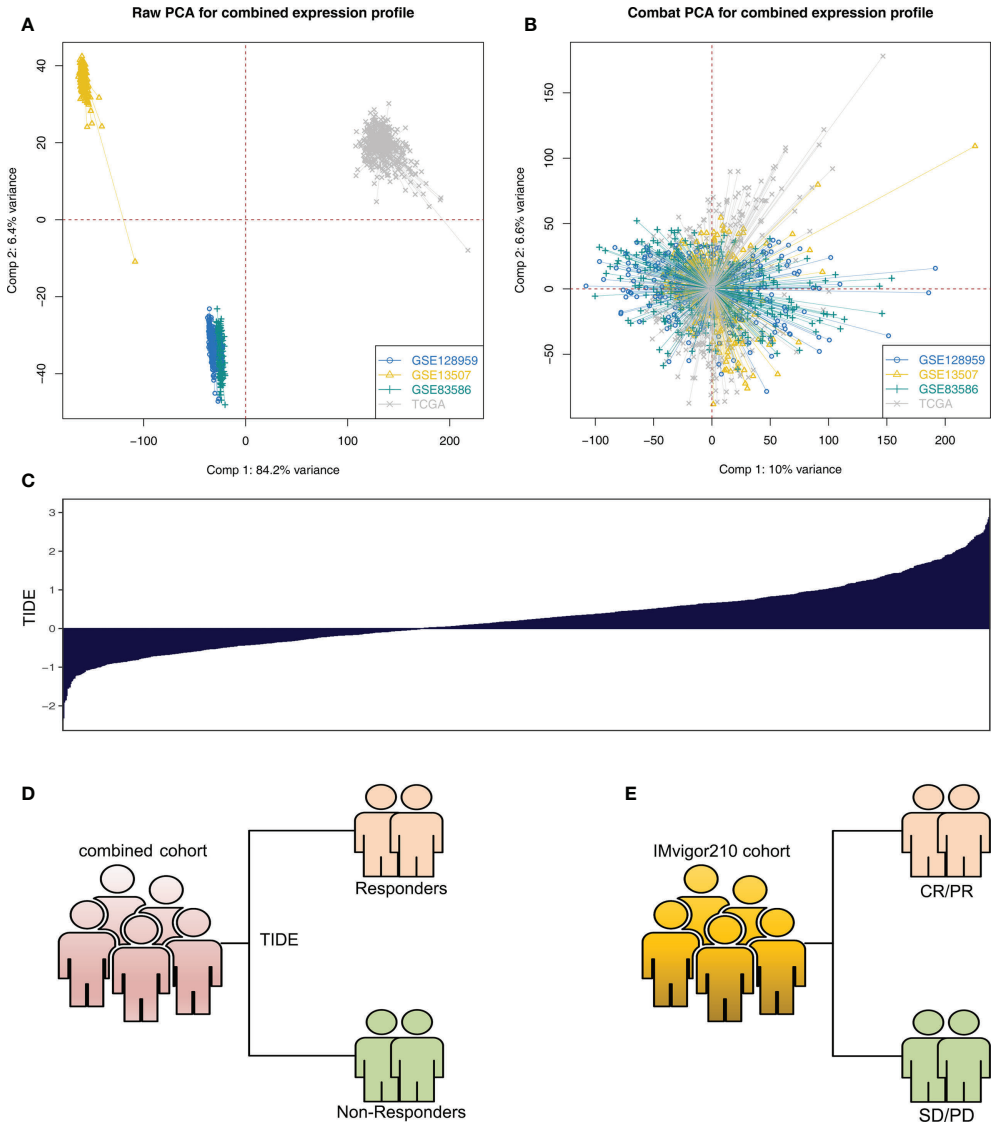


FIGURE 1
Data preparation **(A)** Significant batch effect of four independent bladder cancer cohorts was observed, namely, TCGA, GSE128959, GSE13507, and GSE83586. **(B)** Sva package was used for data combination and removal of batch difference. **(C)** Overview of TIDE score in the combined cohort (TCGA + GSE128959 + GSE13507 + GSE83586). **(D)** According to TIDE algorithm, the patients in the combined cohort were divided into immunotherapy responders and non-responders. **(E)** SD/PD and CR/PR patients in the IMvigor210 cohort.

transduction, DNA replication, natural killer cell-mediated cytotoxicity, and spliceosome were significantly enriched (Figure S2).

M2 macrophages is associated with immunotherapy response

Immune cell infiltration was quantified using the CIBERSORT algorithm, which is shown in Figure 3A. Results indicated that in the combined cohort, immunotherapy non-responders might have a higher level of follicular helper T cells, monocytes, M2 macrophages, activated dendritic cells, and neutrophils, but a lower level of Tregs (Figure 3B). Also, in the IMvigor210 cohort, the SD/PD patients might have higher activated memory CD4 T cells, follicular helper T cells, delta gamma T cells, activated NK cells, M2 macrophage, and activated dendritic cells (Figure 3C). In the combined and IMvigor210 cohorts, M2 macrophages showed a consistent trend (Figures 3D, E). Correlation analysis also indicated a positive correlation between M2 macrophages and TIDE score (Figure 3F, $R = 0.232$). The results showed that M2 macrophages might hamper the sensitivity of bladder cancer immunotherapy.

Identification of the important molecules involved in bladder cancer immunotherapy

Next, we identified differentially expressed gene (DEG) analysis between SD/PD and CR/PR in the IMvigor210 cohort. Under the threshold of $|\log FC| > 0.5$ and adjusted p-value < 0.05, 30 downregulated and 24 upregulated genes were determined (Figure 4A). Additionally, we identified that the genes remarkably correlated with TIDE score in the combined cohort (Table S1). Then, we intersected the molecules meeting the criteria of "Positive with TIDE and downregulated in CR/PR group" and "Negative with TIDE and upregulated in CR/PR group" based on the machine learning algorithm (Figures 4B, C, and Figure S3). Finally, three genes were identified, including PTHLH, BHMT2, and NGFR. Receiver operating characteristic (ROC) curves indicate a good prediction ability of these three molecules on patients' immunotherapy. In the combined cohort, the AUCs of PTHLH, BHMT2, and NGFR were 0.720, 0.735, and 0.661, respectively (Figures 4D–F). A logistic regression model was constructed with the formula "−4.9842 + 0.3738*PTHLH + 0.675*BHMT2 + 0.1128*NGFR", which showed satisfactory prediction efficiency (Figure 4G, AUC = 0.775). Also, in the IMvigor210 cohort, the AUCs of PTHLH, BHMT2, NGFR, and logistics scores were 0.679, 0.697, 0.706, and 0.750, respectively (Figures 4H–K). A higher level of PTHLH, BHMT2, and NGFR was observed in immunotherapy non-responders (combined cohort) and SD/PD patients (IMvigor210 cohort) (Figures 4L, M).

The logistics score affects patients' immunotherapy and prognosis

The overview of logistics calculated with the above formula is shown in Figures 5A, B. In the combined cohort, patients in the immunotherapy non-responder group had a higher logistics score (Figure 5C). Moreover, in patients with a high logistics score, a higher proportion of immunotherapy non-responders was found (Figure 5D, 88.6% *vs.* 65.8%). Meanwhile, in the IMvigor210 cohort, we observed the same trend in SD/PD patients (Figures 5E, F, 88.6% *vs.* 65.8%). Kaplan–Meier survival curve indicated that the logistics score was correlated with worse survival in both combined and IMvigor210 cohorts and dead cases might have a higher logistics score (Figures 5G–J). Clinical correlation analysis revealed that the logistics score was associated with a more progressive clinical stage and grade (Figure 5K).

Immune microenvironment difference in patients with a high and those with a low logistics score

Immune checkpoint is tightly correlated with the immunotherapy response. Therefore, we evaluated the level of hub immune checkpoints (CD274, PDCD1, CTLA, and PDCD1LG2) in patients with a high and those with a low logistics score. Results indicated that all these immune checkpoints had a higher expression in patients with a high logistics score (Figures 6A–D). We also investigated the immune difference in patients with a high and those with a low logistics score. In the combined and IMvigor210 cohorts, we all observed a higher level of M2 macrophages in patients with a high logistics score (Figures 6E, F). Next, results showed that M2 macrophages were positively correlated with PTHLH, BHMT2, NGFR, and logistics score (Figures 6G–J).

Patients with a high logistics score might have a higher genomic instability

Moreover, we observed a higher stromal score, immune score, TMB score, and MSI score in patients with a high logistics score (Figures 7A–D). The mRNAsi obtained is shown in Figure 7E. Patients with a high logistics score might have a higher mRNAsi (Figure 7F). Meanwhile, we found that the patients with a high logistics score might have a lower level of amplification frequency in the 3p, 3q, and 21q sites in the chromosome (Figure 7G). In the deletion frequency, a lower level in the 1q, 2q, 5p, 9q, 14q, and 16q sites were found in patients with a high logistics score (Figure 7G). Meanwhile, the difference in SCNA level was also illustrated (Figure 7H).

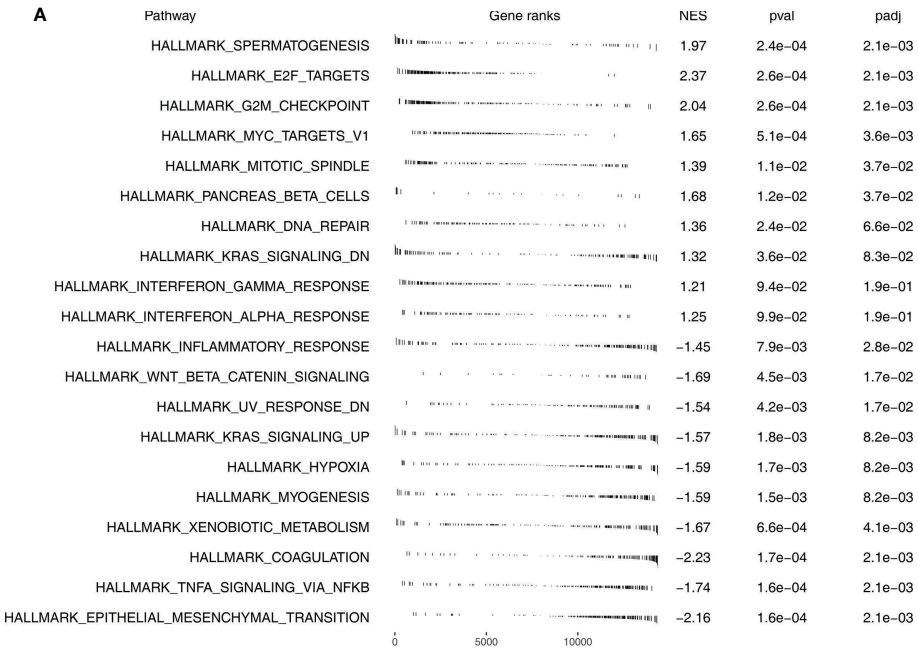


FIGURE 2
Biological enrichment between immunotherapy responders and non-responders **(A)** GSEA between immunotherapy responders and non-responders based on the Hallmark gene set. **(B)** GO analysis using the clusterprofiler package.

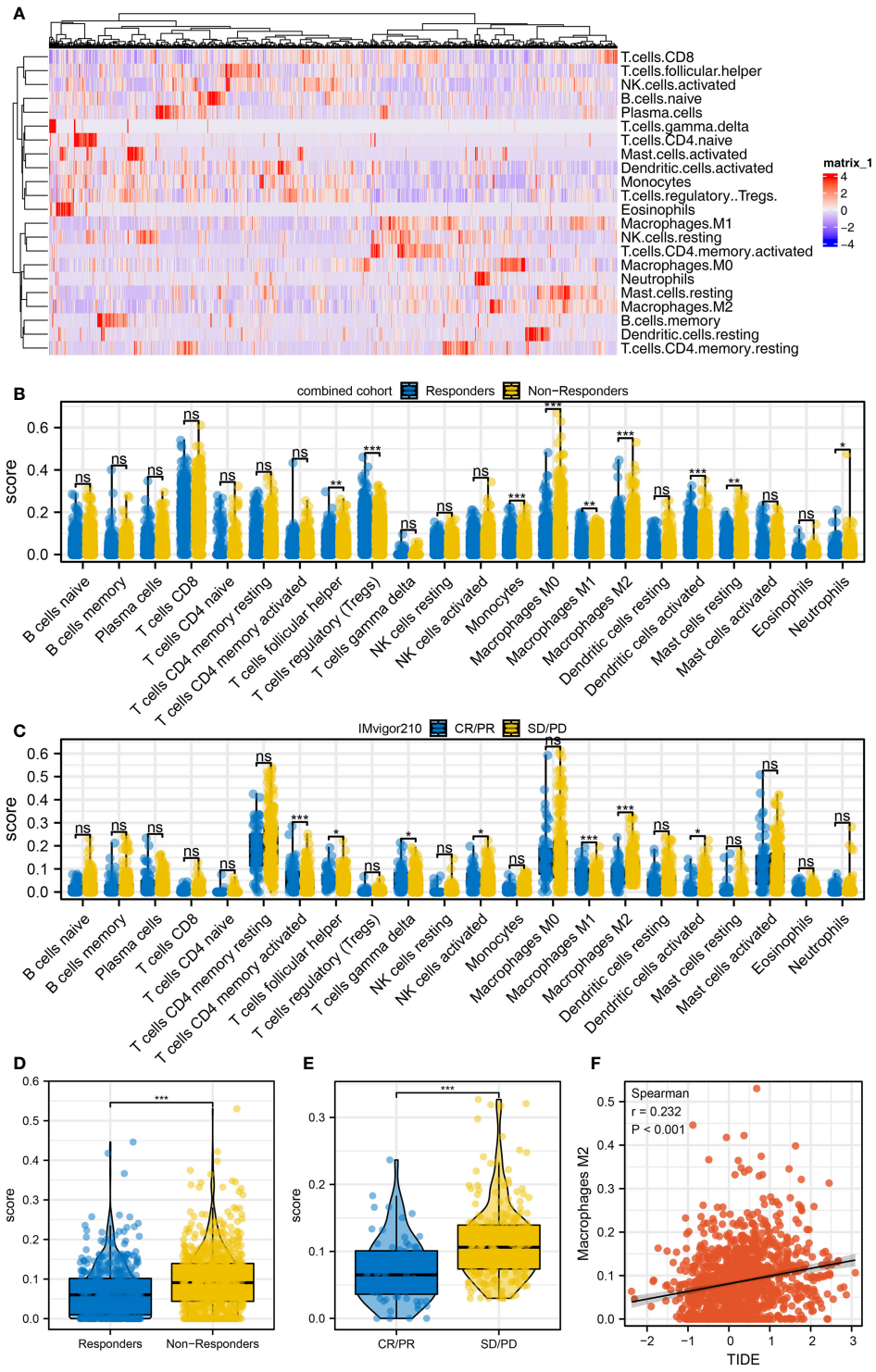


FIGURE 3
Immune microenvironment difference between immunotherapy responders and non-responders **(A)** The CIBERSORT algorithm was utilized to quantify the immune microenvironment of bladder cancer tissue. **(B)** The level of quantified immune cells in the immunotherapy responders and non-responders (combined cohort), ns = $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(C)** The level of quantified immune cells in the immunotherapy CR/PR and SD/PD patients (IMvigor210 cohort), ns = $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(D)** The level of M2 macrophage in immunotherapy responders and non-responders, ***$p < 0.001$. **(E)** The level of M2 macrophage in immunotherapy SD/PD and CR/PR patients, ***$p < 0.001$. **(F)** Correlation between the M2 macrophage and TIDE score.

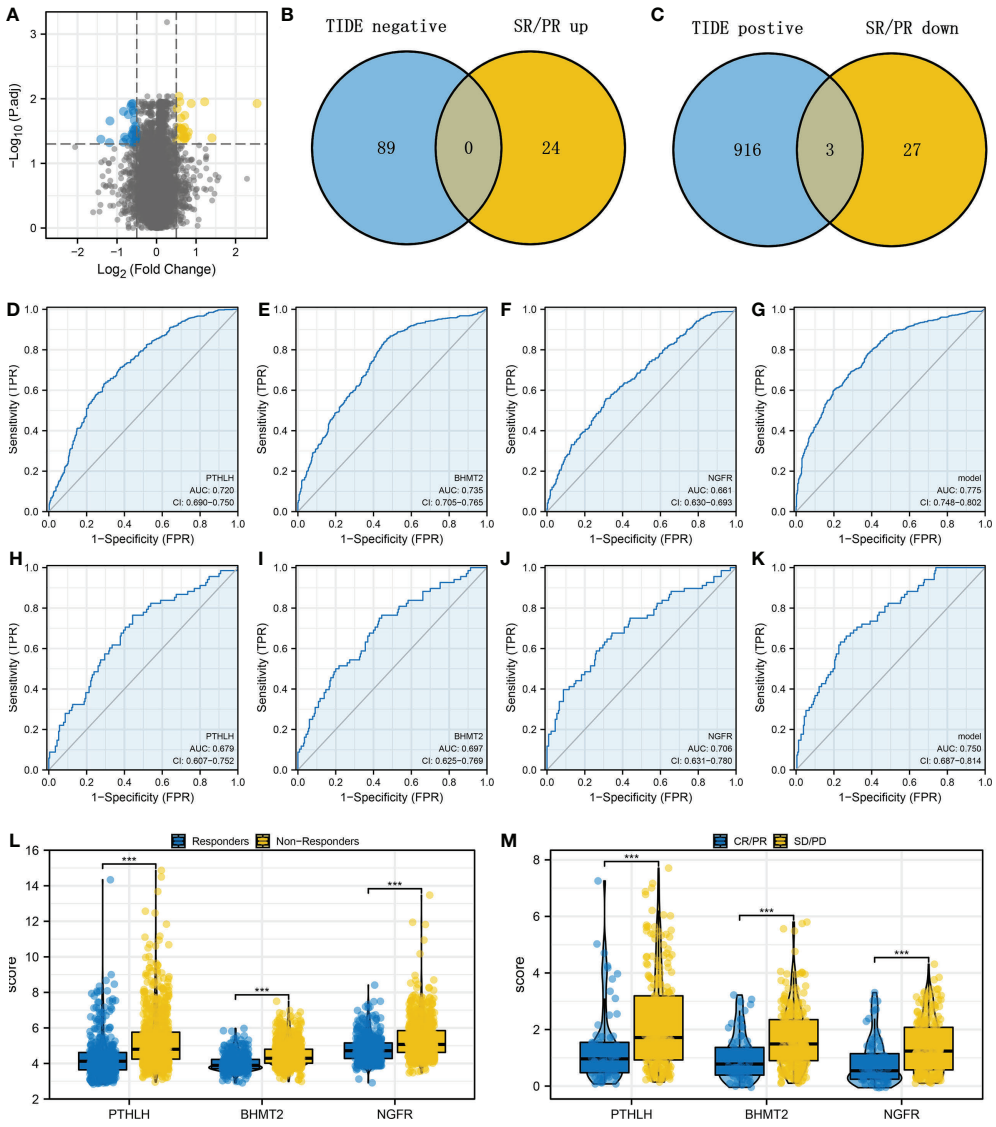


FIGURE 4
Identification of the PTHLH, BHMT2, and NGFR **(A)** Limma package was used for DEG analysis between the SD/PD and CR/PR patients with the threshold of |logFC| > 0.5 and adjusted *p*-value < 0.05. **(B)** Intersection of the genes negatively correlated with TIDE score and upregulated in CR/PR (LASSO logistics regression). **(C)** Intersection of the genes positively correlated with TIDE score and downregulated in CR/PR (LASSO logistics regression). **(D–G)** Performance of PTHLH, BHMT2, NGFR, and the logistics model in predicting bladder cancer immunotherapy (combined cohort). **(H–K)** ROC curves were used to evaluate the performance of PTHLH, BHMT2, NGFR, and the logistics model in predicting bladder cancer immunotherapy (IMvigor210 cohort). **(L)** The level of PTHLH, BHMT2, and NGFR in immunotherapy responders and non-responders. **(M)** The level of PTHLH, BHMT2, and NGFR in immunotherapy SD/PD and CR/PR patients. ***P < 0.001.

PTHLH was associated with patients' survival and M2 macrophage polarization

Following this, we investigated the prognostic role of PTHLH, BHMT2, and NGFR. Results showed that all these three genes were risk factors for overall survival and disease-free survival, but not for progression-free survival (Figures 8A–C and Figure S4). Then, we try to identify the expression difference of these molecules in bladder cancer tissue. No significant difference was found in PTHLH mRNA level between normal and bladder cancer tissue (Figure 8D). However, the HPA database showed a high protein level of PTHLH in bladder cancer tissue (Figure 8E). For the mRNA level, BHMT2 was lower in bladder cancer tissue, but not in protein levels (Figures 8F, G). Also, the same trend was noticed in NGFR

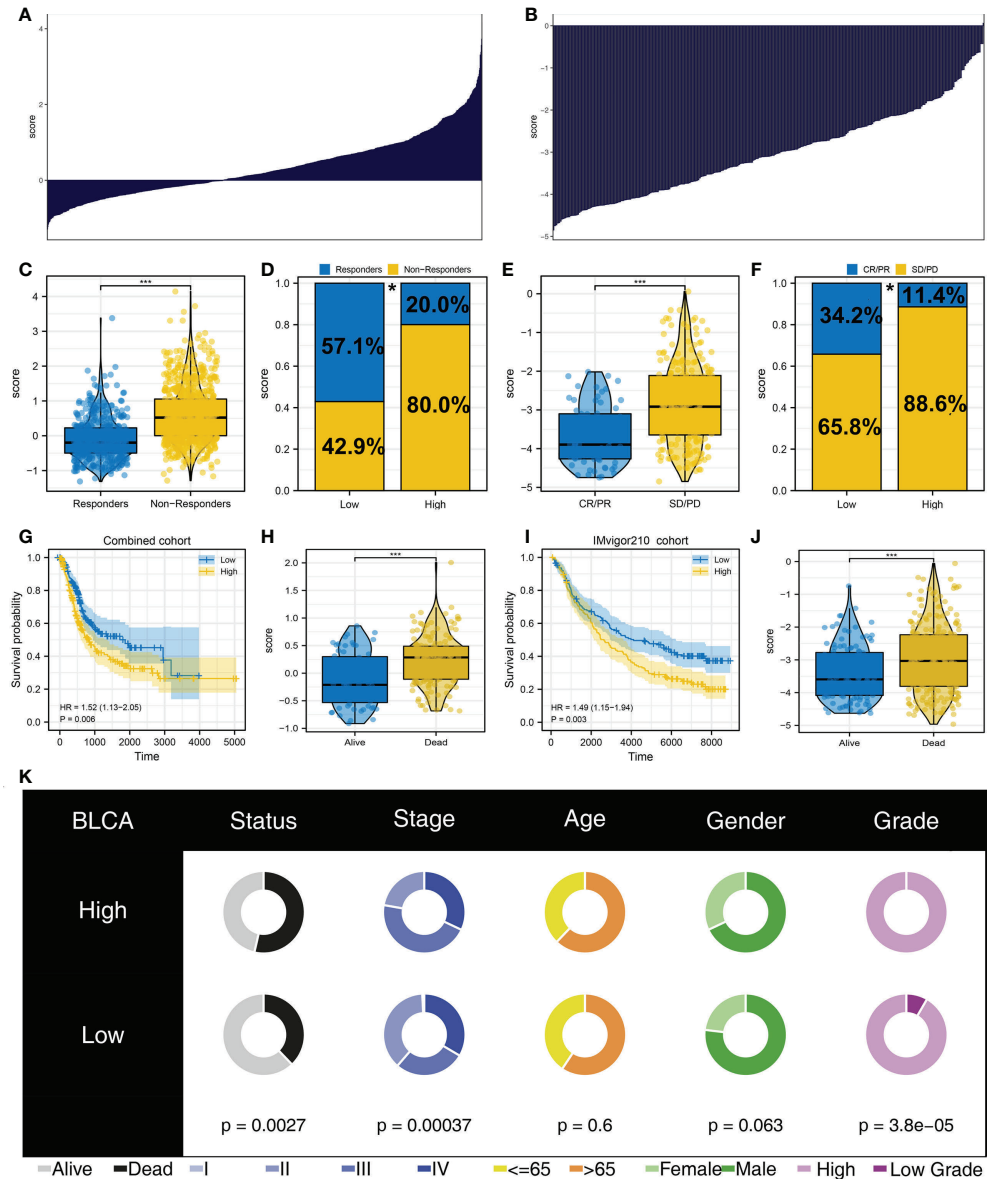


FIGURE 5
Logistics score was associated with patient immunotherapy response and prognosis **(A)** Calculated logistics score in the combined cohort. **(B)** Calculated logistics score in the IMvigor210 cohort. **(C)** Logistics score in immunotherapy responders and non-responders (combined cohort). **(D)** Percentage of immunotherapy responders and non-responders in patients with a high and those with a low logistics score (combined cohort), *$p < 0.05$. **(E)** Logistics score in immunotherapy SD/PD and CR/PR patients (IMvigor210 cohort). **(F)** Percentage of immunotherapy SD/PD and CR/PR in patients with a high and those with a low logistics score (IMvigor210 cohort), *$p < 0.05$. **(G)** Kaplan–Meier survival curves in patients with a high and those with a low logistics score (combined cohort). **(H)** The level of logistics score in alive and dead cases in the combined cohort. **(I)** Kaplan–Meier survival curves in patients with a high and those with a low logistics score (IMvigor210 cohort). **(J)** The level of logistics score in alive and dead cases in the IMvigor210 cohort. **(K)** Clinical correlation analysis of logistics score. ***P < 0.001.

(Figures 8H, I). Considering the high protein level of PTHLH in bladder cancer tissue, we next try to validate its association with M2 macrophages. The knockdown efficiency is shown in Figure S5. Immunofluorescence showed that knockdown of PTHLH in bladder cancer cells can significantly inhibit the M2 polarization of co-culture M0 macrophages (Figures 8J, K).

Discussion

There is no doubt that bladder cancer represents one of the greatest health problems globally (18). In recent years, immunotherapy, such as immune checkpoint inhibitors, has improved bladder cancer treatment options.

10.3389/fimmu.2022.1051063

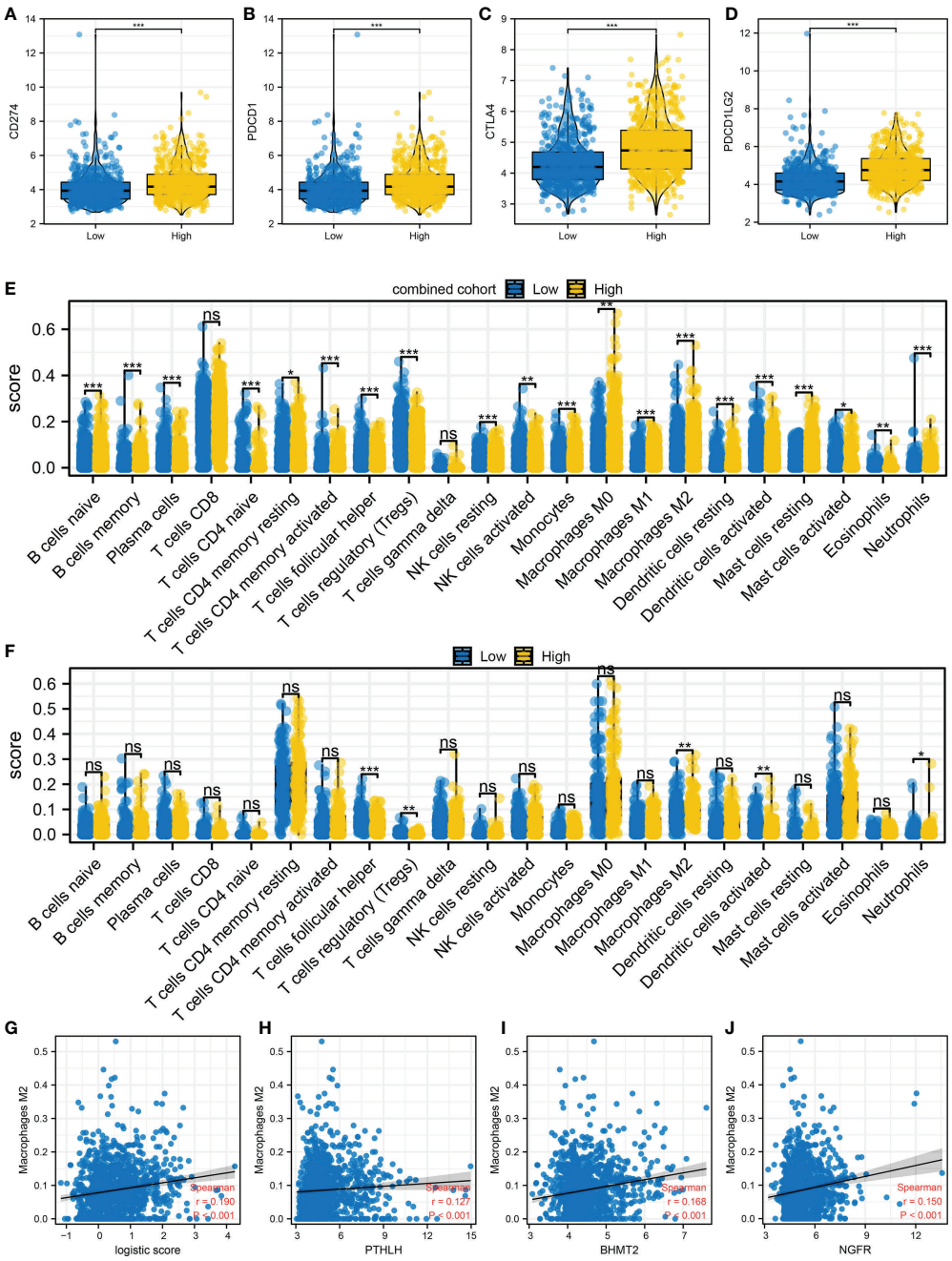


FIGURE 6
Immune microenvironment difference in patients with a high and those with a low logistics score **(A–D)** Hub immune checkpoints in patients with a high and those with a low logistics score. **(E)** The level of quantified immune cells in patients with a high and those with a low logistics score (combined cohort). **(F)** The level of quantified immune cells in patients with a high and those with a low logistics score (IMvigor210 cohort). **(G)** Correlation between logistics score and M2 macrophages. **(H–J)** Correlation between PTHLH, BHMT2, NGFR, and M2 macrophages. *P < 0.05, **P < 0.01, ***P < 0.001. ns: P > 0.05.

Here, through the combined (TCGA, GSE128959, GSE13507, and GSE83586) and IMvigor210 cohorts, we comprehensively investigated the biological and immune microenvironment differences in patients with diverse immunotherapy responses. Meanwhile, we found that M2 macrophage could affect the sensitivity of bladder cancer patients to immunotherapy. Moreover, PTHLH, BHMT2, and NGFR were identified, which all have good prediction abilities

10.3389/fimmu.2022.1051063

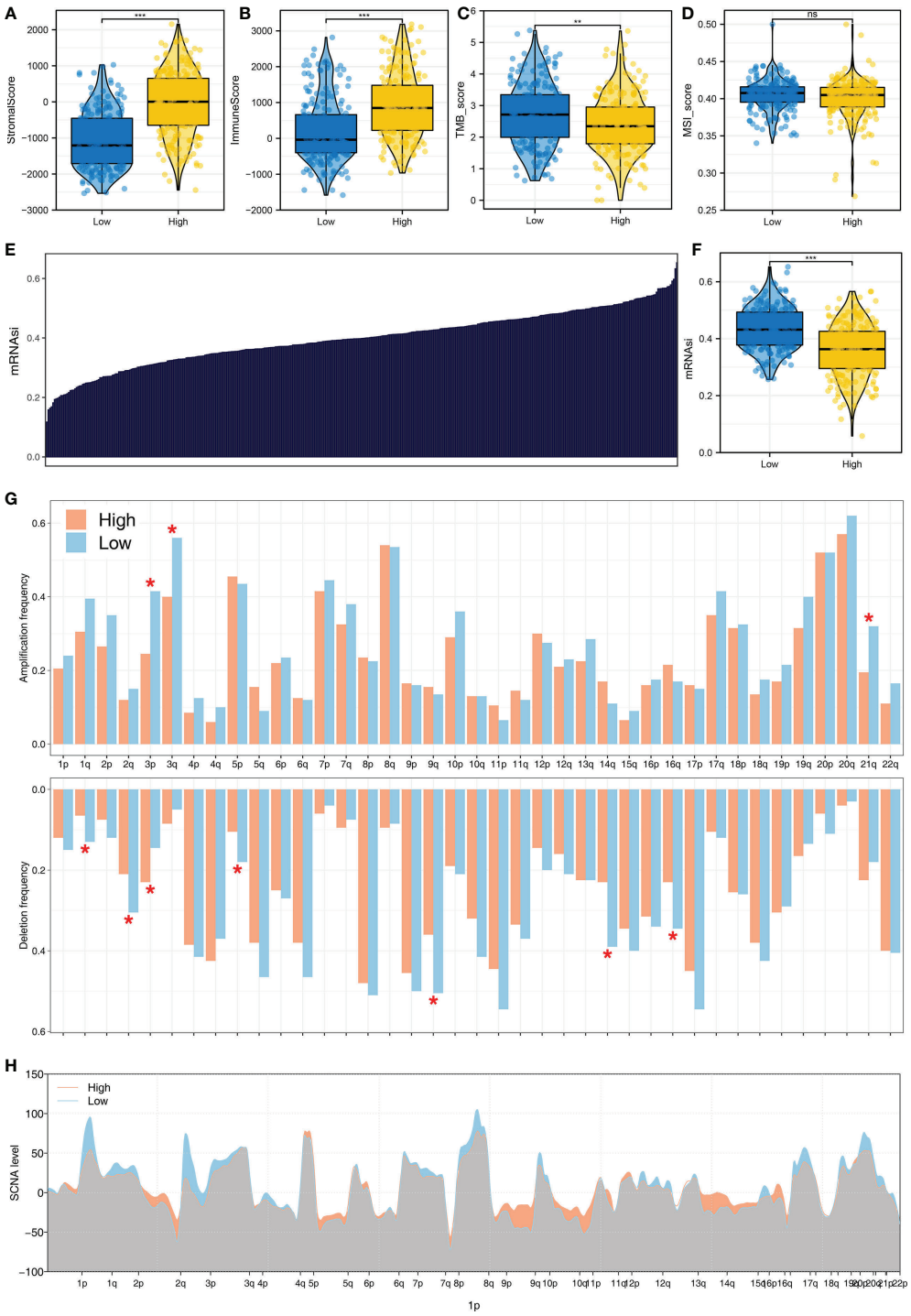


FIGURE 7
The genomic difference in patients with a high and those with a low logistics score **(A)** Level of stromal score in patients with a high and those with a low logistics score. **(B)** Level of immune score in patients with a high and those with a low logistics score. **(C)** Level of TMB score in patients with a high and those with a low logistics score. **(D)** Level of stromal score in patients with a high and those with a low logistics score. **(E)** The obtained mRNAsi in the TCGA database. **(F)** Level of mRNAsi score in patients with a high and those with a low logistics score. **(G, H)** Comparisons of arm-level amplification and deletion frequencies and focal-level amplification and deletion levels in patients with high and those with low logistics scores. *P < 0.05, **P < 0.01, ***P < 0.001. ns: P > 0.05.

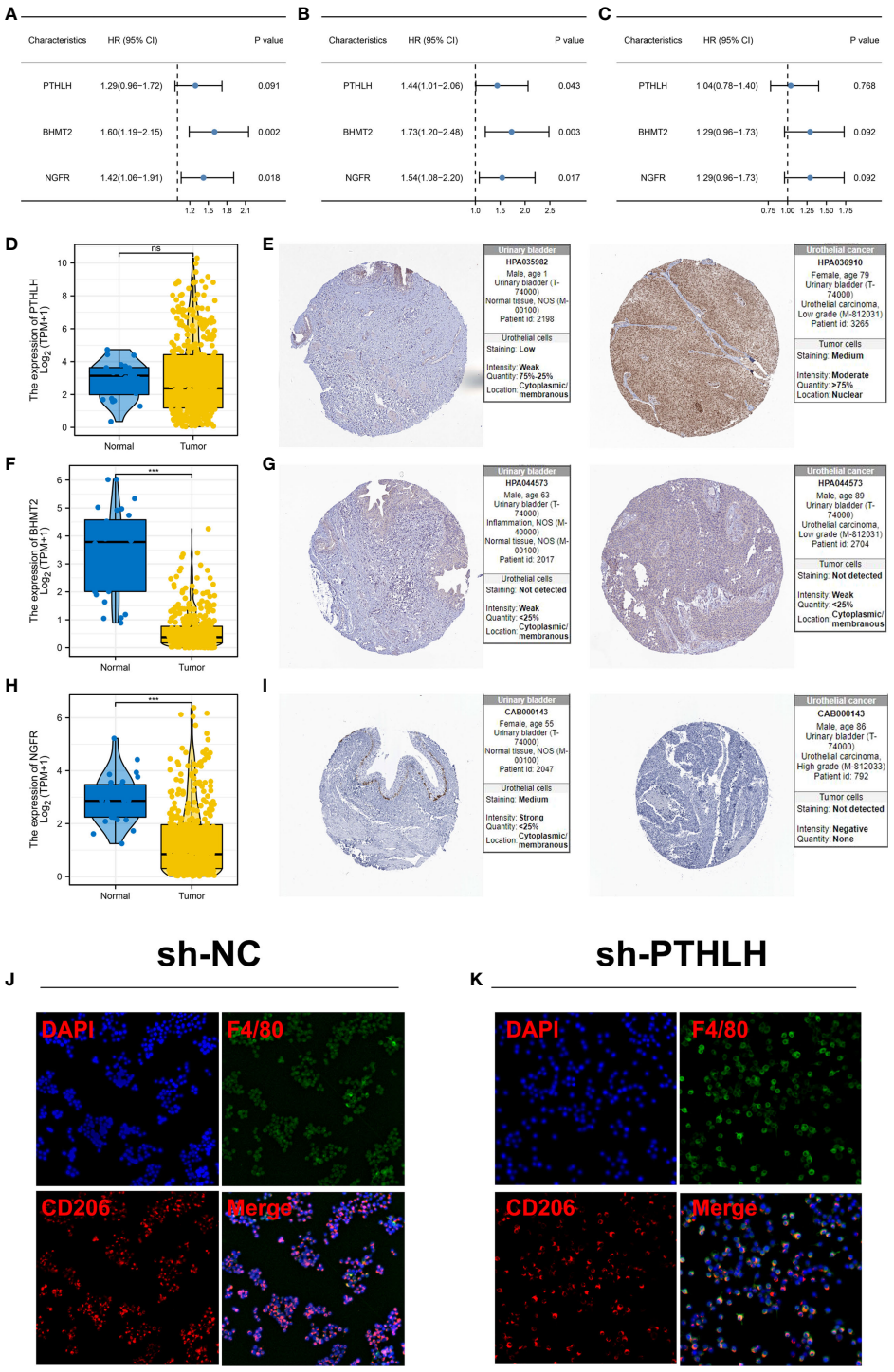


FIGURE 8
Further exploration of PTHLH, BHMT2, and NGFR **(A)** Prognosis correlation of PTHLH, BHMT2, and NGFR in overall survival. **(B)** Prognosis correlation of PTHLH, BHMT2, and NGFR in disease-free survival. **(C)** Prognosis correlation of PTHLH, BHMT2, and NGFR in progression-free survival. **(D)** Expression level of PTHLH in normal bladder and bladder cancer tissue in mRNA level. **(E)** Representative immunohistochemistry image of PTHLH protein level in normal bladder and bladder cancer tissue. **(F)** Expression level of BHMT2 in normal bladder and bladder cancer tissue in mRNA level. **(G)** Representative immunohistochemistry image of BHMT2 protein level in normal bladder and bladder cancer tissue. **(H)** Expression level of NGFR in normal bladder and bladder cancer tissue in mRNA level.; **(I)** Representative immunohistochemistry image of NGFR protein level in normal bladder and bladder cancer tissue. **(J, K)** Knockdown of PTHLH in bladder cancer cells can significantly inhibit the M2 polarization of co-culture M0 macrophages. ***P < 0.001. ns: P > 0.05.

for patients' immunotherapy. Then, a logistics regression model was established based on PTHLH, BHMT2, and NGFR, and each patient was assigned a logistics score. Afterwards, we explored the difference in patients with a high and those with a low logistics score, including biological enrichment, immune microenvironment, and genomic characteristics. Meanwhile, data from the HPA database indicated a higher protein level of PTHLH in bladder cancer tissue. Immunofluorescence showed that the knockdown of PTHLH in bladder cancer cells can significantly inhibit the M2 polarization of co-culture M0 macrophages.

Based on the immune analysis, we found that M2 macrophages might affect bladder cancer immunotherapy. Tissue homeostasis is mediated by macrophages. Meanwhile, cancer-associated macrophages show the ability to hamper T-cell recruitment and its function, as well as regulate other aspects of tumor immunity (19). Previous studies have reported the potential effects of macrophages in tumor immunotherapy (20). The effect of TAMs on naive T-cell proliferation has been demonstrated in numerous studies, which suggests that macrophages can suppress T-cell function directly (21). An example is that arginase-1 is a common marker of M2 macrophage in mice, which was also involved in antitumor activity and T-cell fitness (22). Meanwhile, in ovarian cancer, Curiel et al. found that macrophages could mediate the recruitment of Treg cells through secreting CCL22, contributing to immunosuppression (23). In breast cancer, macrophages characterized by the expression of IL10 could hamper CD8+ T-cell-dependent responses by inhibiting IL-12 expression in intratumoral dendritic cells (24).

We identified three genes, PTHLH, BHMT2, and NGFR, involved in bladder cancer immunotherapy. Also, the logistics model based on these three genes can effectively indicate immunotherapy sensitivity and patient prognosis. Moreover, we found that all these three genes were positively correlated with M2 macrophage infiltration. In renal cancer, Yao et al. demonstrated that the upregulation of PTHLH can indicate more progressive clinical features and poor prognosis (25). In pancreatic cancer, the protein named PTHrP (encoded by PTHLH) can drive the growth of primary and metastatic tumors in mice (26). In colon cancer, Chen et al. indicated that NGFR could act as a tumor suppressor by activating S100A9, thus leading to the enhanced apoptotic and autophagic effects of 5-fluorouracil (27). Huang et al. found that the NGFR-FOXP3 positive feedback loop contributes to ICOTINIB resistance in non-small cell lung cancer (28). Also, we found that PTHLH can induce M2 macrophage polarization. Our results indicated that PTHLH, BHMT2, and NGFR might be novel targets for bladder cancer immunotherapy and prognosis. In real practice, detecting the relative level of PTHLH, BHMT2, and NGFR through a customized chip or absolute real-time PCR can indicate the sensitivity of immunotherapy and therefore contribute to the therapy option.

Also, we noticed a higher genomic instability in patients with a high logistics score, including TMB, MSI, and tumor stemness index. It is widely believed that TMB is related to the response to immunotherapy for tumors (29). Zhu et al. found that the mutations in EP300 lead to TMB and promote antitumor immunity in bladder cancer (30). Meanwhile, Zhan et al. found that SOX2OT can promote bladder cancer stemness, as well as the malignant phenotype through modulating SOX2 (31). These results indicated that the patients with a high logistics score might have a poor prognosis due to the higher genomic instability.

Even though our research is based on reliable data and analysis, there are some limitations to consider. Firstly, our analysis primarily had a Western population, and therefore, underlying race bias might reduce the credibility of conclusions. Secondly, the loss of probes when combining the TCGA and GSE cohorts might result in information loss.

Data availability statement

Publicly available datasets were analyzed in this study, the names of the repositories and accession numbers are available within the article/Supplementary Material.

Author contributions

WJ, HX, and DG performed the analysis. WJ and BY performed the experiments. CM designed this work. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Key research and development project of science and technology department of Sichuan province (Number 2022YFS0156).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1051063/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The flowchart of whole study.

SUPPLEMENTARY FIGURE 2
KEGG analysis of immunotherapy responders.

SUPPLEMENTARY FIGURE 3
Machine learning algorithm – LASSO logistics regression.

SUPPLEMENTARY FIGURE 4
Prognostic role of PTHLH, BHMT2, and NGFR **(A–C)** Kaplan–Meier survival curves of PTHLH, BHMT2, and NGFR in overall survival. **(D–F)** Kaplan–Meier survival curves of PTHLH, BHMT2, and NGFR in disease-free survival. **(G–I)** Kaplan–Meier survival curves of PTHLH, BHMT2, and NGFR in disease-free survival.

SUPPLEMENTARY FIGURE 5
Knockdown efficiency of PTHLH.

References

1. Dobruch J, Daneshmand S, Fisch M, Lotan Y, Noon AP, Resnick MJ, et al. Gender and bladder cancer: A collaborative review of etiology, biology, and outcomes. *Eur Urol* (2016) 69(2):300–10. doi: 10.1016/j.eururo.2015.08.037

2. Patel VG, Oh WK, Galsky MD. Treatment of muscle-invasive and advanced bladder cancer in 2020. *CA: Cancer J Clin* (2020) 70(5):404–23. doi: 10.3322/caac.21631

3. Pfail JL, Katims AB, Alerasool P, Sfakianos JP. Immunotherapy in non-Muscle-Invasive bladder cancer: Current status and future directions. *World J Urol* (2021) 39(5):1319–29. doi: 10.1007/s00345-020-03474-8

4. Lenfant L, Rouprêt M. [Immunotherapy and bladder cancer]. *Biol aujourd'hui* (2018) 212(3-4):81–4. doi: 10.1051/jbio/2018028

5. Smith SG, Zaharoff DA. Future directions in bladder cancer immunotherapy: Towards adaptive immunity. *Immunotherapy* (2016) 8(3):351–65. doi: 10.2217/imt.15.122

6. deLeeuw RJ, Kost SE, Kakal JA, Nelson BH. The prognostic value of Foxp3+ tumor-infiltrating lymphocytes in cancer: A critical review of the literature. *Clin Cancer res: an Off J Am Assoc Cancer Res* (2012) 18(11):3022–9. doi: 10.1158/1078-0432.Ccr-11-3216

7. Shi R, Wang X, Wu Y, Xu B, Zhao T, Trapp C, et al. Apobec-mediated mutagenesis is a favorable predictor of prognosis and immunotherapy for bladder cancer patients: Evidence from pan-cancer analysis and multiple databases. *Theranostics* (2022) 12(9):4181–99. doi: 10.7150/thno.73235

8. Groeneveld CS, Fontugne J, Cabel L, Bernard-Pierrot I, Radvanyi F, Allory Y, et al. Tertiary lymphoid structures marker Cxcl13 is associated with better survival for patients with advanced-stage bladder cancer treated with immunotherapy. *Eur J Cancer (Oxford England: 1990)* (2021) 148:181–9. doi: 10.1016/j.ejca.2021.01.036

9. Yi X, Zheng X, Xu H, Li J, Zhang T, Ge P, et al. Igfbp7 and the tumor immune landscape: A novel target for immunotherapy in bladder cancer. *Front Immunol* (2022) 13:898493. doi: 10.3389/fimmu.2022.898493

10. Carosella ED, Ploussard G, LeMaoult J, Desgrandchamps F. A systematic review of immunotherapy in urologic cancer: Evolving roles for targeting of ctla-4, pd-1/Pd-L1, and hla-G. *Eur Urol* (2015) 68(2):267–79. doi: 10.1016/j.eururo.2015.02.032

11. Daro-Faye M, Kassouf W, Souhami L, Marcq G, Cury F, Niazi T, et al. Combined radiotherapy and immunotherapy in urothelial bladder cancer: Harnessing the full potential of the anti-tumor immune response. *World J Urol* (2021) 39(5):1331–43. doi: 10.1007/s00345-020-03440-4

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

13. McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic lasso regression for dietary intakes and breast cancer. *Nutrients* (2020) 12(9):2652. doi: 10.3390/nu12092652

14. Ren X, Chen X, Zhang X, Jiang S, Zhang T, Li G, et al. Immune microenvironment and response in prostate cancer using Large population cohorts. *Front Immunol* (2021) 12:686809. doi: 10.3389/fimmu.2021.686809

15. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An r package for comparing biological themes among gene clusters. *Omics: J Integr Biol* (2012) 16 (5):284–7. doi: 10.1089/omi.2011.0118

16. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with cibersort. *Methods Mol Biol* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12

17. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* (2018) 173(2):338–54.e15. doi: 10.1016/j.cell.2018.03.034

18. de Braud F, Maffezzini M, Vitale V, Bruzzi P, Gatta G, Hendry WF, et al. Bladder cancer. *Crit Rev oncology/hematol* (2002) 41(1):89–106. doi: 10.1016/s1040-8428(01)00128-7

19. Loyher PL, Hamon P, Laviron M, Meghraoui-Kheddar A, Goncalves E, Deng Z, et al. Macrophages of distinct origins contribute to tumor development in the lung. *J Exp Med* (2018) 215(10):2536–53. doi: 10.1084/jem.20180534

20. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol* (2019) 19(6):369–82. doi: 10.1038/s41577-019-0127-6

21. Doedens AL, Stockmann C, Rubinstein MP, Liao D, Zhang N, DeNardo DG, et al. Macrophage expression of hypoxia-inducible factor-1 alpha suppresses T-cell function and promotes tumor progression. *Cancer Res* (2010) 70(19):7465–75. doi: 10.1158/0008-5472.Can-10-1439

22. Geiger R, Rieckmann JC, Wolf T, Basso C, Feng Y, Fuhrer T, et al. L-arginine modulates T cell metabolism and enhances survival and anti-tumor activity. *Cell* (2016) 167(3):829–42.e13. doi: 10.1016/j.cell.2016.09.031

23. Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, et al. Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nat Med* (2004) 10(9):942–9. doi: 10.1038/nm1093

24. Ruffell B, Chang-Strachan D, Chan V, Rosenbusch A, Ho CM, Pryer N, et al. Macrophage il-10 blocks Cd8+ T cell-dependent responses to chemotherapy by suppressing il-12 expression in intratumoral dendritic cells. *Cancer Cell* (2014) 26 (5):623–37. doi: 10.1016/j.ccell.2014.09.006

25. Yao M, Murakami T, Shioi K, Mizuno N, Ito H, Kondo K, et al. Tumor signatures of pthlh overexpression, high serum calcium, and poor prognosis were observed exclusively in clear cell but not non clear cell renal carcinomas. *Cancer Med* (2014) 3(4):845–54. doi: 10.1002/cam4.270

26. Pitarresi JR, Norgard RJ, Chiarella AM, Suzuki K, Bakir B, Sahu V, et al. Pthrp drives pancreatic cancer growth and metastasis and reveals a new therapeutic vulnerability. *Cancer Discov* (2021) 11(7):1774–91. doi: 10.1158/2159-8290.Cd-20-1098

27. Chen H, Huang J, Chen C, Jiang Y, Feng X, Liao Y, et al. Ngfr increases the chemosensitivity of colorectal cancer cells by enhancing the apoptotic and autophagic effects of 5-fluorouracil *Via* the activation of S100a9. *Front Oncol* (2021) 11:652081. doi: 10.3389/fonc.2021.652081

28. Huang J, Yu Q, Zhou Y, Chu Y, Jiang F, Zhu X, et al. A positive feedback loop formed by ngfr and Foxp3 contributes to the resistance of non-small cell lung

cancer to icotinib. *Trans Cancer Res* (2020) 9(2):1044–52. doi: 10.21037/tcr.2019.12.60

29. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann Oncol: Off J Eur Soc Med Oncol* (2019) 30(1):44–56. doi: 10.1093/annonc/mdy495

30. Zhu G, Pei L, Li Y, Gou X. Ep300 mutation is associated with tumor mutation burden and promotes antitumor immunity in bladder cancer patients. *Aging* (2020) 12(3):2132–41. doi: 10.18632/aging.102728

31. Zhan Y, Chen Z, He S, Gong Y, He A, Li Y, et al. Long non-coding rna Sox2ot promotes the stemness phenotype of bladder cancer cells by modulating Sox2. *Mol Cancer* (2020) 19(1):25. doi: 10.1186/s12943-020-1143-7

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 November 2022
DOI 10.3389/fonc.2022.1024915

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Hu Qian,
Zunyi Medical University, China
Changgang Sun,
Affiliated Hospital of Weifang Medical University, China
Ming Liu,
Institute of Cancer and Basic Medicine, Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Xu Cao
hughcaoxu@hotmail.com

†These authors have contributed equally to this work

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

RECEIVED 22 August 2022
ACCEPTED 19 October 2022
PUBLISHED 11 November 2022

CITATION
Wang X, Xia G, Xiao S, Wu S, Zhang L, Huang J, Zhang W and Cao X (2022) A ferroptosis-related gene signature associated with immune landscape and therapeutic response in osteosarcoma.
Front. Oncol. 12:1024915.
doi: 10.3389/fonc.2022.1024915

A ferroptosis-related gene signature associated with immune landscape and therapeutic response in osteosarcoma

Xinxing Wang[1†], Guang Xia[1†], Shilang Xiao[2], Song Wu[1], Lina Zhang[1], Junjie Huang[1], Wenxiu Zhang[1] and Xu Cao[1*]

[1]Department of Orthopaedics, The Third Xiangya Hospital, Central South University, Changsha, China, [2]Department of Gastroenterology, The Third Xiangya Hospital, Central South University, Changsha, China

Background: The role of ferroptosis in tumor progression and immune microenvironment is extensively investigated. However, the potential value of ferroptosis regulators in predicting prognosis and therapeutic strategies for osteosarcoma (OS) patients remains to be elucidated.

Methods: Here, we extracted transcriptomic and survival data from Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Gene Expression Omnibus (GEO) to investigate the expression and prognostic value of ferroptosis regulators in OS patients. After comprehensive analyses, including Gene set variation analysis (GSVA), single-sample gene-set enrichment analysis (ssGSEA), Estimated Stromal and Immune cells in Malignant Tumor tissues using Expression (ESTIMATE), single-cell RNA sequencing, and biological experiments, our constructed 8-ferroptosis-regulators prognostic signature effectively predicted the immune landscape, prognosis, and chemoradiotherapy strategies for OS patients.

Results: We constructed an 8-ferroptosis-regulators signature that could predict the survival outcome of OS. The signature algorithm scored samples, and high-scoring patients were more prone to worse prognoses. The tumor immune landscape suggested the positive relevance between risk score and immunosuppression. Interfering HILPDA and MUC1 expression would inhibit tumor cell proliferation and migration, and MUC1 might improve the ferroptosis resistance of OS cells. Moreover, we predicted chemoradiotherapy strategies of cancer patients following ferroptosis-risk-score groups.

Conclusion: Dysregulated ferroptosis gene expression can affect OS progression by affecting the tumor immune landscape and ferroptosis resistance. Our risk model can predict OS survival outcomes, and we propose that HILPDA and MUC1 are potential targets for cancer therapy.

KEYWORDS

ferroptosis, osteosarcoma, immune microenvironment, prognostic risk model, sing-cell sequencing

Introduction

Osteosarcoma (OS), the most common bone tumor, is a highly aggressive malignancy that frequently occurs in childhood and adolescence and has a worldwide annual incidence rate of 1~3 cases per million (1). OS originates from primitive mesenchymal cells in bone and rarely in soft tissue and progresses to pulmonary metastasis, whose subsequent relapse remains the primary cause of OS-related death (2). The current treatment strategy for OS patients includes neoadjuvant chemotherapy combined with surgical removal of the primary lesions and evidenced metastatic lesions, followed by additional adjuvant chemotherapy (3). Compared with management regimens before 1970, multiagent chemotherapy has considerably improved the long-term survival of localized OS patients from 20% to 70%. However, metastatic and recurrent OS patients still have a significantly low survival rate (4). Unfortunately, since the mid-1970s, little progress has been made in improving standard management strategies and increasing the survival rate of OS patients (3). The therapeutic outcome of OS is significantly impacted by intrinsic cellular heterogeneity and complex immunogenic mechanisms (5). Immune checkpoint inhibitors have made breakthroughs in the immunotherapy of various cancers (6, 7), whereas the therapeutical effect of targeting TILs and PD-L1 in managing OS is inconsistent (8–11). These suggest that OS might have a complex immune status that helps cancer cells evade the immune surveillance-mediated cell death. Therefore, identifying novel effective immune therapeutic targets to benefit treatment for OS is needed. Recently, three newly identified types of cell death, including ferroptosis, necroptosis, and pyroptosis, have been suggested to have crosstalk with antitumor immunity (12). As a research hotspot, ferroptosis was involved in multiple antitumor mechanisms. However, the relationship between ferroptosis and OS immune microenvironment remains to be elucidated.

Ferroptosis, distinguished from traditional cell death-like apoptosis, cell autophagy, or necroptosis, is a novel programmed cell death characterized by iron-dependent lipid peroxidation (13). Previous studies have suggested that ferroptosis regulators, including GPX4 (14), FANCD2 (15), P53 (16), and HSPB1 (17), are related to oncogenesis and progression. Increasing evidence has identified the pivotal role of ferroptosis in tumor therapies (18–20), in addition to the sensitivity of various tumors to ferroptosis, such as ovarian cancer (21), hepatocellular carcinoma (22), and adrenocortical carcinomas (23). Notably, the anti-tumorigenesis effect of ferroptosis is likely propelled by the immune system. Wang et al. (24) reported that CD8+ T cells released interferon-gamma (IFNγ) could induce ferroptosis activity in cancer cells. On the contrary, ferroptosis-induced regulatory factors and the release of micromolecules may contribute to immunosuppression and tumor growth (25). Hence, the regulatory network between immune responses and ferroptosis as it relates to tumor immunotherapy remains unclear. In attempts to address this gap, several studies have suggested a correlation between ferroptosis regulators and antitumor drug sensitivity in treating OS (26–29).

In this study, we collected data from Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and Gene Expression Omnibus (GEO) to investigate the expression and prognostic value of ferroptosis regulators in OS patients. Risk signatures were constructed based on selected ferroptosis genes to evaluate the prognostic value of ferroptosis in risk stratification. Single-cell sequencing analysis was performed to explore the interaction between ferroptosis regulators and the immune microenvironment. Additionally, we investigated the predictive value of ferroptosis signature in anticancer chemotherapy. We further verified the cancer promotion function of pivotal genes HILPDA and MUC1 and revealed the probable association between them and ferroptosis. Therefore, this study aimed to comprehensively assess the effect of ferroptosis regulators on the immune microenvironment, prognosis, and therapeutic efficacy in OS.

Materials and methods

This study protocol was approved by the institutional review board (IRB) of the Third Xiangya Hospital, Central South University (No: 2020-S221). All experiments involving human tissues were performed based on guidelines approved by the IRB. Each sample was processed only after receiving a signed informed consent form.

Abbreviations: OS, osteosarcoma; TARGET, Therapeutically Applicable Research to Generate Effective Treatments; GEO, Gene Expression Omnibus; NMF, Non-negative matrix factorization; MAD, Median absolute deviation; PCA, Principal component analysis; GSVA, Gene set variation analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TME, tumor microenvironment; ssGSEA, single-sample gene-set enrichment analysis; ESTIMATE, Estimated Stromal and Immune cells in Malignant Tumor tissues using Expression; LASSO, least absolute shrinkage and selection operator; HR, Hazard ratio; UMAP, Uniform Manifold Approximation and Projection for Dimension Reduction; siRNA, Small interfering RNA; RT-qPCR, Real-time quantitative polymerase chain reaction; GSDC, Genomics of Drug Sensitivity in Cancer; IC50, half-maximal inhibitory concentration; ROC, Receiver operating characteristic; AUC, area under the curve; DEGs, differentially expressed genes; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ATF4, activating transcription factor 4; HILPDA, hypoxia inducible lipid droplet associated; ATM, ATM serine/threonine kinase; CBS, cystathionine beta-synthase; MUC1, mucin 1, cell surface associated; MT1G, metallothionein 1G; PML, PML nuclear body scaffold; ARNTL, aryl hydrocarbon receptor nuclear translocator like.

Data collection

Expression array profiling of 9 normal cell lines (5 normal osteoblast cells and 4 normal bone cells) and 103 patient-derived OS cell lines were extracted from GSE42352 (30) and GSE36001 on the GEO (https://www.ncbi.nlm.nih.gov/geo/). The batch effect was eliminated using the "removeBatchEffect" function in the R package "limma." Expression heatmaps were visualized with the "pheatmap" R package, while boxplot was constructed using the "ggpubr" R package. TARGET-OS RNA-seq data of 84 OS patients with available clinical characteristics extracted from the UCSC Xena website (https://xenabrowser.net/) were analyzed as the training cohort (Table S1). Furthermore, 53 OS samples extracted from GSE21257 (31) in the GEO database were validation cohorts (Table S2). In each cohort, we used the following criteria to exclude unqualified samples: (a) follow-up time < 1 month; (b) lack of survival data; (c) histopathological type is not OS. These count matrixes were standardized using the "DEseq2" package. Single-cell RNA sequencing datasets containing two primary OS lesions, "BC21" and "BC22", two metastatic OS lesions "BC10" and "BC17", and two recurrent OS lesions "BC11" and "BC20" were collected from GSE152048 (32) in GEO database. Ferroptosis regulators, including 108 driver genes and 69 suppressor genes, were obtained from the FerrDb website (Table S3) (http://www.zhounan.org/ferrdb) (33).

Non-negative matrix factorization clustering for ferroptosis regulators

One hundred seventy-three ferroptosis-related genes were extracted and analyzed in the TARGET-OS training cohort. Candidate regulators with a high median absolute deviation (MAD > 0.5) value across the OS patients were selected for subsequent NMF clustering analysis. Unsupervised NMF clustering was performed using the "NMF" R package based on the 132 candidate genes (34). When the coexistence correlation coefficient k = 2, we observed the clearest boundary and most appropriate consistency; thus, 84 patients were clustered into two subclusters. In addition, principal component analysis (PCA) was used to validate the subcluster distribution with the expression of candidate ferroptosis regulators.

Gene set variation analysis and functional annotation

To explore the difference between ferroptosis-related subclusters in biological processes, we conducted a GSVA enrichment analysis using the "GSVA" R package (35). Two gene sets, "c2.cp.kegg.v7.4.symbols" and "c5.go.bp.v7.4.symbols" were obtained from MSigDB database for performing GSVA enrichment. Moreover, Gene Ontology (GO) term enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, and annotation were also conducted with "clusterProfiler" and "org.Hs.eg.db" R packages. Finally, histograms were developed with the "ggplot2" R package.

Assessment of tumor microenvironment cell infiltration

We conducted a single-sample gene-set enrichment analysis (ssGSEA) algorithm to assess the expression abundance of 28 specific infiltrating immune cell types in the OS TME. Marker gene sets for these TME infiltrating immune cells were collected from previous studies, covering multiple immune cell types, including activated B cell, CD8+ T cell, macrophage, natural killer T cell, and others (36, 37). Estimated Stromal and Immune cells in Malignant Tumor tissues using Expression (ESTIMATE) analysis was performed using the "estimate" R package to evaluate the infiltration of stromal cells and immune cells. The ESTIMATE score based on stromal and immune scores was used to evaluate tumor purity (38), and Scatter diagrams were developed using the "ggplot2" R package.

Construction of ferroptosis risk signature

Based on the 132 ferroptosis regulators for NMF clustering, we identified 22 independent prognosis-related genes with univariate Cox regression analysis ($P < 0.05$). Then, the least absolute shrinkage and selection operator (LASSO) algorithm filtered out 11 ferroptosis regulators that met the minimum lambda value. Finally, stepwise multivariate Cox regression analysis confirmed 8 genes with optimal collinearity, and a risk signature was constructed. A risk score of each OS patient in the TARGET training cohort and GEO validation cohort was calculated with the following algorithm:

Risk score = 0.705×ATF4 + 0.503×ATM + 0.616×HILPDA + 0.323×MUC1 + 0.417×CBS + 0.238×MT1G + (-0.969)×ARNTL + (-0.553)×PML.

Hazard ratios (HRs) were used to distinguish protective (HR < 1) and risky elements (HR > 1). Forest plots were developed using the "ggplot2" R package.

Single-cell RNA sequencing analysis

scRNA-seq analysis was conducted as previously described (39, 40). All single-cell expression matrixes of primary, metastatic, and occurrent OS patients from GSE152048 were processed by the "Seurat" R package. Firstly, "NormalizedData"

was applied to normalize these expression data, then we performed "FindVariableFeatures" to identify the 1,000 most variable genes. After PCA with "RunPCA," we conducted a K-nearest neighbor graph *via* "FindNeighbors," while cells were combined with the "FindClusters" function. Subsequently, Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (41) was used for visualization. Moreover, we performed a "Single R" R package to annotate cells when feature genes for all concerned cell categories were obtained from reported studies (32). Then, the "FindMarkers" function was performed to find differentially expressed genes for identified risk clusters.

Immunohistochemistry

Five pairs of formalin-fixed paraffin-embedded OS tissue and para-carcinoma tissue blocks (all post-chemotherapy) from 5 patients with OS were made into 5 μm paraffin sections. IHC was performed following the Mouse/rabbit enhanced polymer method detection system (ZSGB-BIO, PV-9000, China). The slides were deparaffinized and rehydrated using xylene and gradient-concentration ethyl alcohol, followed by antigen retrieval with sodium citrate at 95°C. At room temperature, the slides were blocked using an endogenous peroxidase blocker for 10 min. Samples were incubated with primary antibodies against HILPDA (Proteintech, China) and MUC1 (Proteintech, China) overnight at 4°C, reaction enhancer for 20 min at 37°C, and enhanced enzyme-conjugated sheep anti-mouse/rabbit IgG polymer for 20 min at 37°C. Then the slides were stained with 3, 30-diaminobenzidine tetrahydrochloride (DAB) and counterstained with hematoxylin. Images were captured with a magnification of 20x.

Cells culture

Two osteosarcoma cell lines (U2OS and MNNG/HOS) were kindly provided by Procell Life Science & Technology Co., Ltd. U2OS and MNNG/HOS were correspondingly cultured in McCoy's 5A (Procell, China), and MEM (Procell, China), both supplemented with 10% fetal bovine serum (Gibco, USA) and 1% penicillin-streptomycin solution (Biosharp, China) at 37°C with saturated humidity and 5% CO2. The average time of culture medium exchange was 24-48h. The cells were digested with trypsin-EDTA (Gibco, USA) and passaged when cell adhesion exceeded 80% confluency.

Small interfering RNA transfection

Human HILPDA siRNA (si-HILPDA), MUC1 siRNA (si-MUC1), and their nonspecific control siRNA (si-NC) were synthesized by JTSBio (Wuhan, China). The siRNAs were transfected into cells using jetPRIME transfection reagent (Polyplus, France) following the manufacturer's protocol. The siRNAs sequences were listed in Table S4. RNA extraction and cell proliferation assay were performed 48h after transfection.

Western blot

A mixture of RIPA (Beyotime, China) and a final concentration of 1mM PMSF (Beyotime, China) was used to lyse cells for protein extraction. Loading Buffer (Biosharp, China) was added to the protein supernatant, and then the sample was boiled to denature the protein. Then proteins were separated using SDS–PAGE gel (Biosharp, China), transferred to PVDF membranes (Millipore, USA), and blocked in 5% skimmed milk for 1h. Then membranes were incubated overnight at 4°C with primary antibodies, including HILPDA (Proteintech, China), MUC1 (Proteintech, China), ASCL4 (Affinity, China), GPX4 (Affinity, China), xCT (Affinity, China) and GAPDH. The membranes were incubated with fluorophore-conjugated secondary antibody (LI-COR Corp, NE) the following day. Protein bands were captured with an enhanced LI-COR Odyssey infrared imaging system (LI-COR Corp, NE), and the protein levels were normalized to the GAPDH levels.

Real-time quantitative polymerase chain reaction

RT-qPCR primers are listed in Table S4. Total RNA from cultured cells was extracted using Rnafast200 (Fastagen, Japan), and cDNA was synthesized using HiScript II Q RT SuperMix for qPCR (Vazyme, China). ChamQ Universal SYBR qPCR Master Mix (Vazyme, China) was used to conduct RT-qPCR based on the manufacturer's protocol. All steps for RT-qPCR reaction were conducted as follows: initial denaturation at 95°C for 30s, one cycle; denaturation at 95°C for 10s, 40 cycles; dissolution curve at 95°C for 15s, 60°C for 60s, 95°C for 15s, one cycle. Gene expression levels were normalized to those of GAPDH and calculated using lg2–△△Ct method.

EdU incorporation assay

Proliferating OS cells were identified using the Click-iT Plus EdU Alexa Fluor 488 Imaging Kit (Invitrogen, USA), and cell nuclei were stained using Hoechst (Invitrogen, USA). Image Pro-Plus version 6.0 (Media Cybernetics, USA) was applied to calculate the counts and percentage of EdU-positive cells.

Cell migration assay

OS cell migration was assayed using a Transwell chamber (Corning, USA) with polycarbonic membranes (6.5 mm in diameter and 8 μm pore size). Cells in a serum-free medium were added into the upper chamber at the density of 5×10^5 cells/ml (200 μl/well), and an OS-conditioned medium with 10% FBS was added to the lower chamber. After incubating for 48h at 37°C, non-migrated cells on the membrane were removed with a cotton swab. Cells that penetrated to the lower surface were stained with 0.1% crystal violet. Then cells in five random fields per well were counted under 200×magnification as n=1 for the assay in triplicate.

Cell viability detection

The cells were seeded into 96-well plates at a density of 5,000 cells/well with specific-concentration RSL3 (Selleck, China). After 24h, 1/10 volume of CCK-8 reagent (Proteintech, USA) was added to the wells, and the absorbance value was detected at 450nm after 1h incubation at 37°C. The experiment was repeated three times.

Lipid reactive oxygen species detection

BODIPY 581/591 C11(Invitrogen, D3861, USA) with a final concentration of 2μM was used to detect intracellular and lipid cell membrane ROS. After incubation for 30min at dark 37°C, cells were digested with trypsin and resuspended by PBS to prepare a 300 μl cell suspension to determine lipid oxidation by Flow Cytometry. The fluorescence intensity of the FITC channel was measured by BD FACS Canto II (BODIPY 581/591 C11 at 590 nm in the non-oxidized state and 510 nm in the oxidized state). At least 10,000 cells were analyzed per sample, and data were analyzed using FlwoJo V10.

Potential therapeutic prediction value of ferroptosis signature

We extracted expression matrix and drug response data of over 1,000 cancer cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC, http://www.cancerrxgene.org/) database (42). Afterward, each cell line's risk scores were calculated by conducting a ferroptosis signature. Then, we performed the Spearman method to evaluate the correlation (Cor) between risk scores and half-maximal inhibitory concentration (IC50) of each cell line. $|\text{Cor}| > 0.2$ and $P < 0.05$ were considered statistically significant.

Statistical analysis

All bioinformatics statistical analyses and visualization were performed using R version 4.0.3 (https://www.r-project.org/), and the R script was listed in Supplementary "R_script". Kaplan–Meier and log-rank analysis using "survival" and "survminer" packages were applied to evaluate the survival comparison. Receiver operating characteristic (ROC) and the area under the curve (AUC) were conducted with the "Time ROC" R package. Spearman correlation analysis was applied to evaluate correlations among continuous variables. Wilcoxon and One-way Anova tests were used to compare the difference among groups. A Chi-square test was used to identify the significance of ferroptosis DEGs (differentially expressed genes) among all detected genes. Values in cell experiment are mean ± SD unless otherwise noted and analyzed using Graphpad Prism version 8.0.2.263. Furthermore, the Benjamini-Hochberg method was utilized to adjust p values in functional annotation. *P.adjust* < 0.05 was considered statistically significant.

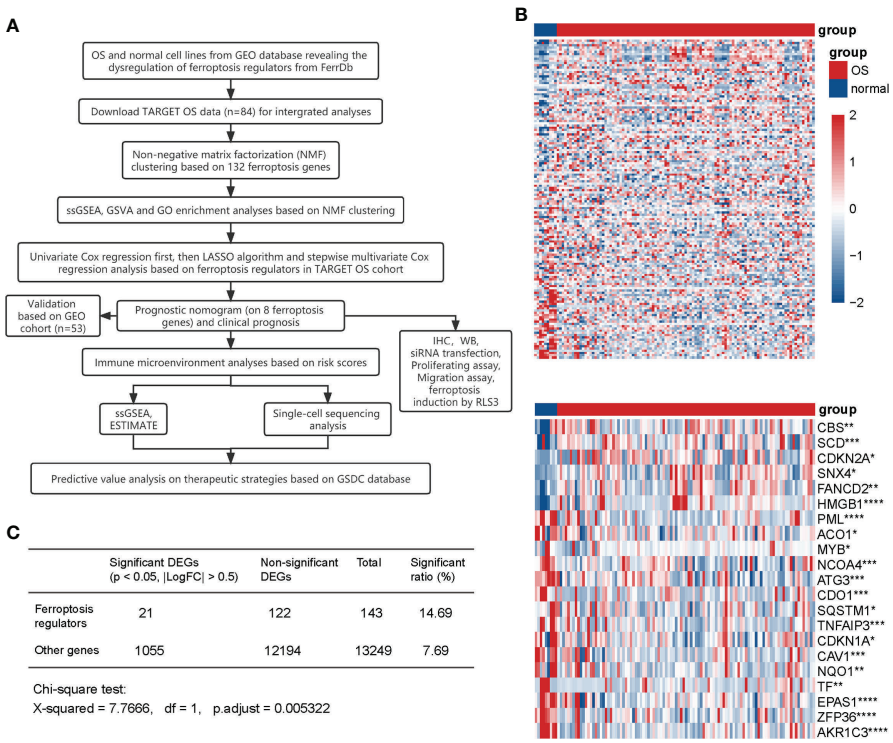
Results

Expression of ferroptosis genes was disordered in OS cells

A flow diagram was generated to systematically describe our study (Figure 1A). We collected 108 driver genes and 69 suppressor genes from FerrDb (http://www.zhounan.org/ferrdb), among which four genes were intersected, then 173 ferroptosis regulators were selected. Of the merged expression matrix containing 9 normal and 103 OS cell lines from GSE42352 and GSE36001, 143 of 173 ferroptosis regulators were detected. Subsequently, the expression of the 143 detected regulators were evaluated and visualized in heatmap, while 21 significant DEGs were identified ($P < 0.05$, $|\log FC| > 0.5$) (Figure 1B). CBS, SCD, CDKN2A, SNX4, FANCD2, and HMGB1 were upregulated in OS, and 15 regulators, including PML, ACO1, MYB, NCOA4, ATG3, CDO1, SQSTM1, TNFAIP3, CDKN1A, CAV1, NQO1, TF, EPAS1, ZFP36, and AKR1C3 were downregulated. The Chi-square test indicated that the ratio of significant ferroptosis-related DEGs was statistically higher than that of other genes (Figure 1C). Therefore, these results indicated that the expression of ferroptosis regulators was dysregulated in OS.

Ferroptosis regulators-based classification correlated with steosarcoma prognosis and immune microenvironment

We downloaded TARGET-OS gene expression profiles from UCSC Xena (https://xenabrowser.net/) and screened out 84



	Significant DEGs ($p < 0.05$, $\|LogFC\| > 0.5$)	Non-significant DEGs	Total	Significant ratio (%)
Ferroptosis regulators	21	122	143	14.69
Other genes	1055	12194	13249	7.69

FIGURE 1
Expression of ferroptosis regulators in normal and OS cell lines. **(A)** Flow diagram of the study. **(B)** Heatmaps of the expression of 143 detected ferroptosis regulators (up) and 21 significant DEGs (|LogFC| > 0.5, *P* <0.05) (down). Red represents high expression level and blue represents low expression. The darker the color, the greater the significance. **(C)** Chi-square test for the significance of ferroptosis DEGs. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

patients with analyzable prognostic information as a training cohort. Based on the previously mentioned 173 ferroptosis-regulated genes, a total of 132 genes selected with MAD value > 0.5 (43–45) were applied for NMF clustering analysis. Then, unsupervised NMF clustering was performed to assess potential gene expression features by dividing the original matrix into subclusters. A comprehensive correlation coefficient determined the most appropriate k value. Compared with heatmaps at k values of 3, 4, and 5 (Supplementary Figure 1A), k = 2 generated a heatmap that displayed the clearest boundary and best consistency in every subcluster (Figure 2A and Supplementary Figure 1B). Thus 84 patients were clustered into two subclusters, 50 patients in cluster one and 34 patients in cluster two. The heatmap displays 132 selected ferroptosis regulators' expression levels in clusters one and two (Figure 2B). PCA analysis was performed to verify the consistency of subcluster distribution (Figure 2C), which is highly consistent. Based on the clinical information of these patients in the TARGET cohort, survival analysis (Figure 2D) was constructed and revealed that cluster two OS patients exhibited poor survival outcomes compared with cluster one patients ($p < 0.001$). To investigate the 28 specific infiltrating immune cell types in tumor progression, ssGSEA was conducted (Figure 2E) and showed that cluster one was more positively correlated with immune cell infiltration than cluster two. The specific immune cells in this ssGSEA analysis included activated B cells, activated CD8 T cells, regulatory T cells, macrophages, NK cells, and others. Furthermore, based on the KEGG and GO databases, gene set variation analysis (GSVA) was performed to investigate the activation level of immune-related biological pathways in two subclusters (Supplementary Figures 2A, B). Our results demonstrated that cluster one is more relevant to various immune-related processes and pathways, such as NK cell-mediated cytotoxicity, primary immunodeficiency, T cell receptor signaling pathway, and regulation of macrophage fusion. Additionally, GO enrichment analysis was performed to comprehensively evaluate the biological characteristics in two ferroptosis-related subclusters and indicated that cluster one was closely correlated with immune-related activities (Supplementary Figure 2C), and cluster two was relevant to ion transmembrane channel activity and intercellular adhesion (Supplementary Figure 2D). In summary, these results suggest a significant difference in prognostic outcomes and biological characteristics within ferroptosis-related subclusters, and the difference in prognosis is highly correlated with the immune microenvironment.

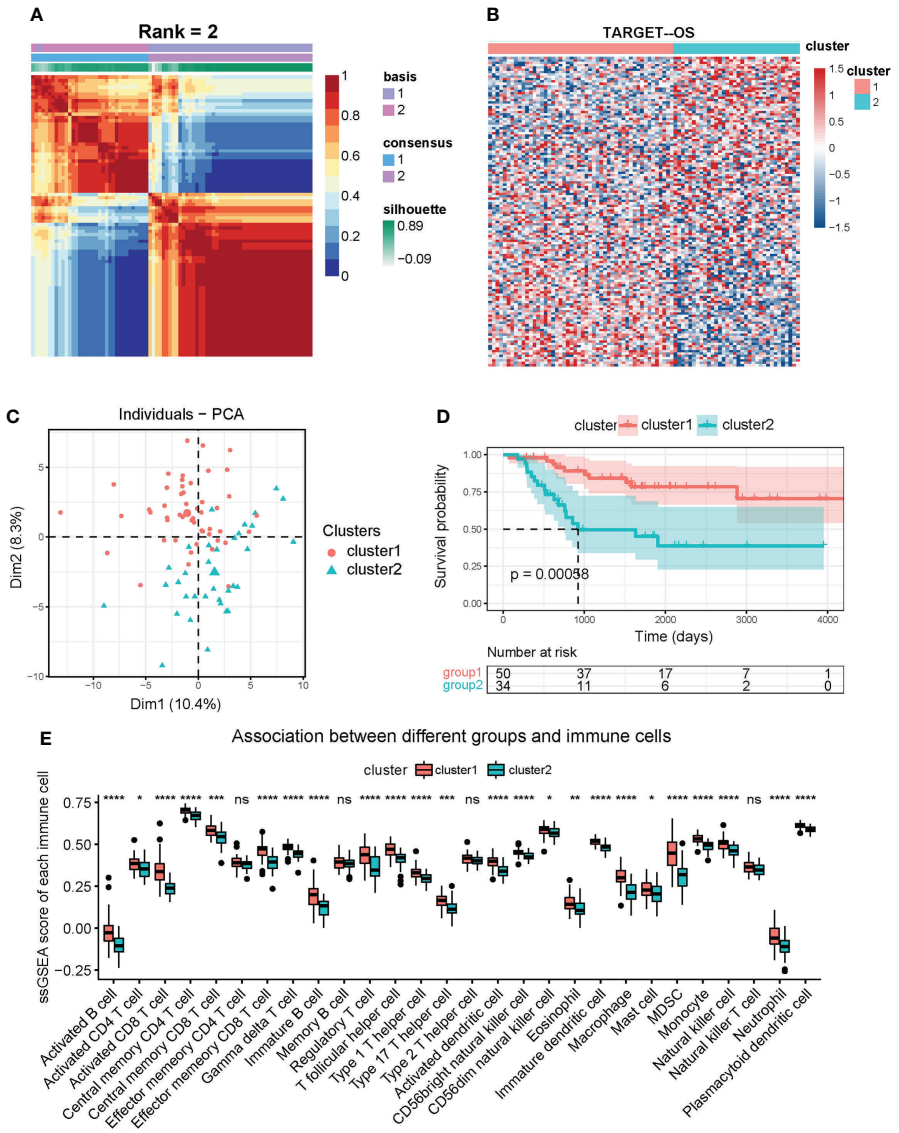


FIGURE 2
Identification of OS subclusters by unsupervised NMF clustering with ferroptosis regulators in TARGET OS cohort. **(A)** NMF clustering heatmap based on 132 ferroptosis genes (MAD > 0.5). NMF clustering divided 84 OS patients into two subclusters (we observed clearest boundary and most appropriate consistency when the coexistence correlation coefficient k = 2). **(B)** Heatmap of expression of the 132 ferroptosis regulators in two subclusters. Red represents high expression level and blue represents low expression. The darker the color, the greater the significance. **(C)** PCA scatter diagram supporting that NMF clustering algorithm divided OS patients into two subclusters. **(D)** Kaplan-Meier analysis for overall survival of OS patients in two subclusters. **(E)** Box diagram of ssGSEA analysis revealing expression of 28 immune cells in two subclusters. Kruskal test * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; ns, no significance.

Construction of prognostic signature based on ferroptosis regulators in osteosarcoma cohort

Cox regression analysis and the LASSO regression algorithm were conducted to determine the prognostic value of ferroptosis regulators in OS. Among the OS patients in the TARGET cohort, initially, 22 independent prognosis-related genes were confirmed by univariate Cox regression analysis ($P < 0.05$) (Figure 3A). Then, the LASSO algorithm filtered 11 ferroptosis regulators that met the minimum lambda value from the 22 genes (Figure 3B). Based on the LASSO results, stepwise multivariate Cox regression analysis was performed to construct a prognostic signature model (Figure 3C), which selected an optimal model containing eight genes: ATF4, HILPDA, ATM, CBS, MUC1, MT1G, PML, and ARNTL. Subsequently, every patient obtained a risk score calculated based on the eight regulators' regression

coefficients and expression levels. Patients were classified into high-risk and low-risk groups using the median risk score (46, 47). The Kaplan-Meier analysis revealed that patients in the high-risk group exhibit poor overall survival compared with low-risk group patients ($P < 0.0001$) (Figure 3D). The expression of the eight risk genes is shown in the heatmap (Figure 3E). The Scatter diagram displayed that the high-risk group correlated more with death incidents (Figure 3E). Time-dependent ROC (Figure 3F) indicated that the area under the curve (AUC) of 1-year, 2-year, 3-year, and 5-year survival was 0.881, 0.945, 0.886, and 0.858, respectively. Notably, it was most accurate for the risk score to predict 2-year survival. Additionally, Kaplan-Meier analyses based on these 8 genes respectively verified their potential to serve as independent prognosis factors (Supplementary Figures 3A–H). These results suggest the potential value of the constructed risk signature in predicting the prognosis of OS patients.

Moreover, ssGSEA analysis revealed that the high-risk group was likely to have less expression of immune cells, including activated B cells, macrophages, and NK cells (Figure 3G). In ESTIMATE analysis, Stroma, Immune, and ESTIMATE scores were prominently lower (T-test $P < 0.05$) in the high-risk group than those in the low-risk group (Figure 4A). Correlation analysis revealed that risk score was negatively correlated with Stromal, Immune, and ESTIMATE scores (Figure 4B). Relative expression of immune checkpoints in two risk groups was also visualized (Figure 4C), in which checkpoints PDCD1LG2, CD274, TIGIT, and CD40LG were observed at relatively low levels in a high-risk group. These results suggest that the risk score based on the ferroptosis prognostic signature was associated with immunosuppression and tumor progression.

In the validation set from GSE21257, a prognostic signature was executed to calculate risk scores. Then, 53 OS patients were divided into high-risk and low-risk groups using the median risk

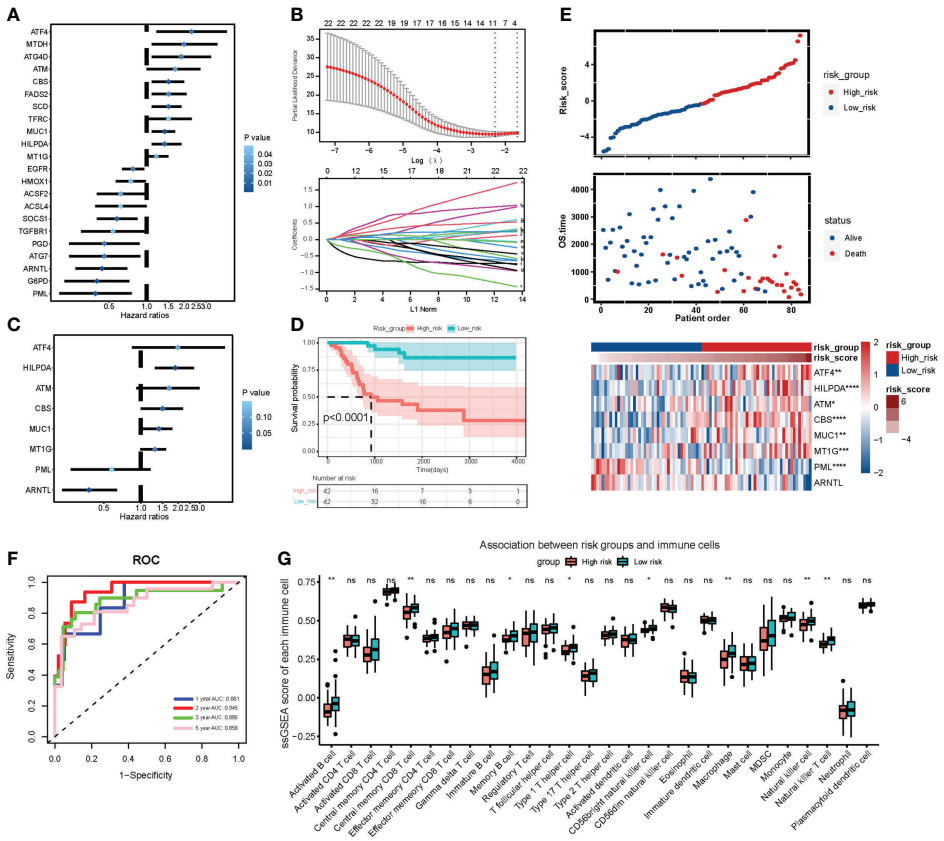


FIGURE 3
Construction and analysis of prognostic signature based on ferroptosis gene expression in TARGET OS cohort. **(A)** Forest plot of 22 independent prognostic genes identified by univariate Cox regression analysis ($P < 0.05$). Blue represents statistical significance. The deeper the blue, the greater the significance. **(B)** LASSO algorithm confirming minimum lambda value. **(C)** Forest plot of 8 regulators in optimal prognostic model selected by stepwise multivariate Cox regression analysis. **(D)** Kaplan-Meier analysis exhibiting the overall survival of OS patients in high-risk group and low-risk group graded by the optimal prognostic model. Red represents high risk group and blue represents low risk group. **(E)** Distribution plots of risk scores and heatmap of signature genes expression in TARGET OS patients. **(F)** Time dependent receiver operating characteristic (ROC) curve of the ferroptosis signature model in predicting prognosis of OS patients. **(G)** Box plot of ssGSEA analysis revealing 28 immune cells expression in two risk subgroups. Kruskal test * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; ns, no significance.

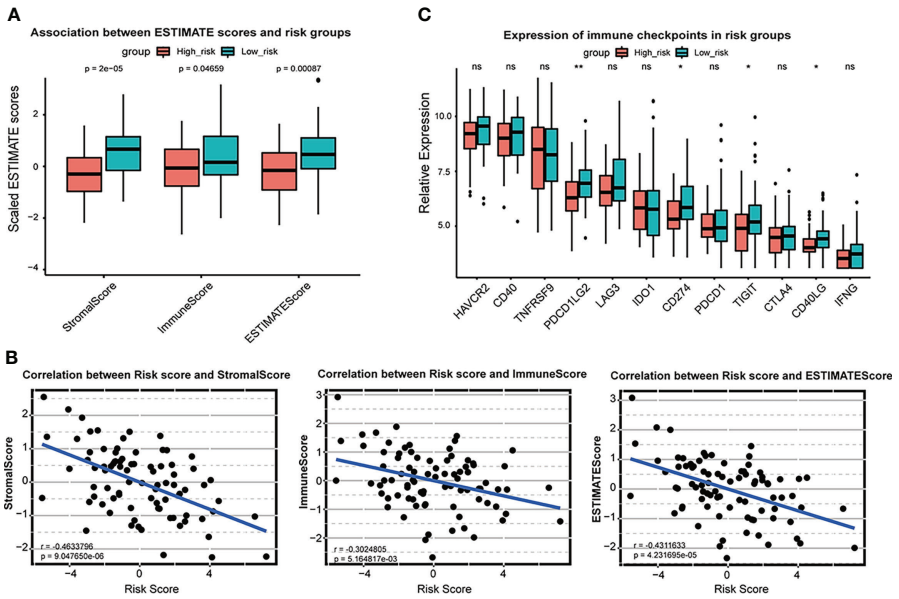


FIGURE 4
Association between ESTIMATE scores and risk scores in TARGET OS cohort. **(A)** Box plot of Stromal score, immune score and ESTIMATE score and in high-risk group and low-risk group. Red represents high risk group and blue represents low risk group. **(B)** Scatter plot shows correlation between risk score and stromal score, immune score and ESTIMATE score. **(C)** Box plot of relative expression of immune checkpoints in risk groups. * $P < 0.05$; ** $P < 0.01$; ns, no significance.

score. The heatmap demonstrated expression of the eight risk genes (Supplementary Figure 4A), and the scatter diagram indicated increased death incidents in the high-risk group (Supplementary Figure 4A). The Kaplan-Meier analysis revealed that high-risk patients possessed poor overall survival compared with low-risk patients ($P < 0.05$) (Supplementary Figure 4B). Additionally, time-dependent ROC (Supplementary Figure 4C) indicated that the area under the curve (AUC) of 1-year, 2-year, 3-year, 5-year, and 8-year survival was 0.658, 0.694, 0.754, 0.718, and 0.689, respectively.

Single cell sequencing investigated the relevance between risk stratification and immunity

To further investigate the correlation between ferroptosis risk signature and immune infiltration in OS, we collected scRNA-seq expression profile from GSE152048 on the GEO database, containing primary, lung metastatic, and recurrent OS lesions. Firstly, 16 cell subclusters were identified by "UMAP" dimension reduction in primary OS lesions expression profile (Supplementary Figures 5A, B). Expression proportions of the eight signature genes among all detected cells in primary OS samples were also visualized (Supplementary Figures 5C, 4D), in which we observed a relatively high proportion in the expression of ATF4, HILPDA, and ATM. All subclusters were annotated with feature genes and visualized into 12 cell clusters, including chondroblastic OS cells, endothelial cells, fibroblasts, M2 macrophages, myeloid cells, NK cells, osteoblastic OS cells, proliferating osteoblastic OS cells, T cells, and novel 1 and novel 2 (Figure 5A). Then, the prognostic signature was applied to calculate the risk scores of all cells and divided into high-risk and low-risk cells by median risk score. We found that chondroblastic OS cells, osteoblastic OS cells, proliferating osteoblastic OS cells, a subset of M2 macrophages, and myeloid cells were identified as high-risk cells, and immune cells, including T cells and NK cells, were identified as low-risk cells (Figure 5B). Subsequently, marker genes in high-risk cells and low-risk cells were distinguished by the "FindMarkers" function of the "Seurat" R package. KEGG enrichment analysis based on these markers indicated that high-risk cells were correlated with several cancer-related pathways, including oxidative phosphorylation, HIF-1 signaling pathway, and glycolysis/gluconeogenesis (Figure 5C). In contrast, low-risk cells were associated with immune-related pathways, including the T cell receptor signaling pathway, PD-L1 expression and PD-1 checkpoint pathway in cancer, NF-κB signaling pathway, NK cell-mediated cytotoxicity, and others (Figure 5D). Moreover, the low-risk group was correlated with ferroptosis and apoptosis (Figure 5D). Notably, these results support a risk score based on constructed prognostic signatures positively associated with tumor progression and negatively associated with immune

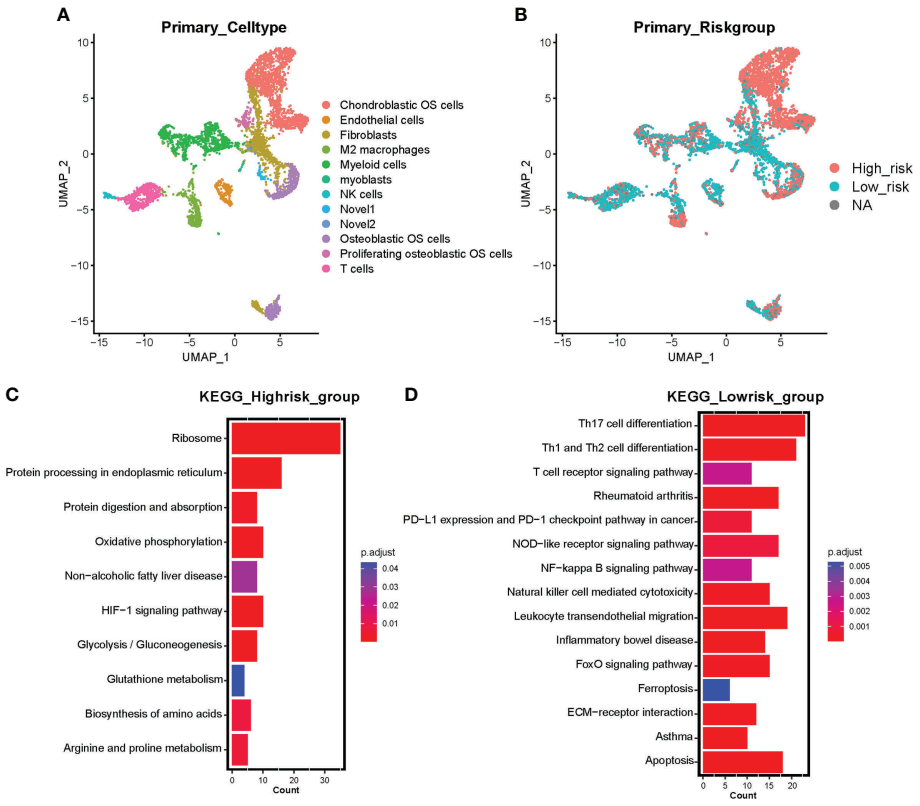


FIGURE 5
Single-cell sequencing investigating the correlation between risk signature and tumor microenvironment in primary OS samples. **(A)** UMAP visualization exhibits 12 annotated cell clusters based on primary OS single cell sequencing. **(B)** Risk cell clustering by ferroptosis signature clusters all cells into high-risk cells and low-risk cells. Red represents high-risk cells and blue represents low-risk cells. NA represents partial signature genes were not expressed in the single-cell sparse matrix. **(C)** KEGG enrichment analysis based on marker genes of high-risk cells. **(D)** KEGG enrichment analysis based on marker genes of low-risk cells. Color represents adjusted p value (Benjamini-Hochberg), the darker the red, the higher the significance; the darker the blue, the lower the significance.

infiltration and programmed cell death like ferroptosis and apoptosis in primary OS lesions.

Additionally, 13 cell subclusters were identified in metastatic OS lesions expression profile (Supplementary Figures 6A, B). The proportion diagram exhibited a markedly elevated expression level of MT1G compared with primary OS lesions (Supplementary Figures 6C, D). Moreover, all subclusters were annotated and visualized into 10 cell clusters, including chondroblastic OS cells, endothelial cells, fibroblasts, M2 macrophages, myoblast, NKT/T cells, osteoblastic OS cells, osteoclasts, proliferating osteoblastic OS cells and B cells (Supplementary Figure 6E). Then the prognostic signature was performed to calculate risk scores and divided all cells into high-risk and low-risk groups. Chondroblastic OS cells, osteoblastic OS cells, and proliferating osteoblastic OS cells were defined as high-risk cells, and B cells, NK T cells, and T cells were low-risk cells (Supplementary Figure 6F). Six-cell clusters were annotated and visualized in recurrence OS lesions, including chondroblastic OS cells, fibroblasts, myeloid cells, NKT/T cells, and novel cells. Prognostic signature classified all recurrence cells into high-risk and low-risk groups, indicating that chondroblastic OS cells, osteoblastic OS cells, and a part of fibroblasts were high-risk cells, and NKT/T cells were low-risk cells (Supplementary Figures 7A–F). The proportion diagram exhibited the up-regulation of HILPDA, MUC1, and MT1G in recurrence OS lesions. Therefore, these findings suggest a vital role of five cancer-promoting genes: ATF4, HILPDA, ATM, MUC1, and MT1G in affecting the OS progression, metastasis, and recurrence.

Knocking down of HILPDA or MUC1 significantly inhibited the proliferation of OS cells

We further analyzed the five cancer-promoting prognostic genes and found that ATF4 (48), ATM (49), and MT1G (50) have been reported in OS, while the functions of HILPDA and

MUC1 remained unclear. We then chose HILPDA and MUC1 as the following research subjects to illustrate their functions in OS. The expression levels of HILPDA (Figure 6A) and MUC1 (Figure 6B) were upregulated in OS tissues compared with paracancerous normal tissues. Then we used small interfering RNA to silence the expression of HILPDA and MUC1 in two OS cell lines. In U2OS cells, si-HILPDA sequence-2 and si-MUC1 sequence-3 had the best interference effect, while in MNNG/HOS cells, si-HILPDA sequence-3 and si-MUC1 sequence-1 were the optimal (Figures 6C, D). Correspondingly, compared to normal control groups, the percentages of Edu-positive OS cells and migrated cell numbers were significantly reduced in si-HILPDA and si-MUC1 groups (Figures 6E, F). The proliferation and migration of OS cells were inhibited considerably after interfering with HILPDA or MUC1 expression.

HILPDA and MUC1 influenced ferroptosis resistance of OS cells

Both HILPDA and MUC1 were reported ferroptosis-related regulators, but mechanisms of how they affect ferroptosis remain to be further investigated. In our subsequent experiments, we used gradient concentration of ferroptosis inducer RSL3 to treat U2OS and MNNG/HOS, and 24h later, the CCK-8 method was

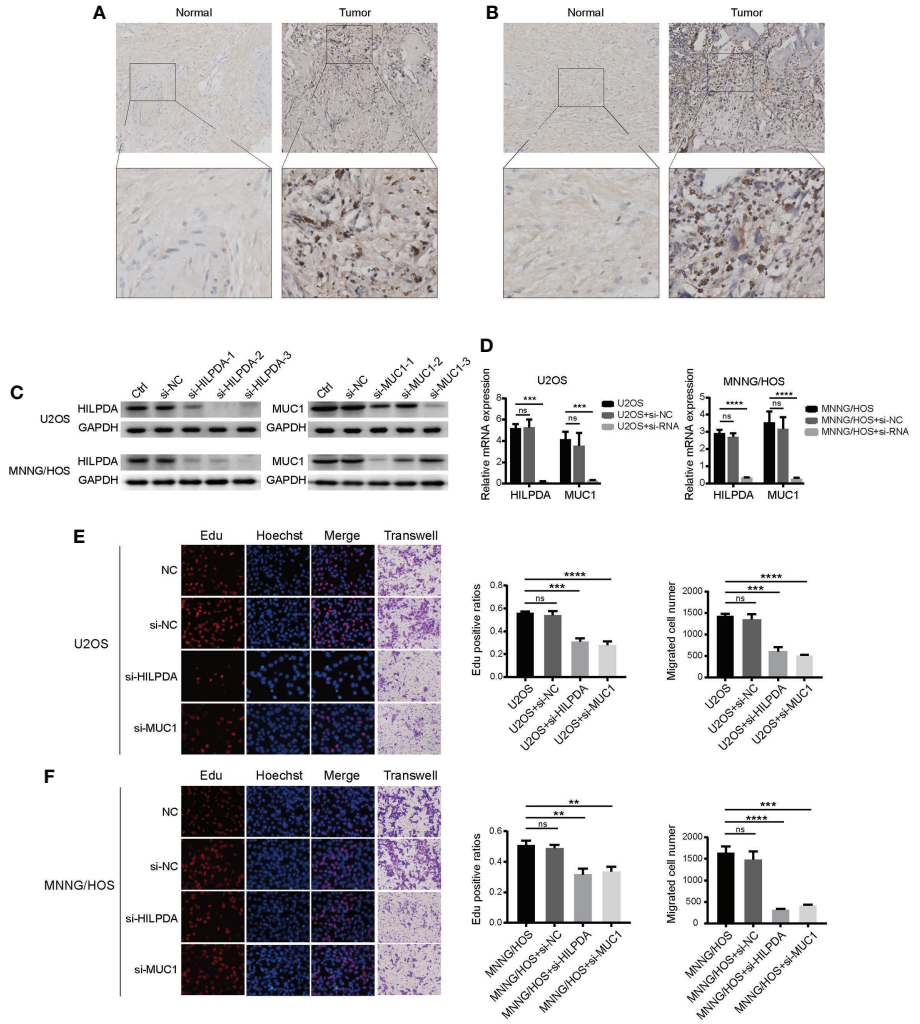


FIGURE 6
Expression of HILPDA or MUC1 influenced the proliferation of OS cells. Representative immunohistochemical images of expressions of HILPDA **(A)** and MUC1 **(B)** in OS and para-carcinoma tissues. **(C)** Relative protein levels of HILPDA and MUC1 after si-RNA transfection (n=3) in U2OS and MNNG/HOS. **(D)** Relative mRNA expression levels of HILPDA and MUC1 using the optimal si-RNA **(E, F)** Representative images of EdU (red), Hoechst staining (blue) and transwell (purple) in U2OS and MNNG/HOS cells after si-RNA transfection.The ratios of EdU-positive (red) cells and migration cell numbers were calculated (n=3) after si-RNA transfection. Student t test ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; ns, no significance.

used to detect the cell viability. Compared with the control group, the si-MUC1 group exhibited poor cell viability. The si-HILPDA group had a higher survival rate, with the greatest difference when RSL3 concentration was 4μM in U2OS (Figure 7A) and 8μM in MNNG/HOS (Figure 7B). Thus, U2OS with 4μM RSL3 and MNNG/HOS with 8μM treatment were used for subsequent experiments. Based on Flow Cytometry, the lipid ROS level was increased in the si-MUC1 group and decreased in the si-HILPDA group (Figures 7C, D), indicating RSL3-induced activity was correlated with lipid peroxidation, the marker of ferroptosis. We further assessed the levels of several ferroptosis-related proteins (Figures 7E, F). Among the control, si-HILPDA, and si-MUC1 groups, ASCL4 exhibited no significant difference, and xCT was decreased in the si-MUC1 groups. Intriguingly, GPX4 seemed to decrease in the si-MUC1 group of U2OS cells while slightly upregulated in the si-HILPDA group of MNNG/HOS cells. This finding might explain the earlier appearance of RSL3-induced ferroptosis in

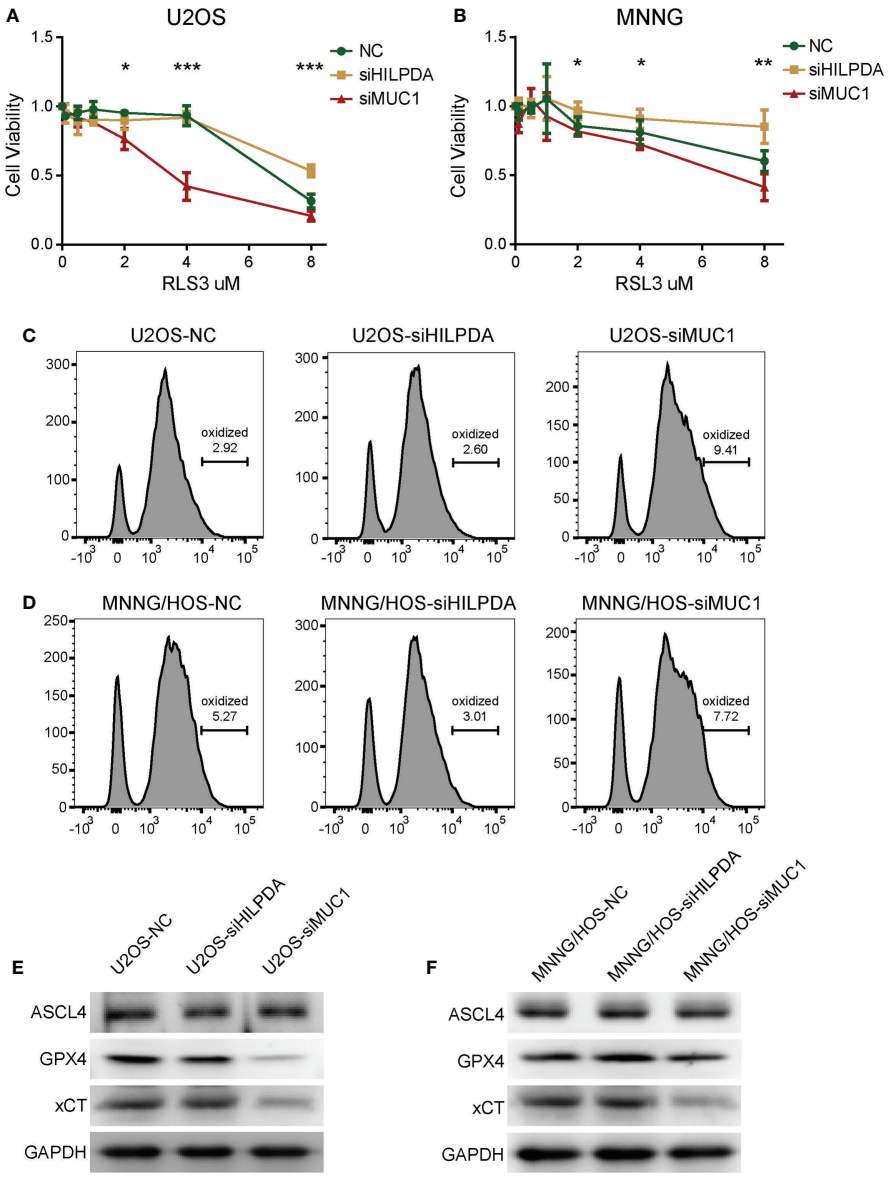


FIGURE 7
The effects HILPDA or MUC1 on ferroptosis resistance in OS cells. CCK-8 method detected the cell viability of U2OS **(A)** and MNNG/HOS **(B)** after treatment by different concerntrations of RSL3 for 24h. One-way Anova test *P< 0.05; **P< 0.01; ***P< 0.001. Lipid ROS levels of RSL3 induced U2OS **(C)** and MNNG/HOS **(D)** after C11 BODIPY incubation based on flow cytometry. Levels of ferroptosis-related proteins including ASCL4, GPX4 and xCT in RSL3 induced U2OS **(E)** and MNNG/HOS **(F)**.

si-MUC1 U2OS cells and the enhanced ferroptosis resistance in si-HILPDA MNNG/HOS cells.

Predictive value on therapeutic strategies of the prognostic signature

To explore the potential value of risk signature in predicting therapeutic strategies, based on the Genomics of Drug Sensitivity in Cancer (GSDC) database, spearman analysis was performed to investigate the correlation between half-maximal inhibitory concentration (IC50) of drugs and risk scores in cancer cell lines. A total of 32 drugs were identified to be significantly associated with the prognostic signature score ($|cor| > 0.15$, $P < 0.05$) (Figure 8A). Among them, drug sensitivity of 10 drugs were determined relevant to the score, including RTK signaling inhibitor BIBF 1120 (cor = -0.22, $P = 0.012$), PI3K/mTOR signaling inhibitor YM201636 (cor = -0.22, $P = 0.019$) and IGF1R signaling inhibitor Linsitinib (cor = -0.16, $P = 0.002$). However, drug resistance of 22 drugs were correlated with risk score, including cell cycle inhibitor CGP-60474 (cor = 0.31, $P = 0.003$), RTK signaling inhibitor Sunitinib (cor = 0.26, $P = 0.013$), and DNA replication inhibitor Bleomycin (cor = 0.23, $P = 0.003$). Additionally, targeted signaling pathways of these drugs were exhibited (Figures 8B, C) and indicated that drugs whose sensitivity was positively related to risk scores mostly target RTK signaling, kinases, IGF1R signaling, and ERK MAPK signaling. However, drugs whose resistance was positively related to risk scores targeted PI3K/mTOR signaling, ERK MAPK signaling, DNA replication, and cell cycle signaling. Therefore, established risk signatures might serve as potential guidance for establishing therapeutic strategies.

Discussion

Therapeutic schedules and outcomes of OS patients have remained significantly unimproved since the 1970s (3). Despite decades of research, molecular exploration still needs to diagnose the disease early, predict the progression and improve the prognosis for OS, especially for lung metastasis and chemotherapy resistance patients (51). Yanlong et al. found that focally amplified long noncoding RNA (lncRNA) expression on chromosome 1 (FAL1) was positively related to

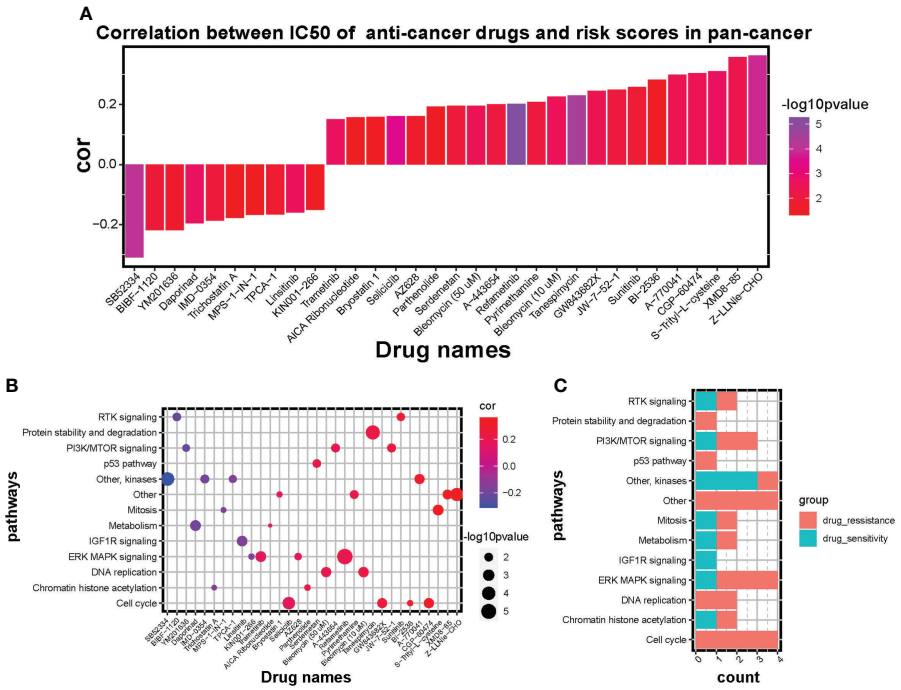


FIGURE 8
Influence of the risk signature on drug sensitivity and resistance in GSDC pan-cancer cell lines. **(A)** Bar diagram of correlation between IC50 of anti-cancer drugs and risk scores in pan-cancer. Altitude represents the correlation, higher the altitude, higher the correlation. Color represents statistical significance (p value), the more purple the color, the greater the significance. **(B)** Scatter diagram of correlation between targeted signaling pathways and IC50 of significant anti-cancer drugs. Size of plots represents statistical significance (p value), the larger the size, the greater the significance. Color of plots represents the correlation between targeted pathways and anti-cancer drugs. Red represents positive correlation and blue represents negative correlation. Purple represents little correlation. **(C)** Bar diagram shows the counts of sensitive drugs and resistant drugs regarding the targeted pathways.

the distance metastasis, tumor stage, and negatively prognosticate outcomes in OS patients (52). Wei et al. also showed that cyclin E1 was a promising prognostic and chemotherapeutic target for OS (53). These studies promoted the potential and significance of exploring molecular biomarkers for the onset and development of OS. In addition to these molecular biomarkers, ferroptosis has been considered a promising antitumor target mechanism in the occurrence and progression of numerous cancers (13, 54, 55). Evidence suggests multiple risk signatures based on ferroptosis genes could effectively predict the diagnosis, prognosis, immune microenvironment, and therapeutic strategies for cancers (56–58). However, few studies reported the correlation between ferroptosis mechanisms and OS progression.

Our study initially showed the dysregulation of ferroptosis regulators with normal and patient-derived OS cell lines, which was statistically significant by the Chi-square test (Figure 1). Whereas biological deviations existed between specific DEGs in cell lines and RNA sequencing results from OS patients, we set MAD > 0.5 as gene screening criteria for following NMF clustering analysis instead of using DEGs from OS cell lines. In doing so, two distinct subclusters were identified with different biological characteristics (Figure 2). OS patients in cluster1 exhibited a more favorable prognosis than those in cluster2, indicating that identified subclusters had significant clinical implications. Meanwhile, ssGSEA and GSVA for immune pathways and GO enrichment analyses suggested a higher degree of immune cell expression and immune response activity in cluster1 (Supplementary Figure 2). Existing studies indicate an association between immune response and tumor progression. Chi et al. (59) revealed that NK T cells promoted antitumor immunity in liver tumors. Mary et al. (60) found that the dysregulation of CD8 T cells would allow for tumor progression. Moreover, the Toll-like receptor signaling pathway benefits immune-related anticancer chemotherapy and radiotherapy (61). Our results are consistent with these dominant perceptions that the immune microenvironment's abundance correlates with better clinical outcomes. Given the above discovery, we speculated that cluster1, having a better prognosis, was more immune-activated than cluster2, and ferroptosis was involved in shaping the immune microenvironment in OS.

Considering the heterogeneity and complexity of individuals, we constructed a risk scoring system, "ferroptosis-based risk signature," to quantify the biological characteristics of OS patients (Figure 3). High-risk scores with worse clinical outcomes exhibited strong relevance to immunosuppression and lower stromal scores (Figure 4). The significance of immune and stromal scores in the ESTIMATE algorithm for tumor classification and clinical outcomes was already testified (62, 63). Hence we speculated that our constructed ferroptosis score was more significant in predicting immunosuppression than in predicting the stromal activation for OS malignancy. Moreover, immune checkpoints like PDCD1LG2, CD274, TIGIT, and CD40LG were upregulated in low-risk groups, reflecting the potential of immunotherapy in managing OS (Figure 4C). These results suggest that ferroptosis-based risk signature is reliable for comprehensively predicting the clinical prognosis, immune response activity, and therapeutic strategy for OS.

Among the 8 independent prognosis factors (Supplementary Figure 3) in the signature, ATF4, HILPDA, ATM, CBS, MUC1, and MT1G were significantly upregulated in the high-risk group, whereas PML was down-regulated (Figure 3E), implying that PML might serve as an antineoplastic factor in OS progression. Chen et al. (64) found that expression of activating transcription factor 4 (ATF4) promoted the malignancy of gliomas and fostered tumor angiogenesis and proliferation, while ATF4 knockdown made cells susceptible to ferroptosis. Hypoxia-inducible lipid droplet-associated (HILPDA) (65, 66) was overexpressed in multiple tumor types, HILPDA was positively correlated with tumor-associated macrophages (TAM) infiltration, and immunosuppressive genes, such as PD-L1, PD-1, TGFB1, and TGFBR1. Notably, Ataxia-Telangiectasia mutated protein (ATM) was reported as a positive regulator for ferroptosis (67). Radiotherapy-activated ATM and IFNγ from immunotherapy-activated CD8+ T cells would synergistically enhance ferroptosis and tumor lipid oxidation, indicating the correlation between ferroptosis agonists and chemoradiotherapy *via* immunotherapy for the first time (68). Li Wang et al. (69) found that inhibition of Cystathionine β-synthase (CBS) triggered ferroptosis in hepatocellular carcinoma and reduced tumor growth. Takahiro et al. (70) showed that the transmembrane mucin MUC1 contributed to immunologic escape in triple-negative breast cancer (TNBC) and that targeting MUC1-C correlated with PD-L1 suppression to activate the immune response and tumor cell killing. Emerging evidence suggests the crucial role of metallothioneins (MTs), including MT1G, in tumor formation, progression, and drug resistance (71). As a tumor suppressor, promyelocytic leukemia (PML) protein was mechanistically capable of inhibiting tumor proliferation, migration, and invasion while promoting cell senescence and apoptosis (72–74). A recent study also reported that ubiquitination of PML promotes lung cancer progression *via* fostering immunosuppression in the tumor microenvironment (75).

Single-cell sequencing analysis further investigated the role of ferroptosis signature in the tumor microenvironment and malignant cell proliferation of OS (Figure 5). Neoplastic cells and M2 macrophages were identified in the high-risk group, while immune cells were mostly identified in the low-risk group. Growing evidence has clarified the crucial role of TAMs in the progression and metastasis of tumors (76, 77). Additionally, Zhou et al. found the preventive effect of inhibiting M2 polarization of TAMs in OS metastasis (78). Moreover, previous work indicated that the infiltration degree of

intratumoral T cells was positively effective in predicting the prognosis of colorectal cancer, ovarian cancer, and melanoma patients (79–81). Existing research also reported the cytotoxic effect of NK cells against tumor progression in multiple cancers (82, 83). Our results implied that the prognostic signature could predict tumor invasion and progression from the M2 polarization of TAMs. The risk score was negatively correlated with anticancer immune cell infiltration in primary OS. However, in the high-risk group, several cancer-promoting pathways were enriched (Figure 5C). Oxidative phosphorylation is upregulated in multiple cancers, including leukemias, melanoma, pancreatic ductal adenocarcinoma lymphomas, and endometrial carcinoma (84). Similarly, high-rate glycolysis can promote tumor proliferation in an aerobic environment (85). Importantly, HIF-1 functions as a crucial signal by coordinating tumorigenesis-related transcription factors and signaling molecules (86); Ni et al. suggests that inhibition of HIF-1α would unleash the activity of tumor-infiltrating NK cells (87). In the low-risk group, immune-related pathways were enriched as expected, including the T cell receptor signaling pathway, PD-L1 expression and PD-1 checkpoint pathway in cancer, NF-κB signaling pathway, and NK cell-mediated cytotoxicity. However, ferroptosis and apoptosis were also correlated with low-risk cells (Figure 5D), suggesting that ferroptosis risk score was negatively relevant to ferroptosis occurrence and ferroptosis occurrence in OS cells associated with immune system activation.

Based on the above findings, we chose two prognostic genes to illustrate our results through functional experiments in OS cells. HILPDA and MUC1 expression were verified to be increased in OS tissues (Figures 6A, B), and we confirmed the knockdown of HILPDA or MUC1 could inhibit the proliferation and migration of OS cells (Figure 6). Notably, Hasegawa et al. reported that MUC1-C forms a complex with xCT, which interacts with xCT and thereby controls GSH levels (88) and that xCT activity drives the expression GPX4 (89). Our results showed that interference targeting MUC1 led to the decrease of xCT, and GPX4 also exhibited downregulation. Therefore, the decline in these two anti-ferroptosis proteins (89) might be the potential mechanism of weakened ferroptosis resistance in MUC1-knockdown cells (Figure 7). However, HILPDA-knockdown cells seemed to have enhanced ferroptosis resistance (Figure 7). Thus, the restraint in OS invasiveness regarding HILPDA knockdown is probably unrelated to the ferroptosis mechanism.

Adverse chemotherapy combined with surgical removal of OS lesions is the primary management strategy for OS patients (3), while chemoresistance has become a pivotal obstacle in improving the therapeutic effect (90). The interaction between ferroptosis and chemoresistance has recently been a topic of investigation, which Zhang et al. (91) reports that cisplatin and paclitaxel facilitated the secretion of miR-522 from cancer-associated fibroblasts, leading to ALOX15 suppression, ferroptosis inhibition, and ultimately chemoresistance. Our analysis for IC50 of anticancer drugs (Figure 8) showed the potential therapeutic efficiency of ferroptosis regulators. The ferroptosis risk score was correlated with sensitivity to drugs targeting RTK, IGF1R signaling, and kinases and with resistance to drugs targeting PI3K/mTOR, ERK/MAPK signaling, DNA replication, and cell cycle signaling. These results imply that patients with higher ferroptosis scores may benefit more from chemotherapy drugs targeting RTK, IGF1R signaling, and kinases. Ferroptosis regulators might be an adequate predictor for evaluating chemoradiotherapy's prognosis or targeted therapies. Therefore, our findings provided new probabilities for improving the management strategies for OS.

There are still some limitations in our study. Firstly, the data capacity for OS in public databases is significantly less than that for other tumor types, obstructing the exploration of OS bioinformatics research. To enlarge the sample capacity of the control group, we extracted control cell lines with inconsistent standards to accomplish the variation analysis, which could result in unpredictable biological deviations. More practicable sequencing data is yet to be discovered. Likewise, the interaction between stromal cell and ferroptosis signature remains unclear, as well as the major function of stromal cells in tumor progression and infiltration. Secondly, checkpoint PD-1/PD-L1 (CD274) has been reported as a pivotal mediator of immunosuppression in the tumor immune microenvironment (92, 93). Zheng et al. (94) demonstrated that PD-L1 was negatively associated with prognosis, while PD-L2 (PDCD1LG2) positively correlated with overall survival in OS. Given our contradictory result that the expression level of checkpoint CD274 was higher in the low-risk group, further inquiry about the molecular mechanisms of CD274 affecting ferroptosis signature and OS prognosis is needed, and PD-L1-related immune therapy on OS remains to be developed. Thirdly, ferroptosis-related gene signature for OS is not a novel subject. Lei et al. (95), Zhao et al. (96), Jiang et al. (97) all reported prognostic ferroptosis signatures, which might make our finding less novel. However, our study appears to be the first to reveal the correlation between immune landscape and ferroptosis signature from the perspective of a single-cell sequence. Notably, we are the first to propose the potential ferroptosis mechanism of specific genes, HILPDA and MUC1 regarding ferroptosis signature. From mechanistic investigations, we confirmed the cancer-promoting function of HILPDA and MUC1. However, the potential mechanisms or detailed pathways between HILPDA, MUC1, and ferroptosis require further exploration. Furthermore, the specific roles of the other six genes and their crosslinking remains to be explored. Generally, existing data and results could only support the predicting value of ferroptosis signature on OS progression, immune activity, and patient prognosis. The activation

mechanism of ferroptosis signature to intervene in the immune system is lacking. Therefore, more experiments are needed to explore the mechanism of ferroptosis signature in OS immunology.

Conclusions

In summary, our study comprehensively evaluated the expression pattern and prognostic value of ferroptosis regulators in OS. Our study's constructed prognostic model based on ferroptosis regulators is promising in predicting tumor progression, immune infiltration, and survival outcome of OS patients. Moreover, the risk stratification had a guidance value on chemoradiotherapy and might be correlated with the efficacy of immunotherapy. We also confirmed the cancer-promoting function of HILPDA and MUC1 and the ferroptosis-resistant related mechanism of MUC1 in OS, which suggested that MUC1 has the potential to become a ferroptosis-related therapeutic target. However, further exploration is necessary to reveal the potential mechanism among these genes in OS progression and therapeutic efficacy.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the institutional review board (IRB) of the Third Xiangya Hospital, Central South University (No: 2020-S221). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

Conception and design: XW, XC and SW. Foundation support: XC and SW. Acquisition and analysis of data: XW. Cell experiment and analysis: XW, LZ and GX. Interpretation of data: XW and SX. Drafting the manuscript and revising for submission quality: XW and XC. Reviewing and approving the final vision: XW, JH, WZ and XC. Study supervision: SW and XC. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the National Natural Science Foundation of China (82072501), Science and Technology Innovation Leading Plan of High-Tech Industry in Hunan Province (2020SK2011), Youth Fund Project of Natural Science Foundation of Hunan Province (2020JJ5848), and Medical Research Development Fund Project (WS865C).

Acknowledgments

The authors thank AiMi Academic Services (www.aimieditor.com) for the English language editing and review services.

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/fonc.2022.1024915/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Unsupervised consensus NMF clustering on TARGET cohort. **(A)** Heatmaps of NMF clustering for k = 3, 4, and 5. **(B)** The correlation among cophenetic, dispersion, evar, residuals, rss, silhouette and sparseness coefficients with reference to different cluster number.

SUPPLEMENTARY FIGURE 2
Enrichment characteristics in two ferroptosis subclusters. **(A, B)** GSVA analysis for NMF clustered ferroptosis subclusters based on KEGG database and GO database. Red represents high expression level and blue represents low expression. The darker the color, the greater the significance. **(C, D)** GO enrichment analysis including biological process, molecular function and cellular component based on high-expression

genes in cluster one and cluster two. The length of bars represents gene counts of GO terms. Color represents adjusted p value (Benjamini-Hochberg), the redder the color, the greater the significance.

SUPPLEMENTARY FIGURE 3
Kaplan–Meier survival analysis for TARGET OS patients based on the expression of ATF4 **(A)**, HILPDA **(B)**, ATM **(C)**, CBS **(D)**, MUC1 **(E)**, MT1G **(F)**, PML **(G)**, ARNTL **(H)**.

SUPPLEMENTARY FIGURE 4
Validation of the ferroptosis prognostic signature in GEO OS cohort. **(A)** Distribution plots of risk scores and heatmap of signature gene expression in GEO OS patients. Red represents high expression level and blue represents low expression. The darker the color, the greater the significance. **(B)** Kaplan-Meier analysis exhibiting the overall survival of GEO OS patients in high-risk group and low-risk group. **(C)** Time dependent ROC curve analysis of the ferroptosis signature model in predicting prognosis of GEO OS patients. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

SUPPLEMENTARY FIGURE 5
Single cell sequencing analysis of primary OS samples. **(A)** UMAP visualized 16 cell subclusters identified using "FindClusters" function. **(B)** Heatmap of top10 feature genes in 16 subclusters. Yellow represents high expression level of genes, purple represents low expression. **(C)** Expression of eight risk signature genes in all identified cells. Purple represents high expression of signature genes. The more purple the color, the higher the expression. **(D)** Expression proportion of eight signature genes among all detected cells in primary OS samples. Red represents the proportion of gene-positive cells and blue represents the proportion of gene-negative cells.

SUPPLEMENTARY FIGURE 6
Single cell sequencing analysis of metastatic OS samples. **(A)** UMAP visualized 13 cell subclusters identified in metastatic OS samples using "FindClusters" function. **(B)** Heatmap of top10 feature genes in the 13 subclusters. Yellow represents high expression level of genes, purple represents low expression. **(C)** Expression of eight risk signature genes in all identified cells. Purple represents high expression of signature genes. The more purple the color, the higher the expression. **(D)** Expression proportion of eight signature genes among all detected cells in metastatic OS samples. Red represents the proportion of gene-positive cells and blue represents the proportion of gene-negative cells. **(E)** UMAP visualization exhibits 10 annotated cell clusters based on metastatic OS single cell sequencing. **(F)** Risk cell clustering by ferroptosis signature clusters all cells into high-risk cells and low-risk cells. Red represents high-risk cells and blue represents low-risk cells. NA represents partial signature genes were not expressed in the single-cell sparse matrix.

SUPPLEMENTARY FIGURE 7
Single cell sequencing analysis of recurrent OS samples. **(A)** UMAP visualized 13 cell subclusters identified in recurrent OS samples using "FindClusters" function. **(B)** Heatmap of top10 feature genes in the 13 subclusters. Yellow represents high expression level of genes; purple represents low expression. **(C)** Expression of eight risk signature genes in all identified cells. Purple represents high expression of signature genes. The more purple the color, the higher the expression. **(D)** Expression proportion of eight signature genes among all detected cells in recurrent OS samples. Red represents the proportion of gene-positive cells and blue represents the proportion of gene-negative cells. **(E)** UMAP visualization exhibits 6 annotated cell clusters based on recurrent OS single cell sequencing. **(F)** Risk cell clustering by ferroptosis signature clusters all cells into high-risk cells and low-risk cells.

References

1. Kansara M, Teng MW, Smyth MJ, Thomas DM. Translational biology of osteosarcoma. *Nat Rev Canc* (2014) 14:722–35. doi: 10.1038/nrc3838

2. Ritter J, Bielack SS. Osteosarcoma. *Ann Oncol* (2010) 21(Suppl 7):vii320–5. doi: 10.1093/annonc/mdq276

3. Isakoff MS, Bielack SS, Meltzer P, Gorlick R. Osteosarcoma: Current treatment and a collaborative pathway to success. *J Clin Oncol* (2015) 33:3029–35. doi: 10.1200/JCO.2014.59.4895

4. Wang Z, Wang Z, Li B, Wang S, Chen T, Ye Z. Innate immune cells: A potential and promising cell population for treating osteosarcoma. *Front Immunol* (2019) 10:1114. doi: 10.3389/fimmu.2019.01114

5. Suehara Y, Alex D, Bowman A, Middha S, Zehir A, Chakravarty D, et al. Clinical genomic sequencing of pediatric and adult osteosarcoma reveals distinct molecular subsets with potentially targetable alterations. *Clin Cancer Res* (2019) 25:6346–56. doi: 10.1158/1078-0432.CCR-18-4032

6. Wang Y, Zhang H, Liu C, Wang Z, Wu W, Zhang N, et al. Immune checkpoint modulators in cancer immunotherapy: Recent advances and emerging concepts. *J Hematol Oncol* (2022) 15:111. doi: 10.1186/s13045-022-01325-0

7. Zhang N, Zhang H, Wu W, Zhou R, Li S, Wang Z, et al. Machine learning-based identification of tumor-infiltrating immune cell-associated lncRNAs for improving outcomes and immunotherapy responses in patients with low-grade glioma. *Theranostics.* (2022) 12:5931–48. doi: 10.7150/thno.74281

8. Le Cesne A, Marec-Berard P, Blay JY, Gaspar N, Bertucci F, Penel N, et al. Programmed cell death 1 (PD-1) targeting in patients with advanced osteosarcomas: results from the PEMBROSARC study. *Eur J Canc* (2019) 119:151–7. doi: 10.1016/j.ejca.2019.07.018

9. Tawbi HA, Burgess M, Bolejack V, Van Tine BA, Schuetze SM, Hu J, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol* (2017) 18:1493–501. doi: 10.1016/S1470-2045(17)30624-1

10. Kim JR, Moon YJ, Kwon KS, Bae JS, Wagle S, Kim KM, et al. Tumor infiltrating PD1-positive lymphocytes and the expression of PD-L1 predict poor prognosis of soft tissue sarcomas. *PLoS One* (2013) 8:e82870. doi: 10.1371/journal.pone.0082870

11. Pollack SM, He Q, Yearley JH, Emerson R, Vignali M, Zhang Y, et al. T-Cell infiltration and clonality correlate with programmed cell death protein 1 and programmed death-ligand 1 expression in patients with soft tissue sarcomas. *Cancer.* (2017) 123:3291–304. doi: 10.1002/cncr.30726

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

13. Mou Y, Wang J, Wu J, He D, Zhang C, Duan C, et al. Ferroptosis, a new form of cell death: opportunities and challenges in cancer. *J Hematol Oncol* (2019) 12:34. doi: 10.1186/s13045-019-0720-y

14. Liu H, Schreiber SL, Stockwell BR. Targeting dependency on the GPX4 lipid peroxide repair pathway for cancer therapy. *Biochemistry.* (2018) 57:2059–60. doi: 10.1021/acs.biochem.8b00307

15. Han B, Shen Y, Zhang P, Jayabal P, Che R, Zhang J, et al. Overlooked FANCD2 variant encodes a promising, portent tumor suppressor, and alternative polyadenylation contributes to its expression. *Oncotarget.* (2017) 8:22490–500. doi: 10.18632/oncotarget.14989

16. Junttila MR, Evan GI. p53–a jack of all trades but master of none. *Nat Rev Canc* (2009) 9:821–9. doi: 10.1038/nrc2728

17. Arrigo AP, Gibert B. HspB1 dynamic phospho-oligomeric structure dependent interactome as cancer therapeutic target. *Curr Mol Med* (2012) 12:1151–63. doi: 10.2174/156652412803306693

18. Gan B. DUBbing ferroptosis in cancer cells. *Cancer Res* (2019) 79:1749–50. doi: 10.1158/0008-5472.CAN-19-0487

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

20. Stockwell BR, Jiang X. A physiological function for ferroptosis in tumor suppression by the immune system. *Cell Metab* (2019) 30:14–5. doi: 10.1016/j.cmet.2019.06.012

21. Carbone M, Melino G. Stearoyl CoA desaturase regulates ferroptosis in ovarian cancer offering new therapeutic perspectives. *Cancer Res* (2019) 79:5149–50. doi: 10.1158/0008-5472.CAN-19-2453

22. Liu Y, Zhang X, Zhang J, Tan J, Li J, Song Z. Development and validation of a combined ferroptosis and immune prognostic classifier for hepatocellular carcinoma. *Front Cell Dev Biol* (2020) 8:596679. doi: 10.3389/fcell.2020.596679

23. Belavgeni A, Bornstein SR, von Massenhausen A, Tonnus W, Stumpf J, Meyer C, et al. Exquisite sensitivity of adrenocortical carcinomas to induction of ferroptosis. *Proc Natl Acad Sci U S A* (2019) 116:22269–74. doi: 10.1073/pnas.1912700116

24. Wang W, Green M, Choi JE, Gijon M, Kennedy PD, Johnson JK, et al. CD8(+) T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature.* (2019) 569:270–4. doi: 10.1038/s41586-019-1170-y

25. Friedmann Angeli JP, Krysko DV, Conrad M. Ferroptosis at the crossroads of cancer-acquired drug resistance and immune evasion. *Nat Rev Canc* (2019) 19:405–14. doi: 10.1038/s41568-019-0149-1

26. Liu Q, Wang K. The induction of ferroptosis by impairing STAT3/Nrf2/GPx4 signaling enhances the sensitivity of osteosarcoma cells to cisplatin. *Cell Biol Int* (2019) 43:1245–56. doi: 10.1002/cbin.11121

27. Lv H, Zhen C, Liu J, Shang P. Beta-phenethyl isothiocyanate induces cell death in human osteosarcoma through altering iron metabolism, disturbing the redox balance, and activating the MAPK signaling pathway. *Oxid Med Cell Longev* (2020) 2020:5021983. doi: 10.1155/2020/5021983

28. Fu J, Li T, Yang Y, Jiang L, Wang W, Fu L, et al. Activatable nanomedicine for overcoming hypoxia-induced resistance to chemotherapy and inhibiting tumor growth by inducing collaborative apoptosis and ferroptosis in solid tumors. *Biomaterials.* (2021) 268:120537. doi: 10.1016/j.biomaterials.2020.120537

29. Lin H, Chen X, Zhang C, Yang T, Deng Z, Song Y, et al. EF24 induces ferroptosis in osteosarcoma cells through HMOX1. *BioMed Pharmacother* (2021) 136:111202. doi: 10.1016/j.biopha.2020.111202

30. Kuijjer ML, Peterse EF, van den Akker BE, Briaire-de Bruijn IH, Serra M, Meza-Zepeda LA, et al. IR/IGF1R signaling as potential target for treatment of high-grade osteosarcoma. *BMC Canc* (2013) 13:245. doi: 10.1186/1471-2407-13-245

31. Buddingh EP, Kuijjer ML, Duim RA, Burger H, Agelopoulos K, Myklebost O, et al. Tumor-infiltrating macrophages are associated with metastasis suppression in high-grade osteosarcoma: a rationale for treatment with macrophage activating agents. *Clin Cancer Res* (2011) 17:2110–9. doi: 10.1158/1078-0432.CCR-10-2047

32. Zhou Y, Yang D, Yang Q, Lv X, Huang W, Zhou Z, et al. Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat Commun* (2020) 11:6322. doi: 10.1038/s41467-020-20059-6

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

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

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

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

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

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

39. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM3rd, et al. Comprehensive integration of single-cell data. *Cell.* (2019) 177:1888–902.e21. doi: 10.1016/j.cell.2019.05.031

40. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* (2019) 20:163–72. doi: 10.1038/s41590-018-0276-y

41. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* (2019) 37:38–44. doi: 10.1038/nbt.4314

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

43. She Y, Kong X, Ge Y, Yin P, Liu Z, Chen J, et al. Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma. *Cancer Cell Int* (2020) 20:22. doi: 10.1186/s12935-020-1104-7

44. Hunt AL, Bateman NW, Barakat W, Makohon-Moore S, Hood BL, Conrads KA, et al. Extensive three-dimensional intratumor proteomic heterogeneity revealed by multiregion sampling in high-grade serous ovarian tumor specimens. *iScience.* (2021) 24:102757. doi: 10.1016/j.isci.2021.102757

45. Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtroder K, et al. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. *Nat Commun* (2021) 12:2301. doi: 10.1038/s41467-021-22465-w

46. Chen H, Yao J, Bao R, Dong Y, Zhang T, Du Y, et al. Cross-talk of four types of RNA modification writers defines tumor microenvironment and pharmacogenomic landscape in colorectal cancer. *Mol Canc* (2021) 20:29. doi: 10.1186/s12943-021-01322-w

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

48. Luo J, Xia Y, Yin Y, Luo J, Liu M, Zhang H, et al. ATF4 destabilizes RET through nonclassical GRP78 inhibition to enhance chemosensitivity to bortezomib in human osteosarcoma. *Theranostics.* (2019) 9:6334–53. doi: 10.7150/thno.36818

49. Shen TS, Hsu YK, Huang YF, Chen HY, Hsieh CP, Chen CL. Licochalcone a suppresses the proliferation of osteosarcoma cells through autophagy and ATM-Chk2 activation. *Molecules.* (2019) 24(13):2435. doi: 10.3390/molecules24132435

50. Trost Z, Trebse R, Prezelj J, Komadina R, Logar DB, Marc J. A microarray based identification of osteoporosis-related genes in primary culture of human osteoblasts. *Bone.* (2010) 46:72–80. doi: 10.1016/j.bone.2009.09.015

51. Verrecchia F, Redini F. Transforming growth factor-beta signaling plays a pivotal role in the interplay between osteosarcoma cells and their microenvironment. *Front Oncol* (2018) 8:133. doi: 10.3389/fonc.2018.00133

52. Wang Y, Zhao Z, Zhang S, Li Z, Li D, Yang S, et al. LncRNA FAL1 is a negative prognostic biomarker and exhibits pro-oncogenic function in osteosarcoma. *J Cell Biochem* (2018) 119:8481–9. doi: 10.1002/jcb.27074

53. Wei R, Thanindratarn P, Dean DC, Hornicek FJ, Guo W, Duan Z. Cyclin E1 is a prognostic biomarker and potential therapeutic target in osteosarcoma. *J Orthop Res* (2020) 38:1952–64. doi: 10.1002/jor.24659

54. Lei G, Zhang Y, Koppula P, Liu X, Zhang J, Lin SH, et al. The role of ferroptosis in ionizing radiation-induced cell death and tumor suppression. *Cell Res* (2020) 30:146–62. doi: 10.1038/s41422-019-0263-3

55. Liu J, Zhang C, Wang J, Hu W, Feng Z. The regulation of ferroptosis by tumor suppressor p53 and its pathway. *Int J Mol Sci* (2020) 21(21):8387. doi: 10.3390/ijms21218387

56. Liang JY, Wang DS, Lin HC, Chen XX, Yang H, Zheng Y, et al. A novel ferroptosis-related gene signature for overall survival prediction in patients with hepatocellular carcinoma. *Int J Biol Sci* (2020) 16:2430–41. doi: 10.7150/ijbs.45050

57. Tang B, Zhu J, Li J, Fan K, Gao Y, Cheng S, et al. The ferroptosis and iron-metabolism signature robustly predicts clinical diagnosis, prognosis and immune microenvironment for hepatocellular carcinoma. *Cell Commun Signal* (2020) 18:174. doi: 10.1186/s12964-020-00663-1

58. Wu G, Wang Q, Xu Y, Li Q, Cheng L. A new survival model based on ferroptosis-related genes for prognostic prediction in clear cell renal cell carcinoma. *Aging (Albany NY)* (2020) 12:14933–48. doi: 10.18632/aging.103553

59. Ma C, Han M, Heinrich B, Fu Q, Zhang Q, Sandhu M, et al. Gut microbiome-mediated bile acid metabolism regulates liver cancer *via* NKT cells. *Science* (2018) 360(6391):eaan5931. doi: 10.1126/science.aan5931

60. Philip M, Fairchild L, Sun L, Horste EL, Camara S, Shakiba M, et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature.* (2017) 545:452–6. doi: 10.1038/nature22367

61. Apetoh L, Ghiringhelli F, Tesniere A, Obeid M, Ortiz C, Criollo A, et al. Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. *Nat Med* (2007) 13:1050–9. doi: 10.1038/nm1622

62. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY)* (2018) 10:592–605. doi: 10.18632/aging.101415

63. Deng X, Lin D, Zhang X, Shen X, Yang Z, Yang L, et al. Profiles of immune-related genes and immune cell infiltration in the tumor microenvironment of diffuse lower-grade gliomas. *J Cell Physiol* (2020) 235:7321–31. doi: 10.1002/jcp.29633

64. Chen D, Fan Z, Rauh M, Buchfelder M, Eyupoglu IY, Savaskan N. ATF4 promotes angiogenesis and neuronal cell death and confers ferroptosis in a xCT-dependent manner. *Oncogene.* (2017) 36:5593–608. doi: 10.1038/onc.2017.146

65. Liu C, Zhou X, Zeng H, Wu D, Liu L. HILPDA is a prognostic biomarker and correlates with macrophage infiltration in pan-cancer. *Front Oncol* (2021) 11:597860. doi: 10.3389/fonc.2021.597860

66. VandeKopple MJ, Wu J, Auer EN, Giaccia AJ, Denko NC, Papandreou I. HILPDA regulates lipid metabolism, lipid droplet abundance, and response to microenvironmental stress in solid tumors. *Mol Cancer Res* (2019) 17:2089–101. doi: 10.1158/1541-7786.MCR-18-1343

67. Chen PH, Wu J, Ding CC, Lin CC, Pan S, Bossa N, et al. Kinome screen of ferroptosis reveals a novel role of ATM in regulating iron metabolism. *Cell Death Differ* (2020) 27:1008–22. doi: 10.1038/s41418-019-0393-7

68. Lang X, Green MD, Wang W, Yu J, Choi JE, Jiang L, et al. Radiotherapy and immunotherapy promote tumoral lipid oxidation and ferroptosis *via* synergistic repression of SLC7A11. *Cancer Discovery* (2019) 9:1673–85. doi: 10.1158/2159-8290.CD-19-0338

69. Wang L, Cai H, Hu Y, Liu F, Huang S, Zhou Y, et al. A pharmacological probe identifies cystathionine beta-synthase as a new negative regulator for ferroptosis. *Cell Death Dis* (2018) 9:1005. doi: 10.1038/s41419-018-1063-2

70. Maeda T, Hiraki M, Jin C, Rajabi H, Tagde A, Alam M, et al. MUC1-c induces PD-L1 and immune evasion in triple-negative breast cancer. *Cancer Res* (2018) 78:205–15. doi: 10.1158/0008-5472.CAN-17-1636

71. Si M, Lang J. The roles of metallothioneins in carcinogenesis. *J Hematol Oncol* (2018) 11:107. doi: 10.1186/s13045-018-0645-x

72. Salomoni P, Ferguson BJ, Wyllie AH, Rich T. New insights into the role of PML in tumour suppression. *Cell Res* (2008) 18:622–40. doi: 10.1038/cr.2008.58

73. Sahin U, Lallemand-Breitenbach V, de The H. PML nuclear bodies: regulation, function and therapeutic perspectives. *J Pathol* (2014) 234:289–91. doi: 10.1002/path.4426

74. Bernardi R, Guernah I, Jin D, Grisendi S, Alimonti A, Teruya-Feldstein J, et al. PML inhibits HIF-1alpha translation and neoangiogenesis through repression of mTOR. *Nature.* (2006) 442:779–85. doi: 10.1038/nature05029

75. Wang YT, Chen J, Chang CW, Jen J, Huang TY, Chen CM, et al. Ubiquitination of tumor suppressor PML regulates prometastatic and immunosuppressive tumor microenvironment. *J Clin Invest* (2017) 127:2982–97. doi: 10.1172/JCI89957

76. Wang D, Wang X, Si M, Yang J, Sun S, Wu H, et al. Exosome-encapsulated miRNAs contribute to CXCL12/CXCR4-induced liver metastasis of colorectal cancer by enhancing M2 polarization of macrophages. *Cancer Lett* (2020) 474:36–52. doi: 10.1016/j.canlet.2020.01.005

77. Sa JK, Chang N, Lee HW, Cho HJ, Ceccarelli M, Cerulo L, et al. Transcriptional regulatory networks of tumor-associated macrophages that drive malignancy in mesenchymal glioblastoma. *Genome Biol* (2020) 21:216. doi: 10.1186/s13059-020-02140-x

78. Zhou Q, Xian M, Xiang S, Xiang D, Shao X, Wang J, et al. All-trans retinoic acid prevents osteosarcoma metastasis by inhibiting M2 polarization of tumor-associated macrophages. *Cancer Immunol Res* (2017) 5:547–59. doi: 10.1158/2326-6066.CIR-16-0259

79. van der Leun AM, Thommen DS, Schumacher TN. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat Rev Canc* (2020) 20:218–32. doi: 10.1038/s41568-019-0235-4

80. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* (2006) 313:1960–4. doi: 10.1126/science.1129139

81. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N Engl J Med* (2003) 348:203–13. doi: 10.1056/NEJMoa020177

82. Chen EB, Zhou ZJ, Xiao K, Zhu GQ, Yang Y, Wang B, et al. The miR-561-5p/CX3CL1 signaling axis regulates pulmonary metastasis in hepatocellular carcinoma involving CX3CR1(+) natural killer cells infiltration. *Theranostics.* (2019) 9:4779–94. doi: 10.7150/thno.32543

83. Lopez-Soto A, Gonzalez S, Smyth MJ, Galluzzi L. Control of metastasis by NK cells. *Cancer Cell* (2017) 32:135–54. doi: 10.1016/j.ccell.2017.06.009

84. Ashton TM, McKenna WG, Kunz-Schughart LA, Higgins GS. Oxidative phosphorylation as an emerging target in cancer therapy. *Clin Cancer Res* (2018) 24:2482–90. doi: 10.1158/1078-0432.CCR-17-3070

85. Wang Z, Dong C. Gluconeogenesis in cancer: Function and regulation of PEPCK, FBPase, and G6Pase. *Trends Canc* (2019) 5:30–45. doi: 10.1016/j.trecan.2018.11.003

86. Balamurugan K. HIF-1 at the crossroads of hypoxia, inflammation, and cancer. *Int J Canc* (2016) 138:1058–66. doi: 10.1002/ijc.29519

87. Ni J, Wang X, Stojanovic A, Zhang Q, Wincher M, Buhler L, et al. Single-cell RNA sequencing of tumor-infiltrating NK cells reveals that inhibition of transcription factor HIF-1alpha unleashes NK cell activity. *Immunity.* (2020) 52:1075–87.e8. doi: 10.1016/j.immuni.2020.05.001

88. Hasegawa M, Takahashi H, Rajabi H, Alam M, Suzuki Y, Yin L, et al. Functional interactions of the cystine/glutamate antiporter, CD44v and MUC1-c oncoprotein in triple-negative breast cancer cells. *Oncotarget.* (2016) 7:11756–69. doi: 10.18632/oncotarget.7598

89. Lee N, Carlisle AE, Peppers A, Park SJ, Doshi MB, Spears ME, et al. xCT-driven expression of GPX4 determines sensitivity of breast cancer cells to ferroptosis inducers. *Antioxidants (Basel)* (2021) 10(2):317. doi: 10.3390/antiox10020317

90. Chen R, Wang G, Zheng Y, Hua Y, Cai Z. Drug resistance-related microRNAs in osteosarcoma: Translating basic evidence into therapeutic strategies. *J Cell Mol Med* (2019) 23:2280–92. doi: 10.1111/jcmm.14064

91. Zhang H, Deng T, Liu R, Ning T, Yang H, Liu D, et al. CAF secreted miR-522 suppresses ferroptosis and promotes acquired chemo-resistance in gastric cancer. *Mol Canc* (2020) 19:43. doi: 10.1186/s12943-020-01168-8

92. Chen S, Crabill GA, Pritchard TS, McMiller TL, Wei P, Pardoll DM, et al. Mechanisms regulating PD-L1 expression on tumor and immune cells. *J Immunother Canc* (2019) 7:305. doi: 10.1186/s40425-019-0770-2

93. Zhang H, Dai Z, Wu W, Wang Z, Zhang N, Zhang L, et al. Regulatory mechanisms of immune checkpoints PD-L1 and CTLA-4 in cancer. *J Exp Clin Cancer Res* (2021) 40:184. doi: 10.1186/s13046-021-01987-7

94. Zheng B, Ren T, Huang Y, Sun K, Wang S, Bao X, et al. PD-1 axis expression in musculoskeletal tumors and antitumor effect of nivolumab in osteosarcoma model of humanized mouse. *J Hematol Oncol* (2018) 11:16. doi: 10.1186/s13045-018-0560-1

95. Lei T, Qian H, Lei P, Hu Y. Ferroptosis-related gene signature associates with immunity and predicts prognosis accurately in patients with osteosarcoma. *Cancer Sci* (2021) 112:4785–98. doi: 10.1111/cas.15131

96. Zhao J, Zhao Y, Ma X, Feng H, Jia L. Outstanding prognostic value of novel ferroptosis-related genes in chemoresistance osteosarcoma patients. *Sci Rep* (2022) 12:5029. doi: 10.1038/s41598-022-09080-5

97. Jiang M, Wang Z, He X, Hu Y, Xie M, Jike Y, et al. A risk-scoring model based on evaluation of ferroptosis-related genes in osteosarcoma. *J Oncol* (2022) 2022:4221756. doi: 10.1155/2022/4221756

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fimmu.2022.1013094

OPEN ACCESS

EDITED BY
Hao Zhang,
Xiangya Hospital Central South University, China

REVIEWED BY
Jianglong Yan,
The University of Chicago, United States
Shengxiu Liu,
Anhui Medical University, China

*CORRESPONDENCE
Hongjuan Cui
hcui@swu.edu.cn;
hongjuan.cui@gmail.com
Ping Liang
liangping868@sina.com

†These authors have contributed equally to this work

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

RECEIVED 06 August 2022
ACCEPTED 19 October 2022
PUBLISHED 17 November 2022

CITATION
Wan S, Moure UAE, Liu R, Liu C, Wang K, Deng L, Liang P and Cui H (2022) Combined bulk RNA-seq and single-cell RNA-seq identifies a necroptosis-related prognostic signature associated with inhibitory immune microenvironment in glioma. *Front. Immunol.* 13:1013094. doi: 10.3389/fimmu.2022.1013094

Combined bulk RNA-seq and single-cell RNA-seq identifies a necroptosis-related prognostic signature associated with inhibitory immune microenvironment in glioma

Sicheng Wan[1,2†], Ulrich Aymard Ekomi Moure[2,3†], Ruochen Liu[1,2], Chaolong Liu[1,2], Kun Wang[1,2], Longfei Deng[1,2], Ping Liang[4*] and Hongjuan Cui[1,2*]

[1]The State Key Laboratory of Silkworm Genome Biology, College of Biotechnology, Southwest University, Chongqing, China, [2]Cancer Center, Medical Research Institute, Southwest University, Chongqing, China, [3]The Ninth People's Hospital of Chongqing, Affiliated Hospital of Southwest University, Chongqing, China, [4]Department of Neurosurgery, Chongqing Children's Hospital, Chongqing, China

Necroptosis is a programmed cell death playing a significant role in cancer. Although necroptosis has been related to tumor immune environment (TIME) remodeling and cancer prognosis, however, the role of necroptosis-related genes (NRGs) in glioma is still elusive. In this study, a total of 159 NRGs were obtained, and parameters such as mutation rate, copy number variation (CNV), and relative expression level were assessed. Then, we constructed an 18-NRGs-based necroptosis-related signature (NRS) in the TCGA dataset, which could predict the patient's prognosis and was validated in two external CGGA datasets. We also explored the correlation between NRS and glioma TIME, chemotherapy sensitivity, and certain immunotherapy-related factors. The two necroptosis-related subtypes were discovered and could also distinguish the patients' prognosis. Through the glioblastoma (GBM) scRNA-seq data analysis, NRGs' expression levels in different GBM patient tissue cell subsets were investigated and the relative necroptosis status of different cell subsets was assessed, with the microglia score culminating among all. Moreover, we found a high infiltration level of immunosuppressive cells in glioma TIME, which was associated with poor prognosis in the high-NRS glioma patient group. Finally, the necroptosis suppressor CASP8 exhibited a high expression in glioma and was associated with poor prognosis. Subsequent

experiments were performed in human glioma cell lines and patients' tissue specimens to verify the bioinformatic analytic findings about CASP8. Altogether, this study provides comprehensive evidence revealing a prognostic value of NRGs in glioma, which is associated with TIME regulation.

KEYWORDS

necroptosis-related gene, necroptosis-related prognostic signature, glioma, tumor immune microenvironment, single cell RNA seq

Introduction

Glioma, a malignant central nervous system (CNS) tumor originating from the glial, is featured with high recurrence and poor prognosis. According to the World Health Organization (WHO), glioma classification relies on different histopathological subtypes and is classified into four grades (I-IV), with grades I-III being the low-grade glioma (LGG) and grade IV representing the aggressive form, GBM (1, 2). Although conventional surgical resection, chemoradiotherapy combined with immunotherapy and electric field therapy improve the prognosis of glioma patients to an extent, the overall prognosis of glioma patients remains poor due to the glioma heterogeneity and epigenetic mutations of intratumoral molecules (including isocitric dehydrogenase (IDH) and epidermal growth factor receptor mutations, 1p19q co-deletion, and MGMT promoter methylation). Accordingly, it is urgent to comprehensively understand the molecular mechanisms underlying glioma recurrence and progression, and discover new biomarkers for a better diagnosis and treatment of this disease (3).

Necroptosis is a novel form of regulated necrosis. It is a cell death pattern originally programmed to protect the host against microbial agents when the caspase-dependent apoptosis pathway is blocked by pathogens. Generally, the classical necroptotic pathway is triggered by extracellular stimuli, such as tumor necrosis factor (TNF), which activates the downstream receptor-interacting serine/threonine kinases 1 and 2 (RIPK1/2), leading to the phosphorylation of the mixed lineage kinase domain-like pseudokinase (MLKL). The latter then translocates to the cytoplasmic membrane to generate the pore complex, resulting in the release of damage-associated molecular patterns (DAMPs) and cellular contents, and membrane rupture (4, 5). Caspase-8 (CASP8) is a necroptosis suppressor that inhibits the necroptotic pathway by cleaving RIPK1 and RIPK3 (6). The two latter molecules show low expression level in multiple cancer types and the conventional perspective reckon that the related activated necroptosis promotes cell necrosis and leads to tumor inhibition. However, recent studies have suggested that necroptosis might play a dual role in tumors. For instance, necroptosis-mediated dying tumor cells induce the C-X-C motif chemokine ligand 1 (CXCL1) and sin3A-associated protein 130 (SAP130) release to aggravate inhibitory TIME (7, 8). Besides, DAMPs released from damaged or dying cells can promote immunosuppressive cell accumulation in the TIME (9).

In this study, we aimed to comprehensively analyze the expression patterns of NRGs in glioma, construct an NRS to predict the prognosis of glioma patients, and explore the relationship between necroptosis and glioma immune microenvironment at a single-cell level. We found the dysregulation of the necroptosis pathway in glioma and certain NRGs displayed abnormal expression, multitype mutations, and CNVs, and certain of the NRGs were associated with TIME regulation. Next, we constructed and validated an NRS that could effectively predict the prognosis and chemosensitivity of patients with glioma. And then, based on the NRS, we explored the immunotherapy difference between the two NRS groups and further discovered two necroptosis-related subtypes, which can also distinguish the patients' prognosis. Later, we also assessed the expression of NRGs in different cell subsets at a single-cell level. Finally, we analyzed the core intersection gene *CASP8* from the perspectives of immune checkpoints, immune cell infiltration, prognosis, and protein expression, which unveiled that CASP8 can be used as a novel potential glioma prognosis biomarker.

Materials and methods

Datasets acquisition and processing

The expression profile with the Toil RNA-seq recompute and related clinical sample information LGG, GBM, and normal brain tissue samples) were downloaded from The Cancer Genome Altas (TCGA) [TCGA-LGG, TCGA-GBM and the Genotype-Tissue Expression (GTEx) database (UCSC Xena repository, an online cancer database designed by the University of California–Santa Cruz, http://xena.ucsc.edu/)] (10). Thereafter, LGG and GBM

RNA-seq data were merged, and totals of 642 glioma samples (LGG, n=499 and GBM, n=143) and 1259 normal brain tissue samples were included in the present study.

For dataset validation, gene expression profile data (mRNAseq-693 and mRNAseq-325) and corresponding clinicopathological information for glioma patients were retrieved from the Chinese Glioma Genome Atlas (CGGA, http://www.cgga.org.cn/) (11).

Corresponding GBM scRNA-seq data from a previous scRNA-seq research (single-cell transcriptome profiles in 10 primary IDH^{wt} type GBM patients, GSE173278) were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). The R-package (Seurat 4.1.0) was used to analyze scRNA-seq data (12–15). The top 20 principal components were used to construct the SNN graph and UAMP embedding. The R-package (harmony 0.1.0) was used for batch correction and cell annotation was performed based on singleR and manual cell type annotation.

Differentially expressed NRGs identification

NRGs differential expression analysis was performed with FPKM and Wilcoxon rank sum test by using R-package (Limma) (16). Gene expression was considered significant when meeting the following criteria: adjusted p-value<0.05 (BH method) and $|\log_2(\text{Fold Change})| > 1.0$.

Analysis of NRGs mutation

NRGs mutation frequency and oncoplot waterfall plot were generated by R-package (maftools) (17). For the gene CNVs, the value >0.2 was defined as "gain" and the value< -0.2 was defined as "loss".

Identification and validation of the NRS

A total of 159 NRGs were obtained from the KEGG necroptosis pathway [https://www.kegg.jp/entry/map04217) (18). By using the R-package (survival (3.2-13)], we performed the univariate Cox regression analysis to identify the NRGs related to glioma patients' overall survival (OS) in the TCGA training set (p<0.05). A total of 126 genes were screened as potential risk factors related to the OS. Then, the LASSO regression algorithm was performed by using the R-package (glmnet) to calculate regression coefficients to further refine the gene set (19, 20). Finally, 15 NRGs were identified as the most valuable OS genes and based on normalized gene expression values and coefficients, each sample risk score was calculated using the following formula:

$$Riskscore = \sum_{i=1}^{n}(\text{exprgenei x coefficientgenei})$$

Based on the median value of all patients' risk scores in the TCGA training set, samples were classified into high- and low-risk cohorts. For the signature validation, the same calculation as described above was employed in two CGGA (mRNAseq-693, mRNAseq-325) validation sets, respectively.

Then, we used R-packages [survival (3.2-13)" and "survminer (0.4.9)] to analyze the survival of two risk groups through the Kaplan-Meier (K-M) curve. The Log-Rank test was conducted to assess survival differences between the two groups. The time-dependent receiver operating characteristic (ROC) curve was plotted by using R-packages [timeROC (0.4)" and "survival (3.2-13)] to evaluate the predictive ability of the NRS for 6 months, 1-, 2-, and 3-year glioma patient survival rates.

Functional enrichment analysis

To clarify the risk score-related to biological functions and pathways, the differentially expressed genes (DEGs) between high and low-risk groups in the training set were identified as described above. DEGs' biological functions and pathways were explored through the gene ontology (GO) and Kyoto Genome Encyclopedia (KEGG) pathway enrichment analysis by using R-package (ClusterProfiler), the FDR p-value<0.05 was used as the cut-off criterion.

The gene set enrichment analysis (GSEA) was used to identify and compare the different cancer hallmarks between high and low-risk groups in the TCGA cohorts (https://www.gsea-msigdb.org/gsea/index.jsp).

Immune cell infiltration analysis

To explore the relationship between the immune cell infiltration and calculated risk score, the correlation between NRS groups and different immune cell infiltration was analyzed by using the CIBERSORT, GSVA, and XCELL methods (21–23). The gene set of immune cell types was obtained from a previous research (24).

Chemotherapeutic drug response analysis

The R-package (oncoPredict) was used to assess the drug response differences between risk score and corresponding drugs derived from the Genomics of Drugs Sensitivity in Cancer (GDSC), Cancer Therapeutics Response Portal (CTRP), and Cancer Cell Line of Encyclopedia (CCLE). Pearson coefficient was used to calculate the correlation between signature score and area under the dose-response curve (AUCs) values.

Construction of an NRS-based nomogram

The R-package (rms) was used to build the NRS-based nomogram to predict glioma patients' 6-months, 1- and 2-year survival probability. To validate the nomogram, the calibration was plotted, which can assess the nomogram prognostic accuracy; the 45° line represents the best prediction. The decision curve analysis (DCA) curve, drawn by R-package (rmad), was used to test the nomogram value for clinical application.

The exploration of new necroptosis subtypes

Based on the 18 NRGs, the consensus non-negative matrix factorization (CNMF) algorithm was performed to identify new necroptosis subtypes in the TCGA glioma cohort by using the R-package (CancerSubtype). We used the silhouette coefficient to evaluate the most optimal cluster number.

Cell culture

Human glioma (LN-229, U87-MG, U118-MG, U251-MG, and A172) and human astrocyte NHA cell lines were obtained from American Type Culture Collection (ATCC, Beijing, China). All cell lines were regularly tested for mycoplasma contamination, and cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco, New York, NY, USA), supplemented with 10% fetal bovine serum (Gibco, New York, NY, USA).

Antibodies

The CASP8 (13423-1-AP) and CD11B/ITGAM (#66519-1-Ig) antibodies were purchased from Proteintech Group (Wuhan, China), and α-Tubulin (ab7291) antibody was purchased from Abcam (Cambridge, UK).

Western blot (WB) assay

Cells were lysed with the cell lysis buffer (Beyotime). WB assay was performed as previously described (25).

Patient specimen immunohistochemistry (IHC) assay

15 pairs of glioma patients' tissue samples were obtained from the Affiliated Hospital of Southwest University (The Ninth People's Hospital of Chongqing), and patients agreed and signed consent. As previously described, paraffin sections were dewaxed, hydrated, and repaired with corresponding antigens (Sangon, Shanghai China), followed by overnight incubation with CASP8 antibody (26). Visualization was made by using a horseradish peroxidase detection system.

Immunofluorescence histochemistry (IFH) assay

After dewaxing, samples were incubated with 10 mM citric acid (pH6.0) for antigen repair and blocked with goat serum for 2 hours. Following primary antibody staining, PBS was washed and incubated with fluorescent secondary antibody for 1 hour, then washed with phosphate-buffered saline (PBS). Then, we repeated the above procedure for the second primary antibody staining. Finally, the tissue slices were mounted with medium containing 4',6-diamidino-2-phenylindole (DAPI). Stained slides were observed by using confocal fluorescence microscopy (Leica, Germany).

Data statistics and analysis

The bioinformatics statistics analysis was accomplished by R (version 4.1.2, Institute for Statistics and Mathematics, Vienna, Austria; https://www.r-project.org). The correlation analysis was conducted by Spearman correlation analysis. The Chi-square test was used to compare the different clinical indicators. Survival status was evaluated by Cox regression analysis and the OS, DSS, and PFS were generated by the Kaplan–Meier method and evaluated by the log-rank test, respectively. The log-rank test was used to assess the difference in immune infiltration and drug response among different NRS groups.

As for the part of experiment validation, all observations were confirmed by at least three independent biological replicates. The results in this study were presented as the means ± standard deviation (SD). Two-tailed Student's t test was performed for paired samples. $P<0.05$ was considered statistically significant.

Results

The necroptosis pathway is dysregulated in glioma

The workflow of this study is depicted in Figure 1. Patients with incomplete clinical information were excluded from the bulk RNA-seq data, and the clinical features of all included patients are summarized in Table 1.

We first investigated the mutation type and occurrence of the 159 NRGs. The results showed a high occurrence of missense and nonsense mutations, splice sites, deletions, and insertions among the top 20 NRGs (Figure 2A). Then, we detected the NRGs' CNV score using the GISTIC software, and found that the *interferon alpha* (*IFNA*) family genes lost their copy numbers (Figure 2C, and Figures S1A-B). Next, we combined data from both TCGA and GTEx databases to identify the DEGs between glioma and normal brain tissues using the following parameters: adjusted P value< 0.05, and $|\log_2(\text{Fold Change})| > 1$. As a result, we screened 29 NRGs with significant differential expressions (Figure 2B, and Figure S1C). Finally, based on the 159 NRGs and TCGA data, we performed the ssGSEA to calculate the sample enrichment score and combined it with survival analysis, where we found that activation of the necroptosis pathway was associated with poor prognosis (Log-rank, p<0.001) (Figure 2D, and Figures S1D-1E).

Establishment and validation of an NRS in glioma

The above findings reveal the dysregulation of the necroptosis pathway in glioma. To construct an effective necroptosis prognosis signature in glioma, we pre-screened 126 potential NRGs closely associated with prognosis through the univariate Cox regression (Table S1). Among these 126 NRGs, we applied the LASSO regression analysis and discovered 18 candidate genes, including the *TIR domain-containing adaptor molecule 2* (*TICAM2*), *interferon beta 1* (*IFNB1*), *H2A.X variant histone* (*H2AX*), *peptidylprolyl isomerase A* (*PPIA*), *interferon-gamma receptor 2* (*IFNGR2*), *H2A clustered histone 11* (*H2AC11*), *interleukin 1 alpha* (*IL1A*), *caspase 8* (*CASP8*), *Z-DNA binding protein 1* (*ZBP1*), *baculoviral IAP repeat containing 3* (*BIRC3*), *phospholipase A2 group IVA* (*PLA2G4A*), *TNF receptor superfamily member 1A* (*TNFRSF1A*), *TNF receptor superfamily member 10 B* (*TNFRSF10B*), *signal transducer and activator of transcription 3* (*STAT3*), *H2A.W histone* (*H2AW*), *BH3 interacting domain death agonist* (*BID*), *macroH2A.2 histone* (*MACROH2A2*), and *glutamate dehydrogenase 1* (*GLUD1*) (Figures 3A–C).

Then, Glioma patients in the TCGA training set were classified into low- (n=321) and high- (n=320) risk groups according to the median risk score. The K-M curves showed that compared to the low-risk group, the high-risk group significantly held a poor prognosis (Log-rank test, p<0.0001) (Figure 3D). Besides, time-dependent ROC curves showed high sensitivity and specificity for 6 months, 1-, 2-, and 3-year survivals (Figure 3E). Finally, we performed the univariate Cox regression analysis on the WHO glioma grades (I-IV), gender, age, and NRS groups, and found that age and risk score were significantly associated with glioma patients' survival, suggesting that the NRS could serve as a prognosis factor for glioma patients (Figure 3F).

Finally, to validate the NRS predictive performance, we assessed the two CGGA datasets (CGGA-mRNAseq_325 and

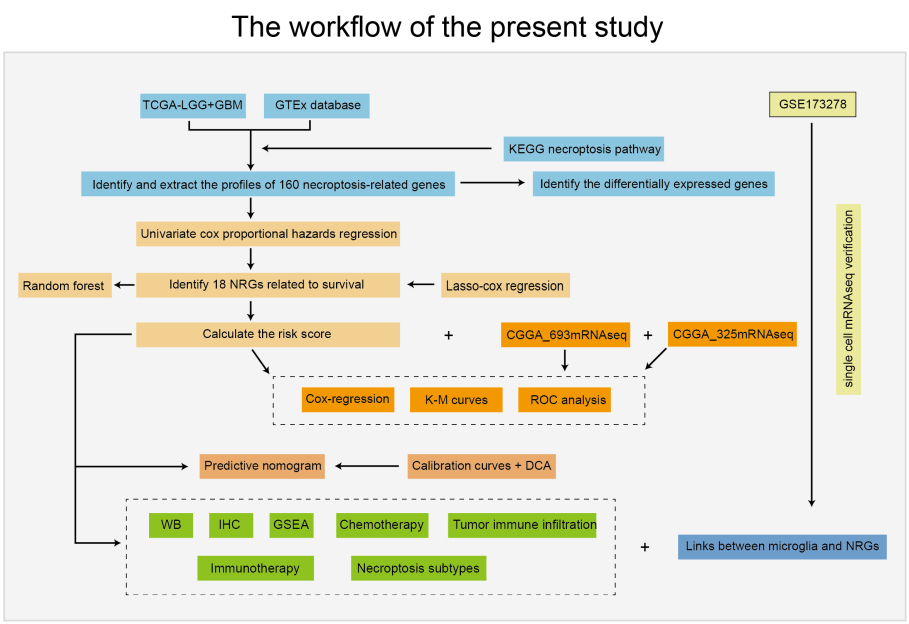


FIGURE 1
Workflow used in this study.

10.3389/fimmu.2022.1013094

TABLE 1 Patients' clinical features from the bulk RNA-seq.

	TCGA (N = 642)	CGGA_693 (N = 692)	CGGA_325 (N = 321)	Overall (N = 1655)
Cancer type				
GBM	143 (22.3%)	249 (36.0%)	139 (43.3%)	531 (32.1%)
LGG	499 (77.7%)	443 (64.0%)	182 (56.7%)	1124 (67.9%)
Age				
Mean	46.7 (15.1)	43.3 (12.4)	43.0 (12.0)	44.5 (13.5)
Median	45.5 [14.0,89.0]	43.0 [11.0,76.0]	42.0 [8.0,79.0]	43.0 [8.0,89.0]
Missing	0 (0%)	1 (0.1%)	0 (0%)	1 (0.1%)
Gender				
Female	271 (42.2%)	294 (42.5%)	122 (38.0%)	687 (41.5%)
Male	371 (57.8%)	398 (57.5%)	199 (62.0%)	968 (58.5%)
Grade				
II	241 (37.5%)	188 (27.2%)	103 (32.1%)	532 (32.1%)
III	258 (40.2%)	255 (36.8%)	79.0 (24.6%)	592 (35.8%)
IV	143 (22.3%)	249 (36.0%)	139 (43.3%)	531 (32.1%)
Radio status				
No	115 (17.9%)	136 (19.7%)	65 (20.2%)	316 (19.1%)
Yes	139 (21.7%)	510 (73.7%)	242 (75.4%)	891 (53.8%)
Missing	388 (60.4%)	46.0 (6.6%)	14.0 (4.4%)	448 (27.1%)

CGGA-mRNAseq_693), which we sorted into both high-risk (mRNAseq_325: n=156, and mRNAseq_693: n=328) and low-risk (mRNAseq_325: n=157, and mRNAseq_693: n=329) groups by using the respective dataset's median risk score as the cut-off value based on the same calculation formula in the TCGA training set. The K-M curves showed that the high-risk group had a shorter survival period than the low-risk group and the ROC curves proved the predictive effect of the NRS (Figures 3G–N).

Identification of potential signaling pathways and biological processes related to the NRS

In order to probe the biological functions related to the NRS, we displayed the differential analysis of the GSVA score on cancer hallmark pathways between different NRS groups, based on the following criteria: FDR< 0.05. Apparently, multiple oncogenic pathways were significantly activated in the high-

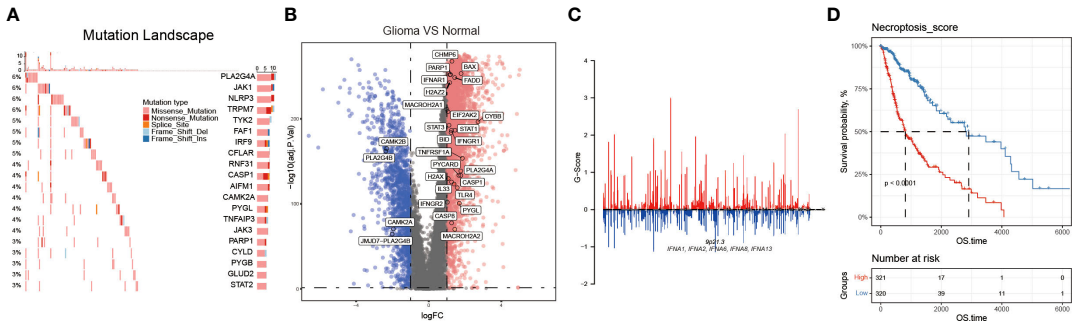


FIGURE 2
The necroptosis pathway is dysregulated in glioma. **(A)** NRGs mutational landscape. The more frequent mutations included missense and nonsense mutations, splice sites, deletions, and insertions. **(B)** Volcano plot showing some NRGs differentially expressed in glioma and normal brain tissues. **(C)** The CNV, gain, and loss of the NRGs and their distribution on human chromosomes. **(D)** Necroptosis score was calculated by using the GSVA method, and through the K-M curves, activated necroptosis was found associated with poor prognosis (OS).

10.3389/fimmu.2022.1013094

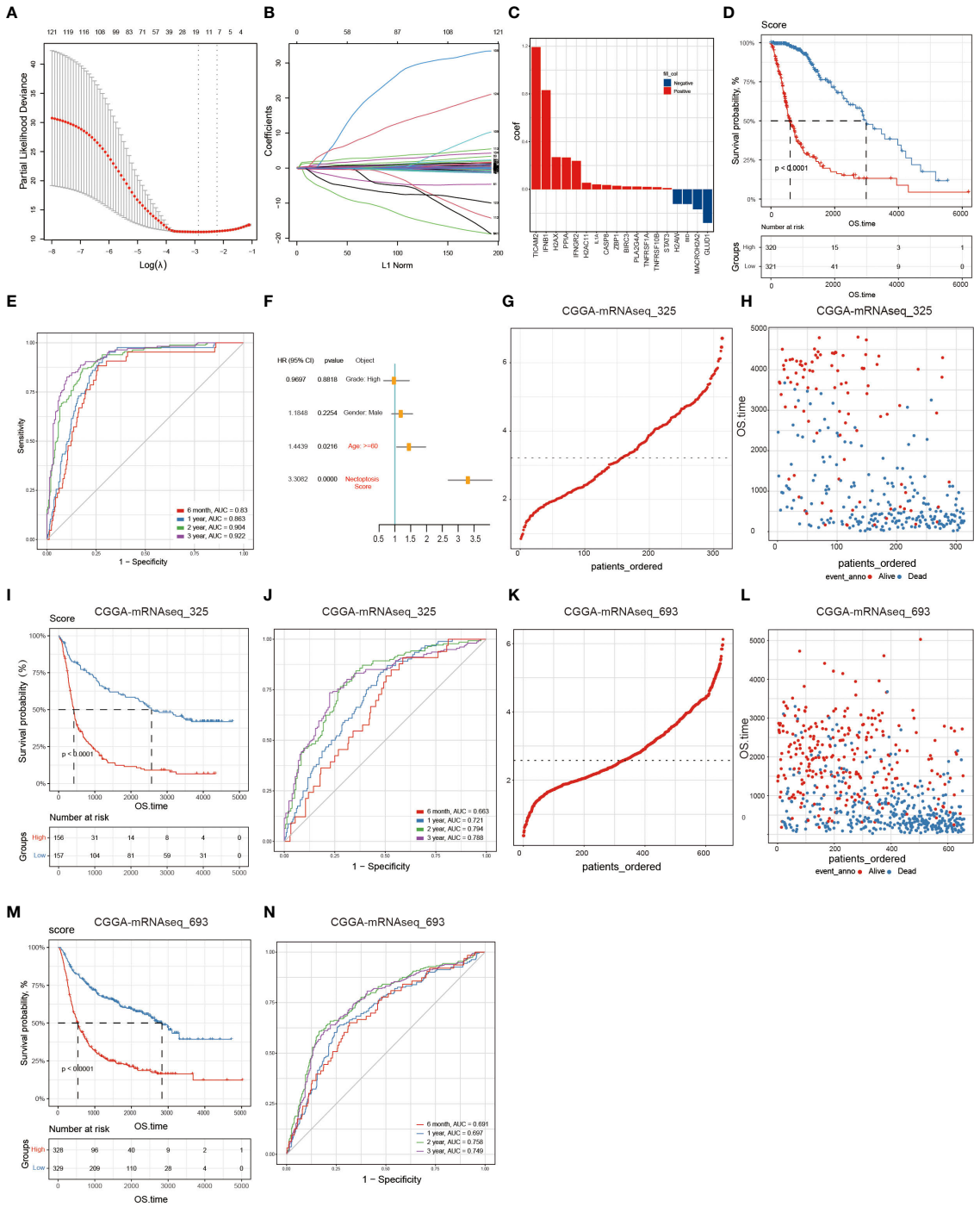


FIGURE 3
Building and validating the NRS. A total of 126 NRGs were selected *via* the univariate Cox regression and Lasso regression and significantly correlated with prognosis. **(A)** The most optimal parameter selected in Lasso regression by using the 10-fold cross-validation. Red dots indicate the likelihood of deviance values, gray lines represent the standard error (SE), and vertical dot lines correspond to optimal values by minimum criteria, and 1-SE, respectively. **(B)** The Lasso coefficient profile of 126 NRGs, with each curve representing a gene. **(C)** A total of 18 NRGs were incorporated for the NRS construction. **(D)** The survival analysis of the NRS in the TCGA training set. **(E)** Verification of the NRS predictive performance using the Time-ROC analysis. **(F)** Univariate Cox regression analysis illustrates that signature and age were the independent prognostic factor for glioma patients. **(G, H)** The distribution of risk scores of glioma patients in CGGA-mRNA_325 and CGGA-mRNA_693 datasets, respectively. **(I, J)** Numbers of alive and dead patients with different risk scores in CGGA-mRNA_325 and CGGA-mRNA_693 datasets, respectively. **(K, L)** Survival analysis. **(M, N)** Time-ROC analysis.

10.3389/fimmu.2022.1013094

risk NRS group, such as angiogenesis, hypoxia, KRAS, Notch, PI3K-AKT-mTOR, and WNT-b-catenin. These pathways are closely related to the poor prognosis of glioma patients (Figure 4A). In addition, the KEGG and GO enrichment analysis were conducted, based on the DEGs between high- and low-risk groups in the TCGA training set. The KEGG and GSEA enrichments indicated that the signature was related to pivotal biological processes, such as lysosomes, antigen processing and presentation, oxidative phosphorylation, DNA replication, and so on. (Figures 4B, C). Besides, the GO enrichment analysis further revealed that a large number of immune-associated biological processes were associated with the NRS, including immune responses mediated by leukocytes, B cells, and T cells (Figure 4D). Notably, many immune-inflammatory biological processes and pathways were significantly enriched in three different enrichment methods, suggesting that the difference may be valid in TIME between high- and low-risk NRS groups. Nevertheless, further studies are needed to compare the specific different immune cell infiltration and TIME between the two NRS groups.

Analysis of tumor immune cell infiltration and chemotherapeutic drug sensitivity between NRS groups

Since the occurrence and progression of cancer considerably rely on the TIME, we sought to investigate the relationship between NRS groups and immune cell infiltration. To reach this aim, three algorithms including XCELL, GSVA, and CIBERSORT were performed to calculate the immune cell infiltration between NRS groups. Regulatory T cells (Tregs) negatively regulate the immunoreaction, and previous evidence has demonstrated that the increased grade of glioma is

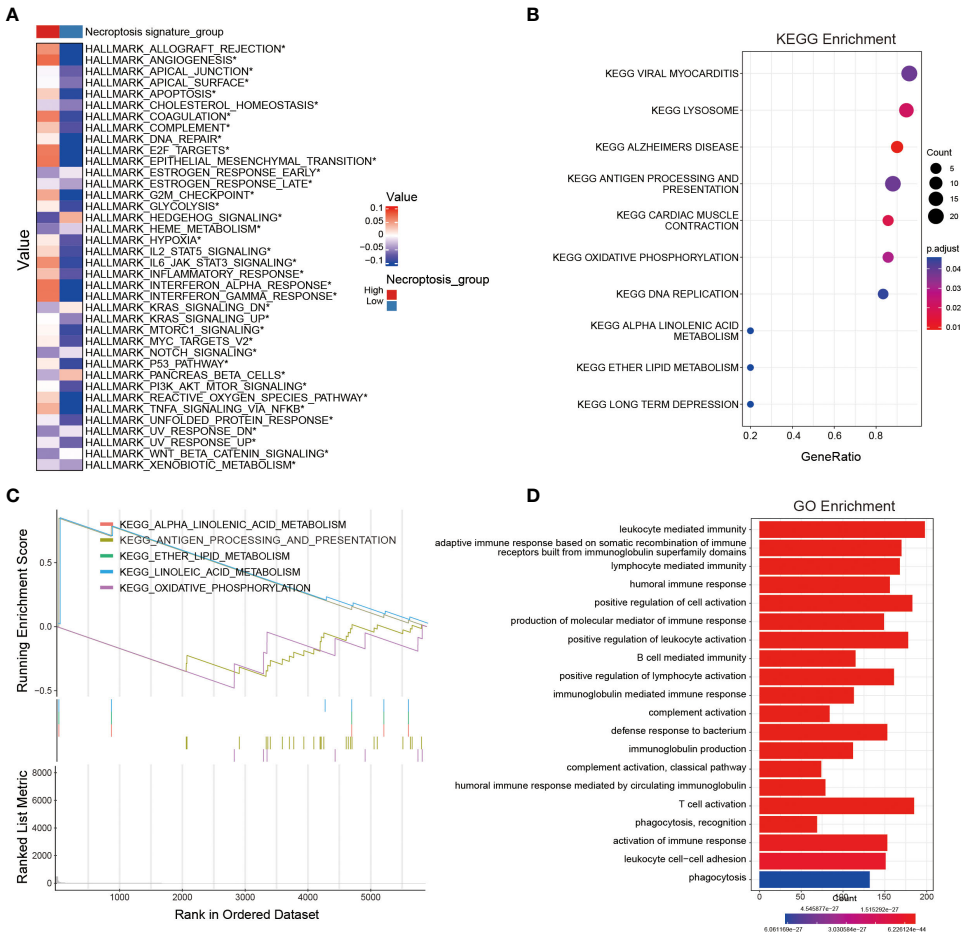


FIGURE 4
Identification of potential signaling pathways and biological processes associated with the NRS. **(A)** The hallmark pathway enrichment analysis between NRS groups. **(B, C)** KEGG and GSEA enrichment analyses revealing the relevant enrichment pathways of DEGs between NRS groups. **(D)** GO enrichment analysis unveiling the enrichment biological process of DEGs between NRS groups. *: p<0.05.

proportional to the number of $CD^{4+}CD^{25+}$Foxop3 Treg cells in the peripheral blood of glioma patients, indicating that malignant progression of glioma might be associated with Treg immunosuppression (27, 28). Macrophages are the major immune cells with high plasticity, and two activated forms of macrophages exist including M1 macrophages and M2 macrophages. Cytokines secreted by glioma can activate the STAT3 signaling in macrophages, down-regulate the surface antigens required for the antigen presentation, and up-regulate M2 macrophages-specific antigens like epidermal growth factor (EGF), vascular endothelial growth factor (VEGF), and matrix metallopeptidase (MMPs), which in turn promote the tumor growth and invasion (29–31). Moreover, myeloid-derived suppressor cells (MDSCs) are a group of phenotypic suppressor cells composed of myeloid progenitor cells and myeloid precursor cells in various differentiation stages, such as granulocytes, macrophages, and dendritic cells, known to extensively infiltrate in tumor tissues and exert an immunosuppressive role in the TIME (32, 33). It is also reported that MDSCs' proportion in peripheral blood mononuclear cells (PBMCs) in GBM patients was evidently higher than that in controls and the accumulation of MDSCs in GBM patients' peripheral blood may inhibit the immune effects of T cells (34, 35). Combining three immune cell infiltration algorithms, we found that the infiltration of activated M2 macrophages, Tregs, and MDSCs in the high-risk NRS group was significantly higher than that in the low-risk NRS group, which partly explained the effect of the immune cell infiltration on the survival between NRS group (Figures 5A–E).

To investigate the chemotherapeutic drug resistance between NRS groups based on the GDSC, CCLE, and CTRP datasets, the R package (oncoPredict) was used to calculate the sensitivity of NRS groups to different chemotherapeutic drugs (Figures 5F; S2A, 2C). We used the correlation analysis to identify the relationship between the signature score and the AUC of 12 GDSC-derived compounds. IOX2, axitinib, AZ628, dasatinib, CGP-60474, and sorafenib were negatively correlated with the signature score, whereas the remaining compounds (n=6) were positively correlated. Moreover, the 12 compounds were found to inhibit the ABL signaling, cell cycle, histone acetylation, DNA replication, ERK MAPK signaling, cell metabolism, RTK signaling, and other signaling pathways by targeting corresponding core molecules. Specific interactions between drugs, target molecules, and pathways are shown in sankey plot (Figures 5G and S2B, D).

Construction of clinical nomograms associated with the NRS

To further incorporate the NRS into the clinical diagnosis of glioma prognosis, we drew an intuitional nomogram to thoroughly comprehend the impact of certain risk factors on patient survival (age, gender, and grade) (Figure 6A). The calibration curve demonstrated a good agreement between the real observed OS and the ideal nomogram-predicted OS in 6 months, 1- and 2-year survival (Figure 6B). Finally, the DCA curves were performed to verify the acceptability of this signature in predicting the probability of glioma patients to survive for 6 months, 1 and 2 years. Our NRS and nomogram had a good predictive performance (Figure 6C). All results validated the high precision and practical utility of the NRS.

Predicting the immunotherapy sensitivity of glioma patients with the present NRS

An increasingly number of clinical trials and researches have reported that glioma patients can benefit from related immune checkpoint inhibitors (ICI), CAR-T therapy, and tumor antigen-related vaccines (36–38). Accordingly, we further explored the relationship between NRS groups and the expression of immunotherapy-related targeted molecules. Firstly, we used the TIDE method to evaluate the potential clinical efficacy of immunotherapy in different NRS groups. Among them, the higher the TIDE prediction score, the greater the possibility of immune evasion, indicating that the patient is less likely to benefit from ICI treatment. However, we found there was no significant difference in the TIDE score between the two groups (Figure 7A) (39). Next, through the differential analysis, we found that the human leukocyte antigen (HLA), checkpoints, chemokines, and costimulatory molecules were highly expressed in the high NRS group, such as the HLA-B, CD44, CXCL14, and TNFRSF1A (Figure 7B). Finally, we found that IFNG, CD8, and CD274 were highly expressed in the high NRS group and the Merck18 score was higher too (Figures 7C–F). Meanwhile, the Exclusion score was lower in the high NRS group (Figure 7G). The above are representative immunological biomarkers. These results indicated that high NRS group patients may be more likely to benefit from anti-tumor immunotherapy.

Exploring new necroptosis subtypes from the 18 NRGs in the present NRS

Based on the 18 NRGs in the present NRS, we identified two new necroptosis-related subtypes in glioma patients by utilizing the consensus non-negative matrix factorization (CNMF) method (Cluster 1 and Cluster 2) (Figures 8A, B) (40). Then, the K-M survival analysis showed that C1 held a better prognosis than C2 (Log-rank test) (Figure 8C). Dimensionally reduction was employed by Principal Component Analysis (PCA), and we found that the 18 NRGs have evident differential expression patterns (Figure 8D). The heatmap showed that the 4 genes, *MACROH2A2*, *GLUD1*,

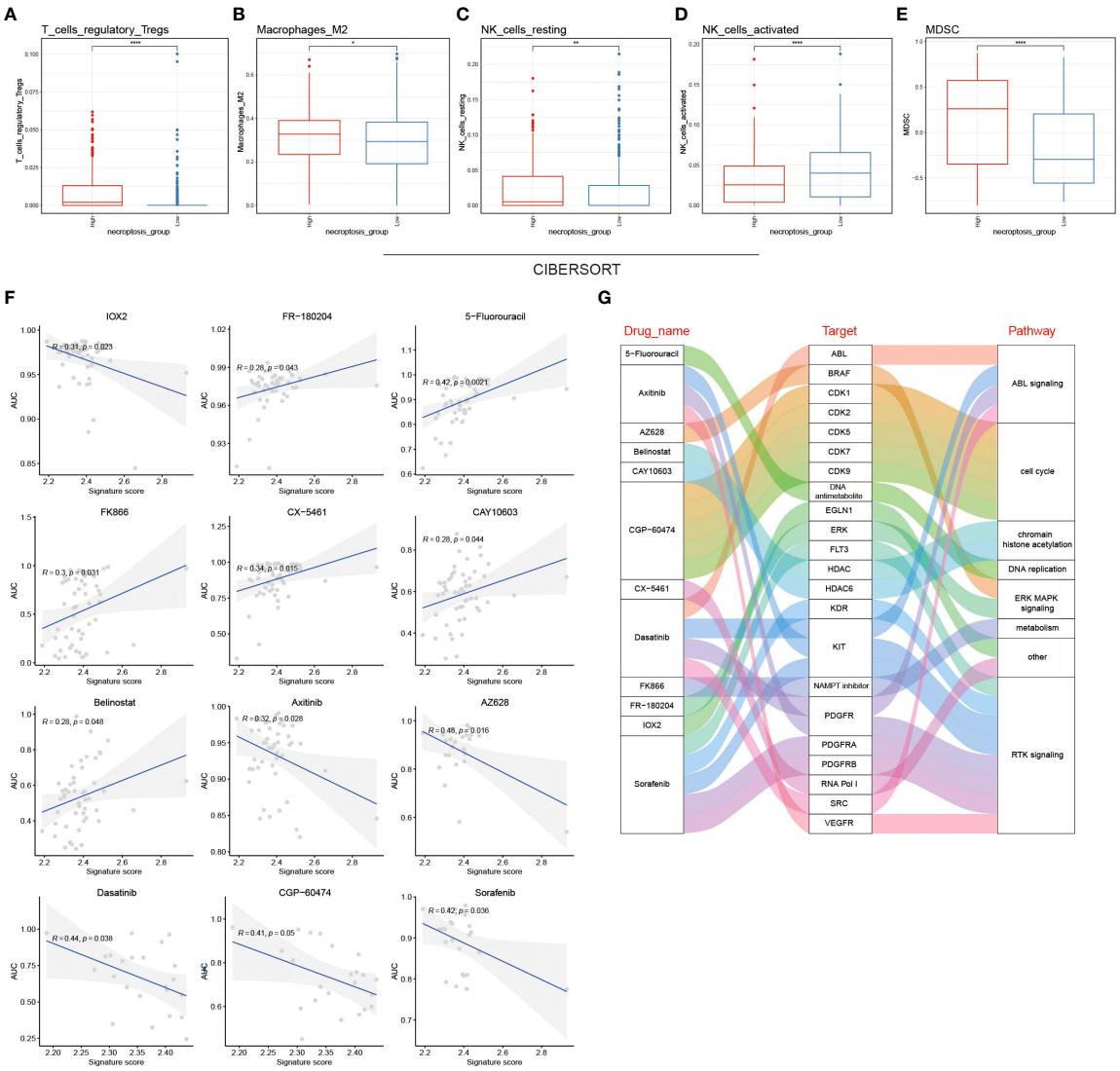


FIGURE 5
Tumor immune cell infiltration analysis and chemotherapeutic drug sensitivity between NRS groups. Infiltration level analysis of Tregs **(A)**, M2 macrophages **(B)**, resting NK cells **(C)**, activated NK cells **(D)**, and MDSCs **(E)** between NRS groups by using the CIBERSORT algorithm. **(F)** Pearson correlation coefficient between the AUC of 12 chemotherapy drugs and the signature score in the GDSC database. **(G)** Sankey plot showing the specific relationship between the 12 drugs in the GDSC database and their target molecules and pathways. * p<0.05; ** p<0.01 and **** p<0.0001.

H2AW, and *BID* were highly expressed in the C1 group, on the contrary, the other 14 genes were highly expressed in the C2 group (Figure 8E). In order to probe the related different pathways of the two subtypes, the GSVA was conducted and we found many immune and inflammatory pathways were significantly up-regulated in the C2 group. (Figure 8F). Finally, we used TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC, 7 different algorithms to evaluate the immune cell infiltration between the two subtypes. By comprehensively comparing the 7 methods, we found that the C2 group held a higher immune cell infiltration (Figure 8G).

Identification of potential genes associated with necroptosis in glioma by scRNA-seq

To resolve the expression patterns of genes in the NRS at a single-cell level, scRNA-seq data in the GSE173278 dataset was selected for analysis. The R-package (Seurat 4.1.0) was used for scRNA-seq analysis and the batch correction between samples was employed by using the harmony algorithm (Figures S3A-3C). We used the uniform manifold approximation and projection (UMAP) for dimensionality reduction visualization and a total of 29339 cells were divided into seven categories: $CENPF^{+}$ malignant (n=5363),

10.3389/fimmu.2022.1013094

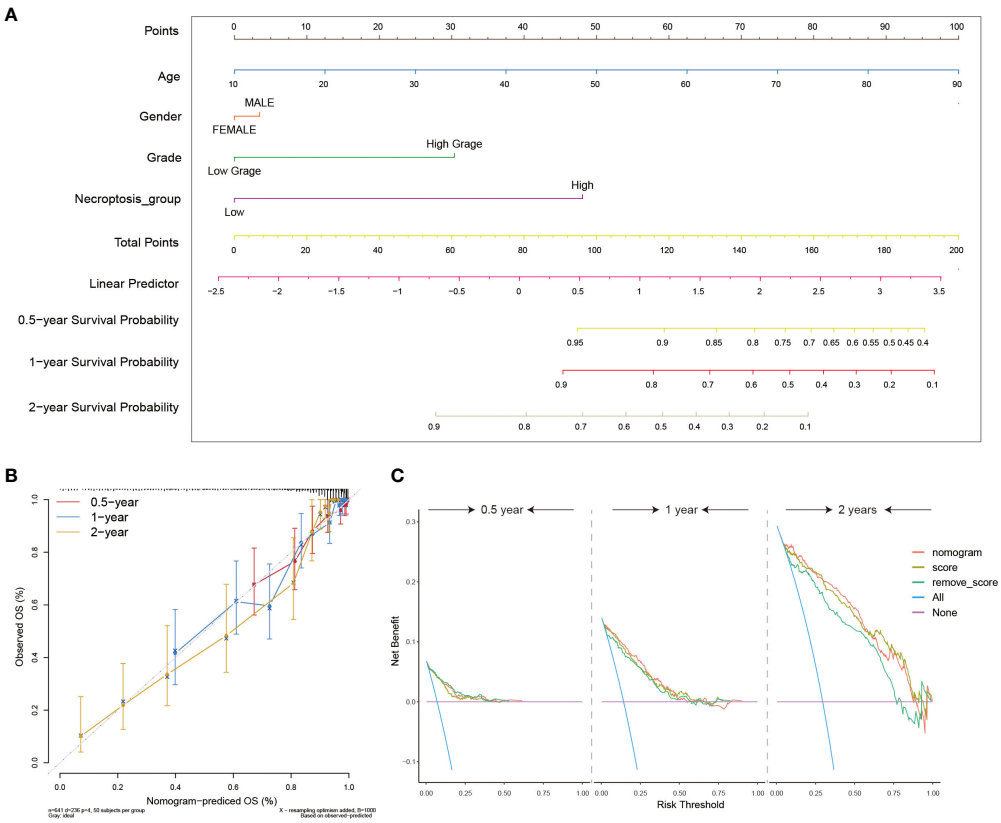


FIGURE 6
Construction of clinical utility nomograms with the NRS. **(A)** The nomogram included patients' age, gender, grade, and NRS groups for predicting the 6 months, 1-, and 2-year survival probability of glioma patients. **(B)** Calibration curves showing accuracy and specification of the nomogram and its association with actual clinical effects. **(C)** The DCA analysis of the nomogram for 6-months, 1- and 2-year survival prediction.

VEGFA$^+$ malignant (n=6446), OLIG1$^+$ malignant (n=11637), microglia (n=3219), endothelial cell (n=919), and oligodendrocyte (n=1020) (Figure 9A). Corresponding molecular markers were used to identify relevant cell subsets (Figure 9B). We analyzed the expression of 18 genes in the NRS in different types of cell subsets (Figure 9E), where BIRC3 and CASP8 were specifically expressed in VEGFA$^+$ malignant cells and microglia cells, respectively (Figures 9C, D). And then, CASP8 expression in microglia was verified by IFH assay in glioma patients' tissues, and we found a significant colocalization of CASP8 and microglia cell marker CD11B (Figure 9H). Because of the specific BIRC3 expression in the VEGFA$^+$ malignant cell subset, the biological function of the VEGFA$^+$ malignant cell subset was investigated. We analyzed the differences between this subset and the other 6 cell subsets and selected the DEGs for GO enrichment analysis. The results showed that the main enriched biological processes were hypoxia and stress responses (Figure 9F). Based on the genes in the NRS, we lastly used the GSVA to evaluate the necroptosis pathway status in each cell, and in general, we found the microglia and CENPF$^+$ malignant cells had a higher activated necroptosis status (Figure 9G). CASP8, a core molecule located in the necrosome, can selectively trigger apoptosis, necrosis, necroptosis, and inflammatory cell death, such as pyroptosis, depending on its status. CASP8 inhibition in the necroptosis pathway will promote the interaction of RIPK1 with RIPK3, which in turn phosphorylates the downstream molecule MLKL, ultimately leading to cell necroptosis (41). Microglia is a type of macrophage that infiltrates in glioma's TIME and has two subtypes (M1 and M2), M2 microglia can induce the immunosuppression, invasion, and angiogenesis of glioma by secreting cytokines (42, 43). Therefore, we hypothesized that the activated necroptosis process in microglia might evade the CASP8's inhibitory effect. These necroptosis microglia, which infiltrate in the glioma TIME induce a immunosuppression, which may be associated with the glioma progression and poor prognosis.

CASP8 is crucial for glioma progression

A random forest algorithm was employed to identify the most characteristic gene in the NRS gene set. Due to the profound impact on survival in glioma patients, 4 genes,

10.3389/fimmu.2022.1013094

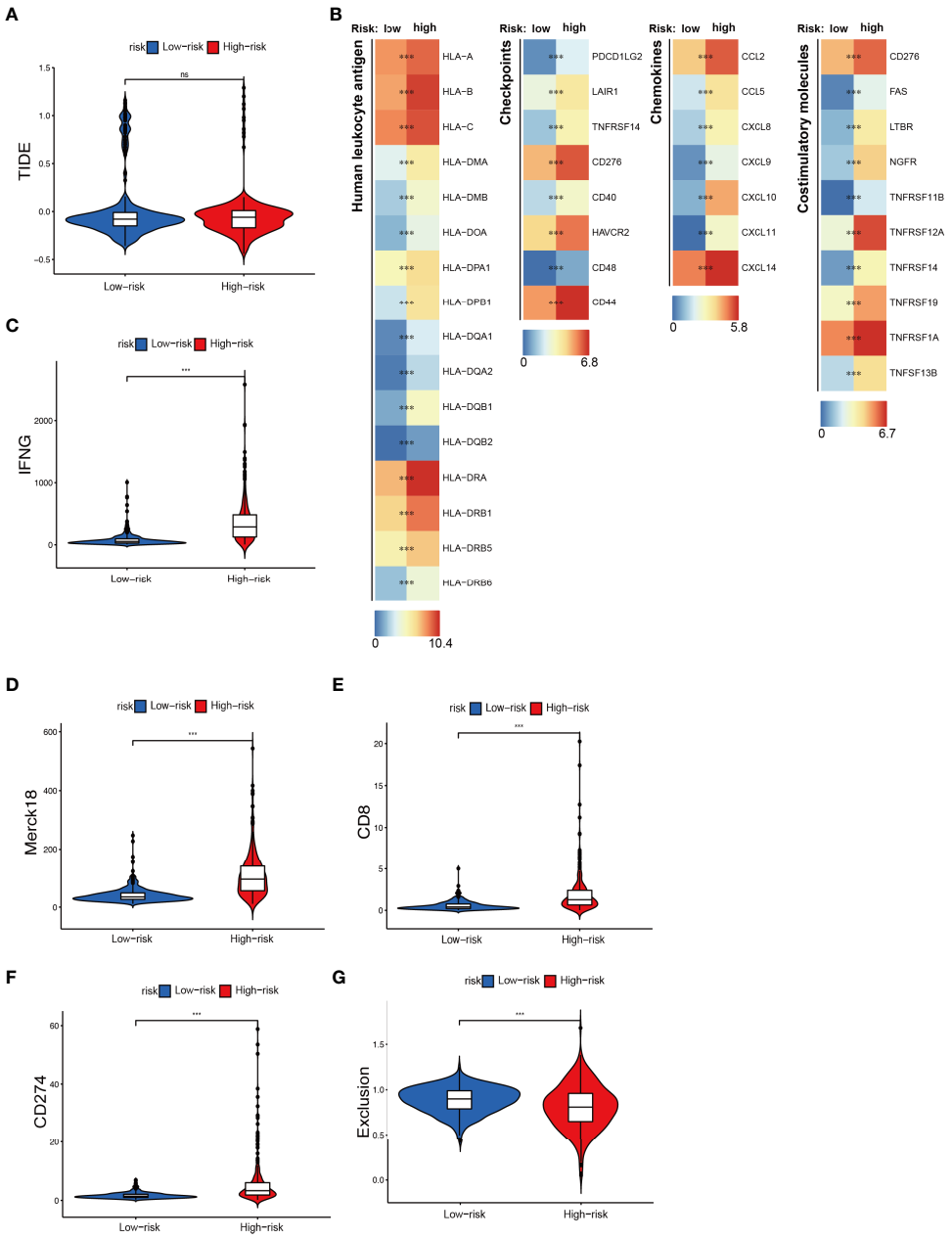


FIGURE 7
Immunotherapeutic responses exploration between NRS groups. **(A)** The TIDE score in NRS groups. **(B)** Expression of different immune-related indexes between NRS groups. **(C–G)** The expression differences of IFNG, CD8, CD274, and different Exclusion and Merck18 score between NRS groups. *** p<0.001; ns, no significance.

namely *IFNGR2*, *GLUD1*, *PPIA*, and *CASP8* were identified (Figure S4A). Both *IFNGR2* and *CASP8* represented the common intersection genes as revealed by the venn plot after combining the DEGs between glioma and normal brain tissues, NRS gene set, and important survival genes from the random forest (Figure 10A). Given the key role of CASP8 in regulating cell death process and its high expression in glioma, we further explored the biological function of *CASP8* in glioma at mRNA level, based on the TCGA database. We, that the expression levels of *CASP8* mRNA were considerably increased in different cancer types, including bladder urothelial carcinoma, esophageal carcinoma, stomach adenocarcinoma, etc. (Figure S4B). Then, from the K-M curves, an association was obtained between *CASP8* expression and the poor prognosis of glioma patients

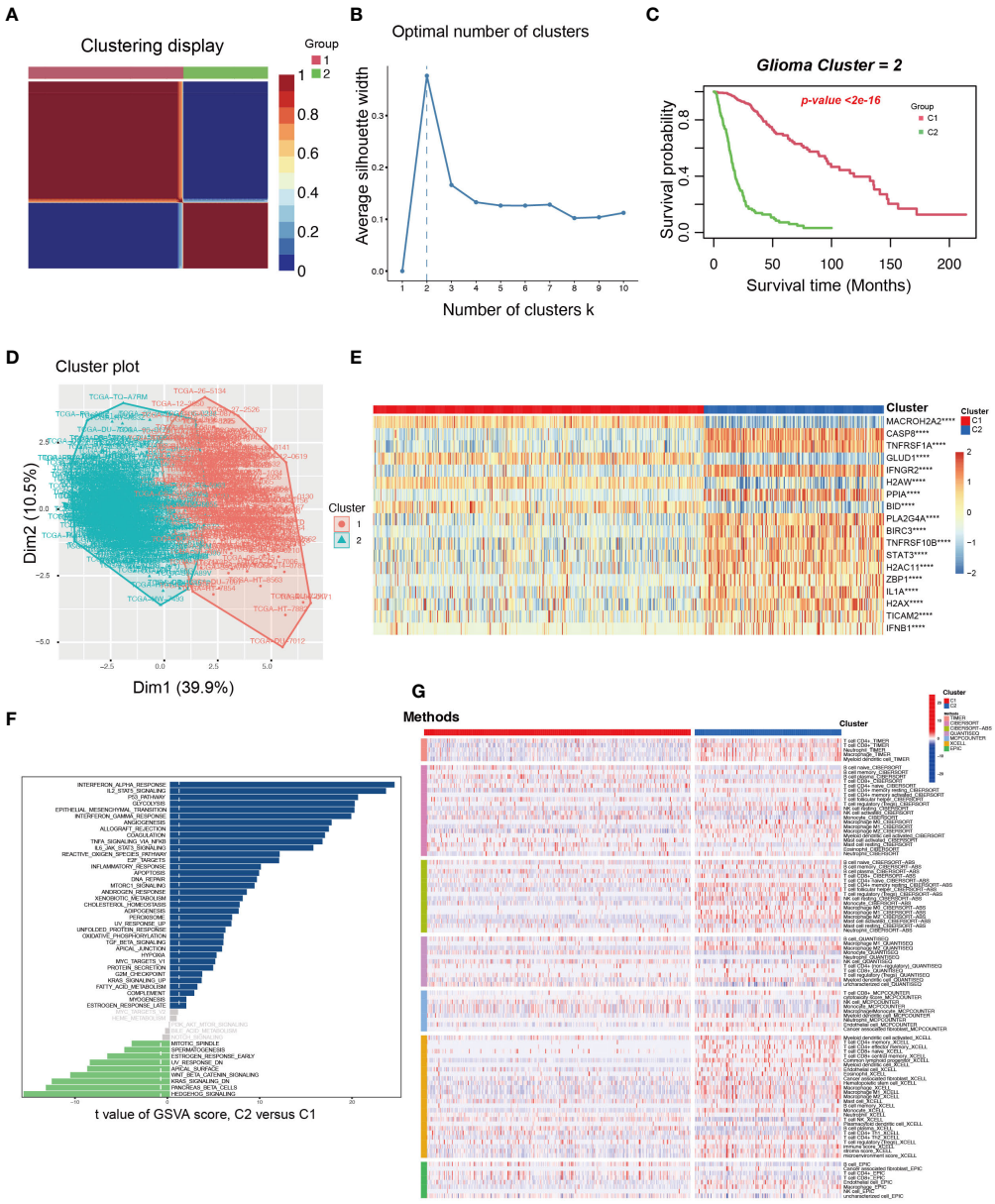


FIGURE 8
The exploration and assessment of two new subtypes in TCGA glioma patients from the 18 NRGs in the present NRS. **(A)** Based on the CNMF method, the C1 and C2 subtype groups were identified. **(B)** When K=2, the silhouette coefficient is the highest. **(C)** The K-M curves showed that the patients in the C1 group had a higher survival probability than the C2 group. **(D)** PCA analysis illustrated that the two subtypes held different NRGs expression patterns. **(E)** Heatmap showed the 18 NRGs expression in the two subtypes. **(F)** The pathway GSVA score of the two subtypes. **(G)** The heatmap showed 7 algorithms to assess the immune cell infiltration between the two subtypes. ****: p<0.0001.

(Figures 10F; S4C, S4D). Similarly, a positive correlation between *CASP8* and increased glioma histopathological grades was observed (Figure 10B). In addition, *CASP8* expression significantly differed between age groups (Age =<60 VS Age >60), while in gender groups there was no difference (Figures S4E, F). Methylation on the gene promoter region, one of the epigenetic modifications, controls gene transcription and expression to a large extent. Thus, based on the TCGA methylation data, we explored the relationship between mRNA and different CASP8 promoter methylation levels [4kb upstream and 100bp downstream of the transcription start site (TSS)] and found the methylation in both of the regions presented an evident negative correlation with *CASP8* expression (Figures S4G, H). We also probed the *CASP8* methylation in normal and

10.3389/fimmu.2022.1013094

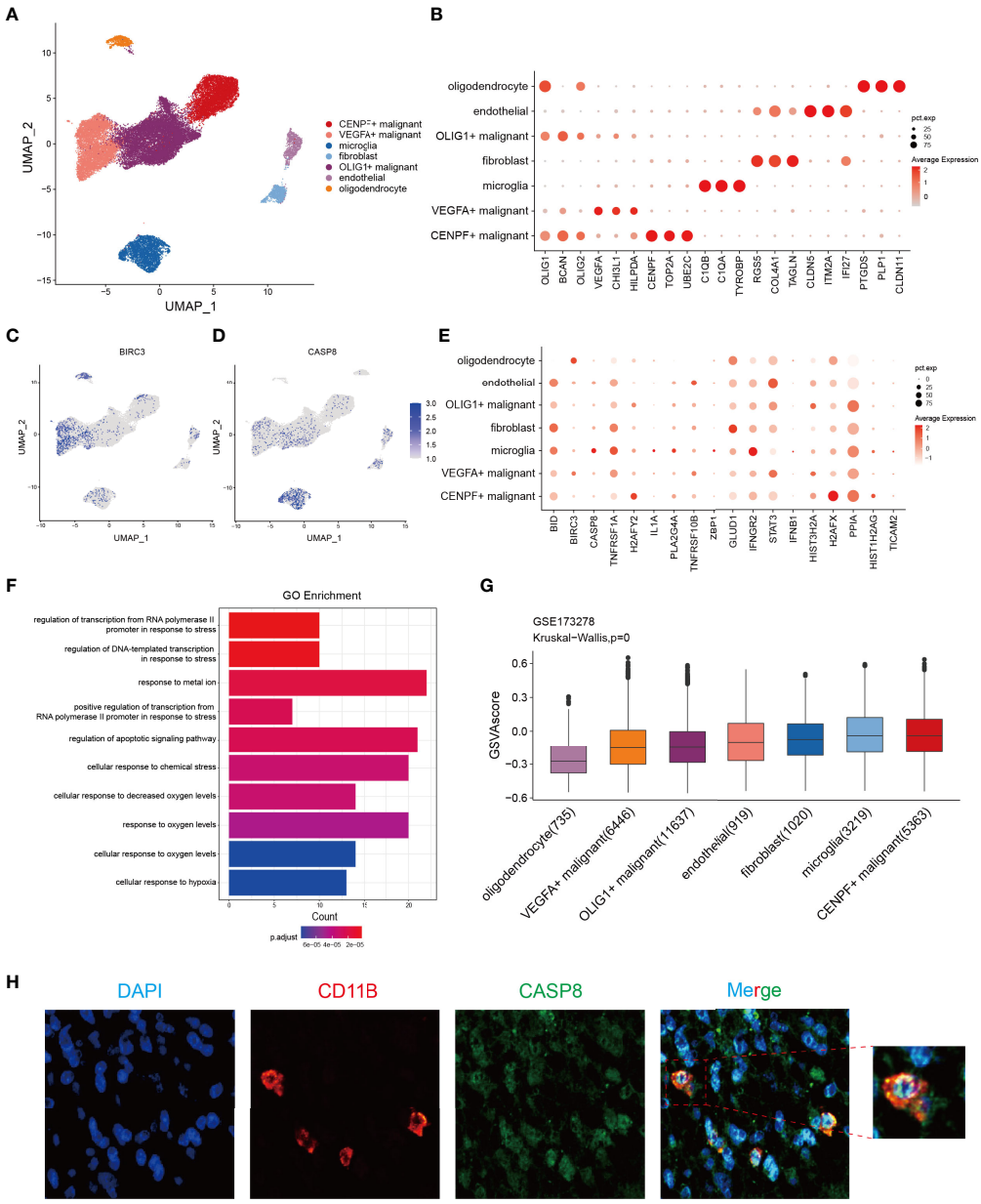


FIGURE 9
Identification of the potential genes associated with necroptosis in glioma through the scRNA-seq analysis. **(A)** Cells were divided into 7 cell subsets based on the marker gene expression. **(B)** Respective marker genes of the 7-cell subsets. **(C, D)** Expression of *BIRC3* and *CASP8* in different cell subsets, with a specific expression in VEGFA^{+} malignant cells and microglia, respectively. **(E)** Expression levels of 18 NRGs in the NRS from 7 different cell subsets. **(F)** GO enrichment analysis of significant DEGs in VEGFA^{+} malignant cell subsets versus other cell subsets. **(G)** The GSVA score of the necroptosis pathway in each cell subset. **(H)** IFH assay confirmed the localization of CASP8 in microglia (red fluorescent: anti-CD11B, microglia marker; green fluorescent: anti-CASP8; blue fluorescent: DAPI, nuclear).

GBM patients according to the TCGA-GBM cohort. In the primary tumor group, *CASP8* had a lower methylation level than that in the normal group, although the number of samples varied considerably between the two groups (Figure S4I). These results about *CASP8* methylation may also reflect the malignant function of *CASP8* in glioma development. Furthermore, a correlation between *CASP8* and chemokines, cytokine receptor interactions, and JAK-STAT signaling pathways was further unveiled using the KEGG and GSEA enrichment analyses (Figures 10D, E). Given the association between *CASP8* and immune signaling pathways, we investigated the correlation between *CASP8* and immune checkpoints (Figures 10C; S5A,

10.3389/fimmu.2022.1013094

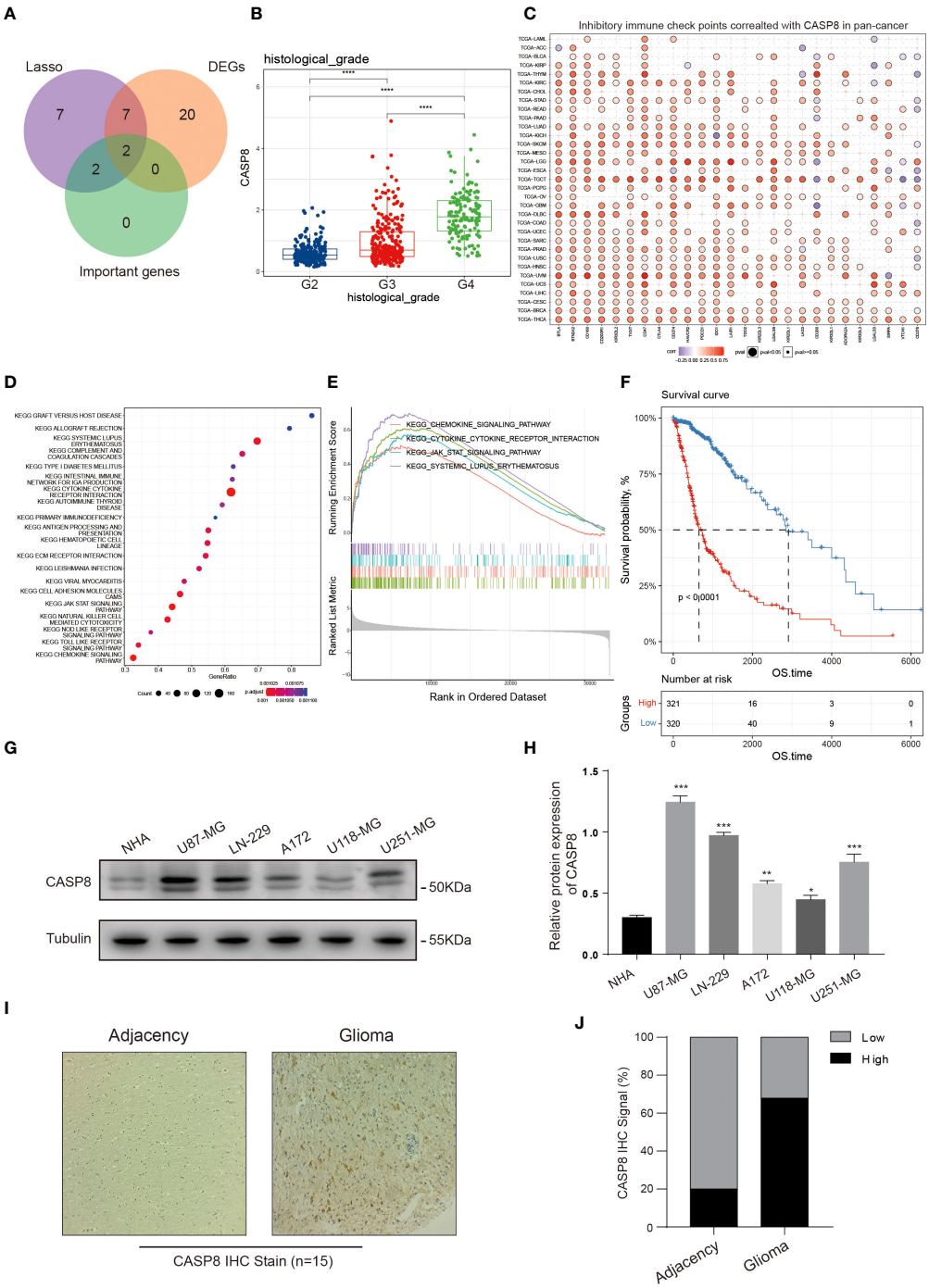


FIGURE 10
CASP8 functional exploration. **(A)** Venn plot portraying the common genes in the NRS, DEGs, and random forest analysis. **(B)** The expression of *CASP8* was positively correlated with glioma patient histopathological grades (II-IV). **(C)** Heat map displaying the relationship between various inhibitory immune checkpoints and *CASP8* expression in pan-cancer. **(D, E)** KEGG and GSEA enrichment analyses showing *CASP8*-related signaling pathways and biological processes in glioma. **(F)** The K-M curves suggested that the high *CASP8* expression was associated with low overall survival of glioma patients. **(G, H)** The protein expression of CASP8 in astrocyte NHA and five human glioma cell lines was detected by WB assay, and the protein expression level was quantified in the histogram. The CASP8 protein expression level is lower in glial cell than glioma cell lines. **(I, J)** IHC stain was performed on the 15 glioma patient tissues. Compared with the adjacency, the positive IHC signal of CASP8 was more in the glioma tissues. * p<0.05; ** p<0.01; *** p<0.001 and **** p<0.0001.

S5B) as well as immune cell infiltration (Figures S5C, D) by using CIBERSORT and GSVA algorithms. Next, the expression analysis of CASP8 in human astrocyte NHA and 5 immortalized human glioma cell lines (LN-229, U87-MG, A172, U118-MG, and U251-MG) showed that CASP8 expression in astrocytes was lower than that in 5 human glioma cell lines (Figures 10G, H). IHC experiment on 15 glioma patient tissues also demonstrated that CASP8 expression in glioma tissues was higher than that in paracancerous tissues (Figures 10I, J), further confirming the above analysis of *CASP8* expression in TCGA.

It has been reported that *CASP8* can initiate apoptosis while inhibiting necroptosis (41). However, our results showed that *CASP8* was highly expressed in glioma. Through the ssGSEA score of glioma patients in the TCGA database based on 159 NRGs, we found that the activated necroptosis was closely related to the poor patients' prognosis, suggesting that necroptosis in glioma might be regulated by other genes than *CASP8*. Therefore, we analyzed the expression patterns of core genes in the KEGG necroptosis pathway between glioma and normal brain tissues (Figure S6) and found several upstream genes, such as *TNF receptor superfamily member 1A* (*TNFR1*), *TNF-related apoptosis-inducing ligand receptor* (*TRAILR*), and *interferon production regulator* (*IFNR*) were highly expressed in glioma. Consistently, two vital effector molecules, RIPK1 and RIPK3 were also highly expressed in glioma, suggesting that initiation of necroptosis may be critically regulated by these highly expressed upstream NRGs in glioma.

Discussion

Mutations in proto-oncogenes and suppressor genes are directly linked to tumor occurrence, which leads to the shutdown of tumor-suppressing signaling and continual activation of tumor-proliferating signaling, resulting in abnormal cell proliferation. At present, with advances in the investigation of the tumor microenvironment (TME) and TIME, researchers have found cancer progression to be related to numerous immunosuppressive cell infiltrations in TIME. Meanwhile, how to inhibit tumor progression by cell death induction has always been a research focus. A variety of small-molecule inhibitors and chemotherapy drugs targeting key regulatory molecules of cell death processes, such as apoptosis, ferroptosis, pyroptosis, and necroptosis, have shown clinical benefits to patients (44, 45). However, the immune suppression caused by the recruitment of immune cells to TIME induced by DAMP release after cell necroptosis has not been fully studied. It has been reported that some drugs can inhibit glioma cell proliferation by inducing necroptosis (46–51). Unfortunately, these findings only rely on *in vitro* studies, which fail to reflect the interaction between glioma necroptosis and TIME. To date, no studies have investigated the relationship between necroptosis and glioma TIME at a single-cell level.

In the present study, we found that some NRGs were highly expressed in glioma and the ssGSEA score suggested an association between activated necroptosis and poor prognosis in glioma patients. This may be consistent with the finding of a previous study that reported an association between RIPK1 overexpression and glioma progression (52). Next, an NRS consisting of 18 NRGs was established and the predictive performance of the NRS was evaluated. Then, the NRS was well validated in two external CGGA datasets, by combining the results, the NRS showed a good performance in assessing the prognosis. Moreover, through the functional enrichment analysis, we found that angiogenesis, KRAS, and WNT signaling pathways were activated in the high-risk group, and KEGG and GO enrichment analysis showed that the DGEs between NRS groups were mainly associated with immune and inflammatory biological functions. Based on these results, we analyzed the different immune cell infiltration to investigate the relationship between necroptosis and TIME in glioma. We found that Tregs, M2 macrophages, and MDSCs were significantly infiltrated in the high-risk group. Combining the TIME analysis and the high expression of NRGs in glioma, we speculated that necroptosis in glioma may enhance the immune response and eliminate own necrotic tumor cells, however, it may probably promote immunosuppression in TIME as well (25, 53–56).

Several studies have reported that injecting necroptotic tumor cells, or engineered cells to specifically overexpress RIPK3 into mouse tumor models leads to killer T cell recruitment to TIME and attack tumor cells. Next, the combination of this therapy with PD-1 immunotherapy could result in long-lasting tumor clearance (57). But, in contrast, some studies have also pointed out that RIPK1 and RIPK3-drived tumor cell necroptosis could induce CXCL1 and SAP130 release that leads to mincle ligation activation and MDSC and M2 macrophage infiltration, which promotes and accelerates immunosuppressive and tumorigenesis, respectively. Our viewpoint is that at an early stage, necroptosis may remove tumor cells, whereas, at a later stage, the immunosuppressive TIME driven by necroptosis contributes to immune escape and results in poor prognosis of glioma patients (8, 25, 58–60). Later, in order to investigate the relationship between immunotherapy and NRS, we analyzed the expression difference of classical immune-associated factors in different NRS groups. Among them, most of the HLA, checkpoints, chemokines, costimulatory molecules, IFNG, CD8, and CD274 were highly expressed in the high-risk group, which showed that the high-risk group may get more chances from the anti-glioma immunotherapy. Such a founding may bring the poor prognosis group some prospect. We discovered two necroptosis-related subtypes by using the CNMF method and assessed them. The two subtypes held different NRGs'

expression patterns, the cluster 2 expressed more immune and inflammatory genes and infiltered a much number of immune cells than did in cluster 1; however, cluster 2 presented a poorer prognosis that may echo the above discussion about the inhibitory glioma TIME associated with necroptosis.

Thereafter, we used the scRNA-seq data to further study the necroptosis in GBM and found through the ssGSEA score that necroptosis was activated in both microglia and certain malignant GBM cell subsets, the necroptotic cells infiltrating in glioma TME may exacerbate the inhibitory TIME. In the end, through differential expression analysis and random forest feature screening, the core intersection gene CASP8 was screened out and the comprehensively functional exploration was conducted on it. We found *CASP8* was significantly overexpressed in a variety of cancers, including glioma. Furthermore, the result indicated that the OS of glioma patient groups with high *CASP8* expression was evidently shortened, and positively correlated with multiple inhibitory immune checkpoints. The three seemingly contradictory findings including the activated necroptosis pathway in glioma patients with poor prognosis and high *CASP8* expression in glioma as well as classical negative regulation of necroptosis by *CASP8* imply that the post-translational modification (PTM) of *CASP8* may play a significant role in necroptotic glioma. As expected, studies have indicated that CASP8 phosphorylation (mediated by ribosomal protein S6 kinase (RSK) recruited into necrosome) at Thr265 can stabilize the necrosome and relieve the inhibitory effect on necroptosis caused by CASP8 (61, 62). However, whether CASP8 expression can be affected by other PTMs, such as ubiquitination, SUMOylation, and glycosylation thereby impacting necroptosis needs further investigation. In addition, the highly expressed *CASP8* in gliomas may be a feedback regulatory mechanism evolved by glioma to avoid necroptosis.

Our study may have some limitations due to the lack of patients' specific clinical information and follow-up data, and all data derived from public databases. Besides, the difference between gene and protein expression levels might be influenced by several complex biological processes. Furthermore, our hypothesis of which immunosuppression is caused by necroptosis needs further verification. We deem that our study may bring more light to our understanding of how the different necroptosis stages impact the glioma progression.

In conclusion, through analyzing gene expressions and clinical characteristics of glioma patients in the TCGA dataset, we found that certain NRGs harbored mutations and overexpressed in glioma. Besides, a novel NRS was developed, which can effectively assessed the prognosis in glioma patients. Our study further indicated that the activated necroptosis pathway was related to poor prognosis, and multiple immunosuppressive cells were highly infiltrated in the high-risk group. We assumed that the poor prognosis caused by necroptosis may be associated with immunosuppressive TIME. Through scRNA-seq data, we also reported that the necroptosis pathway in GBM and microglial was activated, suggesting that necroptosis in the glioma TIME might promote glioma progression. Our findings may provide guidance for the study of immune escape induced by necroptosis in glioma, and give a new scheme for glioma prognosis prediction. CASP8 high expression seems to be one of the mechanism by which glioma escapes necroptosis and therefore it represents a potential biomarker for glioma prognosis.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the ethics committee of Southwest University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

HC and PL supervised this research. SW, UAEM, RL, CL, and KW performed the research. SW and UAEM designed the outlines of this research and wrote the manuscript. LD reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Natural Science Foundation of Chongqing (cstc2022ycjh-bgzxm0145, cstc2022ycjh-bgzxm0016, cstc2019jcyj-zdxmX0033), the pilot program of Southwest University (SWU-XDZD22006), and the Fundamental Research Funds for the Central Universities (SWU120054).

Acknowledgments

Sincerely thanks to the funding support institutions, the members of Prof. Cui's lab, and assistance of single-cell RNA-seq data analysis from Beijing Gaptech 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

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1013094/full#supplementary-material

References

1. Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, et al. Glioma. *Nat Rev Dis Primers* (2015) 1:15017. doi: 10.1038/nrdp.2015.17

2. Nicholson JG, Fine HA. Diffuse glioma heterogeneity and its therapeutic implications. *Cancer Discovery* (2021) 11(3):575–90. doi: 10.1158/2159-8290.CD-20-1474

3. Gusyatiner O, Hegi ME. Glioma epigenetics: From subclassification to novel treatment options. *Semin Cancer Biol* (2018) 51:50–8. doi: 10.1016/j.semcancer.2017.11.010

4. Pasparakis M, Vandenabeele P. Necroptosis and its role in inflammation. *Nature* (2015) 517(7534):311–20. doi: 10.1038/nature14191

5. Vandenabeele P, Galluzzi L, Berghe TV, Kroemer G. Molecular mechanisms of necroptosis: an ordered cellular explosion. *Nat Rev Mol Cell Biol* (2010) 11 (10):700–14. doi: 10.1038/nrm2970

6. Jiang M, Qi L, Li L, Wu Y, Song D, Li Y, et al. Caspase-8: A key protein of cross-talk signal way in "PANoptosis" in cancer. *Int J Cancer* (2021) 149(7):1408–20. doi: 10.1002/ijc.33698

7. Gong Y, Fan Z, Luo G, Yang C, Huang Q, Fan K, et al. The role of necroptosis in cancer biology and therapy. *Mol Cancer* (2019) 18(1):100. doi: 10.1186/s12943-019-1029-8

8. Seifert L, Werba G, Tiwari S, Ly NNG, Alothman S, Alqunaibit D, et al. The necrosome promotes pancreatic oncogenesis *via* CXCL1 and mincle-induced immune suppression. *Nature* (2016) 532(7598):245–9. doi: 10.1038/nature17403

9. Lin S-Y, Hsieh S-Y, Fan Y-T, Wei W-C, Hsiao P-W, Tsai D-H, et al. Necroptosis promotes autophagy-dependent upregulation of DAMP and results in immunosurveillance. *Autophagy* (2018) 14(5):778–95. doi: 10.1080/15548627.2017.1386359

10. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data *via* the xena platform. *Nat Biotechnol* (2020) 38(6):675–8. doi: 10.1038/s41587-020-0546-8

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

12. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* (2021) 184(13):3573–3587.e29. doi: 10.1016/j.cell.2021.04.048

13. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd , et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7):1888–1902.e21. doi: 10.1016/j.cell.2019.05.031

14. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* (2018) 36(5):411–20. doi: 10.1038/nbt.4096

15. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* (2015) 33(5):495–502. doi: 10.1038/nbt.3192

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

17. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28 (11):1747–56. doi: 10.1101/gr.239244.118

18. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. In: *Nucleic acids res*, (NCCary2001 Evans Rd, 27513: Oxford University Press) vol. 49. (2021). p. D545–d551.

19. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models *via* coordinate descent. *J Stat Softw* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01

20. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model *via* coordinate descent. *J Stat Softw* (2011) 39(5):1–13. doi: 10.18637/jss.v039.i05

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

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

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

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

25. Tan R, Zhang G, Liu R, Hou J, Dong Z, Deng C, et al. Identification of early diagnostic and prognostic biomarkers *via* WGCNA in stomach adenocarcinoma. *Front Oncol* (2021) 11:636461. doi: 10.3389/fonc.2021.636461

26. Hou J, Huang P, Lan C, Geng S, Xu M, Liu Y, et al. ZC3H15 promotes gastric cancer progression by targeting the FBXW7/c-myc pathway. *Cell Death Discovery* (2022) 8(1):32. doi: 10.1038/s41420-022-00815-x

27. Tanaka A, Sakaguchi S. Regulatory T cells in cancer immunotherapy. *Cell Res* (2017) 27(1):109–18. doi: 10.1038/cr.2016.151

28. Jordan JT, Sun W, Hussain SF, DeAngulo G, Prabhu SS, Heimberger AB, et al. Preferential migration of regulatory T cells mediated by glioma-secreted chemokines can be blocked with chemotherapy. *Cancer Immunol Immunother* (2008) 57(1):123–31. doi: 10.1007/s00262-007-0336-x

29. Zhou W, Ke SQ, Huang Z, Flavahan W, Fang X, Paul J, et al. Periostin secreted by glioblastoma stem cells recruits M2 tumour-associated macrophages and promotes malignant growth. *Nat Cell Biol* (2015) 17(2):170–82. doi: 10.1038/ncb3090

30. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: Tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol* (2002) 23(11):549–55. doi: 10.1016/S1471-4906(02)02302-5

31. Zhang K, Fu G, Pan G, Li C, Shen L, Hu R, et al. Demethylzeylasteral inhibits glioma growth by regulating the miR-30e-5p/MYBL2 axis. *Cell Death Dis* (2018) 9 (10):1035. doi: 10.1038/s41419-018-1086-8

32. Bruger AM, Dorhoi A, Esendagli G, Barczyk-Kahlert K, van der Bruggen P, Lipoldova M, et al. How to measure the immunosuppressive activity of MDSC: assays, problems and potential solutions. *Cancer Immunol Immunother* (2019) 68 (4):631–44. doi: 10.1007/s00262-018-2170-8

33. Gabrilovich DI. Myeloid-derived suppressor cells. *Cancer Immunol Res* (2017) 5(1):3–8. doi: 10.1158/2326-6066.CIR-16-0297

34. Mi Y, Guo N, Luan J, Cheng J, Hu Z, Jiang P, et al. The emerging role of myeloid-derived suppressor cells in the glioma immune suppressive microenvironment. *Front Immunol* (2020) 11:737. doi: 10.3389/fimmu.2020.00737

35. De Leo A, Ugolini A, Veglia F. Myeloid cells in glioblastoma microenvironment. *Cells* (2020) 10(1). doi: 10.3390/cells10010018

36. Ghouzlani A, Kandoussi S, Tall M, Reddy KP, Rafii S, Badou A, et al. Immune checkpoint inhibitors in human glioma microenvironment. *Front Immunol* (2021) 12:679425. doi: 10.3389/fimmu.2021.679425

37. Sampson JH, Gunn MD, Fecci PE, Ashley DM. Brain immunology and immunotherapy in brain tumours. *Nat Rev Cancer* (2020) 20(1):12–25. doi: 10.1038/s41568-019-0224-7

38. Yang K, Wu Z, Zhang H, Zhang N, Wu W, Wang Z, et al. Glioma targeted therapy: insight into future of molecular approaches. *Mol Cancer* (2022) 21(1):39. doi: 10.1186/s12943-022-01513-z

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

40. Xu T, Le TD, Liu L, Su N, Wang R, Sun B, et al. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics* (2017) 33(19):3131–3. doi: 10.1093/bioinformatics/btx378

41. Fritsch M, Günther SD, Schwarzer R, Albert MC, Schorn F, Werthenbach JP, et al. Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis. *Nature* (2019) 575(7784):683–7. doi: 10.1038/s41586-019-1770-6

42. Hambardzumyan D, Gutmann DH, Kettenmann H. The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci* (2016) 19 (1):20–7. doi: 10.1038/nn.4185

43. Fyfe I. Microglia subset associated with high-grade glioma. *Nat Rev Neurol* (2021) 17(11):660. doi: 10.1038/s41582-021-00580-y

44. Chen X, Kang R, Kroemer G, Tang D. Broadening horizons: the role of ferroptosis in cancer. *Nat Rev Clin Oncol* (2021) 18(5):280–96. doi: 10.1038/s41571-020-00462-0

45. Diepstraten ST, Anderson MA, Czabotar PE, Lessene G, Strasser A, Kelly GL, et al. The manipulation of apoptosis for cancer therapy using BH3-mimetic drugs. *Nat Rev Cancer* (2022) 22(1):45–64. doi: 10.1038/s41568-021-00407-4

46. Zhou J, Li G, Han G, Feng S, Liu Y, Chen J, et al. Emodin induced necroptosis in the glioma cell line U251 *via* the TNF-α/RIP1/RIP3 pathway. *Invest New Drugs* (2020) 38(1):50–9. doi: 10.1007/s10637-019-00764-w

47. Ding Y, He C, Lu S, Wang X, Wang C, Wang L, et al. MLKL contributes to shikonin-induced glioma cell necroptosis *via* promotion of chromatinolysis. *Cancer Lett* (2019) 467:58–71. doi: 10.1016/j.canlet.2019.09.007

48. Pagano C, Navarra G, Coppola L, Avilia G, Pastorino O, Monica RD, et al. N6-isopentenyladenosine induces cell death through necroptosis in human glioblastoma cells. *Cell Death Discovery* (2022) 8(1):173. doi: 10.1038/s41420-022-00974-x

49. Feng Y, Wang W, Zhang Y, Fu X, Ping K, Zhao J, et al. Synthesis and biological evaluation of celastrol derivatives as potential anti-glioma agents by activating RIP1/RIP3/MLKL pathway to induce necroptosis. *Eur J Med Chem* (2022) 229:114070. doi: 10.1016/j.ejmech.2021.114070

50. Zhang G, Zhu Q, Yan X, Ci M, Zhao E, Hou J, et al. HECTD3 promotes gastric cancer progression by mediating the polyubiquitination of c-MYC. *Cell Death Discovery* (2022) 8(1):185. doi: 10.1038/s41420-022-01001-9

51. Li C, Zhang K, Pan G, Ji H, Li C, Wang X, et al. Dehydrodiisoeugenol inhibits colorectal cancer growth by endoplasmic reticulum stress-induced autophagic pathways. *J Exp Clin Cancer Res* (2021) 40(1):125. doi: 10.1186/s13046-021-01915-9

52. Park S, Hatanpaa KJ, Xie Y, Mickey BE, Madden CJ, Raisanen JM, et al. The receptor interacting protein 1 inhibits p53 induction through NF-kappaB activation and confers a worse prognosis in glioblastoma. *Cancer Res* (2009) 69 (7):2809–16. doi: 10.1158/0008-5472.CAN-08-4079

53. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol Cancer* (2021) 20(1):131. doi: 10.1186/s12943-021-01428-1

54. Ma Q, Long W, Xing C, Chu J, Luo M, Wang HY, et al. Cancer stem cells and immunosuppressive microenvironment in glioma. *Front Immunol* (2018) 9:2924. doi: 10.3389/fimmu.2018.02924

55. Zhang G, Zhu Q, Fu G, Hou J, Hu X, Cao J, et al. TRIP13 promotes the cell proliferation, migration and invasion of glioblastoma through the FBXW7/c-MYC axis. *Br J Cancer* (2019) 121(12):1069–78. doi: 10.1038/s41416-019-0633-0

56. Hu X, Pei D, Ci M, Zhang G, Li B, Wang J, et al. ACTL6A deficiency induces apoptosis through impairing DNA replication and inhibiting the ATR-Chk1 signaling in glioblastoma cells. *Biochem Biophys Res Commun* (2022) 599:148–55. doi: 10.1016/j.bbrc.2022.01.124

57. Snyder AG, Hubbard NW, Messmer MN, Kofman SB, Hagan CE, Orozco SL, et al. Intratumoral activation of the necroptotic pathway components RIPK1 and RIPK3 potentiates antitumor immunity. *Sci Immunol* (2019) 4(36). doi: 10.1126/sciimmunol.aaw2004

58. Li C, Deng C, Pan G, Wang X, Zhang K, Dong Z, et al. Lycorine hydrochloride inhibits cell proliferation and induces apoptosis through promoting FBXW7-MCL1 axis in gastric cancer. *J Exp Clin Cancer Res* (2020) 39(1):230. doi: 10.1186/s13046-020-01743-3

59. Pan G, Zhang K, Geng S, Lan C, Hu X, Li C, et al. PHF14 knockdown causes apoptosis by inducing DNA damage and impairing the activity of the damage response complex in colorectal cancer. *Cancer Lett* (2022) 531:109–23. doi: 10.1016/j.canlet.2022.01.002

60. Hu X, Liu R, Hou J, Peng W, Wan S, Xu M, et al. SMARCE1 promotes neuroblastoma tumorigenesis through assisting MYCN-mediated transcriptional activation. *Oncogene* (2022) 41(37):4295–306. doi: 10.1038/s41388-022-02428-1

61. Yang ZH, Wu XN, He P, Wang X, Wu J, Ai T, et al. A non-canonical PDK1-RSK signal diminishes pro-caspase-8-Mediated necroptosis blockade. *Mol Cell* (2020) 80(2):296–310.e6. doi: 10.1016/j.molcel.2020.09.004

62. Mandal R, Barrón JC, Kostova I, Becker S, Strebhardt K. Caspase-8: The double-edged sword. *Biochim Biophys Acta Rev Cancer* (2020) 1873(2):188357. doi: 10.1016/j.bbcan.2020.188357

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 29 November 2022
DOI 10.3389/fimmu.2022.1007461

OPEN ACCESS

EDITED BY
Ping Zheng,
The University of Melbourne, Australia

REVIEWED BY
Shanze Chen,
Jinan University, China
Dinglan Wu,
Southern Medical University, China

*CORRESPONDENCE
Quan Cheng
chengquan@csu.edu.cn
Renjun Peng
prjscience@csu.edu.cn

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

RECEIVED 30 July 2022
ACCEPTED 04 October 2022
PUBLISHED 29 November 2022

CITATION
Xiao G, Wang K, Wang Z, Dai Z, Liang X, Ye W, Luo P, Zhang J, Liu Z, Cheng Q and Peng R (2022) Machine learning-based identification of SOX10 as an immune regulator of macrophage in gliomas. *Front. Immunol.* 13:1007461. doi: 10.3389/fimmu.2022.1007461

Machine learning-based identification of SOX10 as an immune regulator of macrophage in gliomas

Gelei Xiao[1,2,3,4], Kaiyue Wang[5], Zeyu Wang[1,6], Ziyu Dai[1], Xisong Liang[1], Weijie Ye[7], Peng Luo[8], Jian Zhang[8], Zaoqu Liu[9], Quan Cheng[1,4,7]* and Renjun Peng[1,4]*

[1]Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, China, [2]Diagnosis and Treatment Center for Hydrocephalus, Xiangya Hospital, Central South University, Changsha, China, [3]Hunan International Scientific and Technological Cooperation Base of Brain Tumor Research, Xiangya Hospital, Central South University, Changsha, China, [4]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [5]Xiangya School of Medicine, Central South University, Changsha, China, [6]MRC Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, United Kingdom, [7]Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China, [8]Department of Oncology, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [9]Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Gliomas, originating from the glial cells, are the most lethal type of primary tumors in the central nervous system. Standard treatments like surgery have not significantly improved the prognosis of glioblastoma patients. Recently, immune therapy has become a novel and effective option. As a conserved group of transcriptional regulators, the Sry-type HMG box (SOX) family has been proved to have a correlation with numerous diseases. Based on the large-scale machine learning, we found that the SOX family, with significant immune characteristics and genomic profiles, can be divided into two distinct clusters in gliomas, among which SOX10 was identified as an excellent immune regulator of macrophage in gliomas. The high expression of SOX10 is related to a shorter OS in LGG, HGG, and pan-cancer groups but benefited from the immunotherapy. It turned out in single-cell sequencing that SOX10 is high in neurons, M1 macrophages, and neural stem cells. Also, macrophages are found to be elevated in the SOX10 high-expression group. SOX10 has a positive correlation with macrophage cytokine production and negative regulation of macrophages' chemotaxis and migration. In conclusion, our study demonstrates the outstanding cluster ability of the SOX family, indicating that SOX10 is an immune regulator of macrophage in gliomas, which can be an effective target for glioma immunotherapy.

KEYWORDS

gliomas, SOX10, immunotherapy, immune infiltration, immune microenvironment

10.3389/fimmu.2022.1007461

Introduction

Gliomas, originating from the glial cells, are the most lethal type of primary tumors in the central nervous system (CNS) (1). According to the WHO classification criteria, they are classified into four types inferred by malignancy (2). In depth, grade II and III gliomas are classified as lower-grade gliomas (LGG), grade IV (glioblastoma, GBM) as higher-grade gliomas (HGG), by The Cancer Genome Atlas (TCGA). For most cases, LGG with the isocitrate dehydrogenase (IDH) mutant for the metabolic enzymes has a conspicuously better prognosis than the IDH wild type, which are generally GBMs. To date, the standard treatment contains surgery has not significantly improved the prognosis and median overall survival (OS) of GBM patients (3). As a consequence, a new and effective therapy is of urgency.

Recent studies have found that as a constitutive part of the tumor microenvironment (TME), tumor cells, stromal cells, and infiltrating immune cells all serve a variety of biologically important roles in glioma proliferation, progression, and prognosis (4). Moreover, we and others have previously suggested several immune-related prognostic biomarkers to predict prognosis and immunotherapy efficacy perfectly (5, 6). These may all contribute to the immune therapy of glioma.

Sry-type HMG box (SOX) family proteins are a conserved group of transcriptional regulators depending on the high-mobility group (HMG) domain to bind with DNA (7). The SOX family has been revealed to have the correlation with numerous diseases (8). Almost all SOX genes, for instance, SOX1, SOX2, SOX7, and SOX10, have been found to have the potential to regulate the progression of glioma, whose expression levels are also related to the prognosis of patients (9–12). SOX genes play an important role in this regulation, which are found to be involved in the maintenance of the stemness or differential initiation of glioma stem cells (13). For example, knockdown of SOX1 expression in glioma stem cells has been found to impair the self-renewal, proliferation, viability, and tumorigenesis ability of glioma cells, while the overexpression of SOX1 promoted the malignant phenotype of glioma (9). However, the overexpression of SOX11 prevents tumorigenic ability in glioma-initiating cell-like cells and human glioma-initiating cells derived from malignant gliomas by inducing neuronal differentiation (14). Moreover, previous studies have confirmed that SOX is closely associated with the TME (15). SOX genes in tumor cells influence the infiltration of immune cells *via* paracrine signals, and vice versa (16). By giving tumor cells the ability to evade NK cells, SOX2 and SOX9 have been found to promote the immune evasion of tumor cells (17, 18). Therefore, the SOX family is crucial to the development of gliomas, especially in the aspect of the TME and immunotherapy. However, it remains unclear which one, as well as the detailed function, of the SOX family plays the leading role in glioma.

Herein, our study extracted data from bulk tumor (The Cancer Genome Atlas, TCGA; the Chinese Glioma Genome Atlas, CGGA) and single-cell mRNA-seq databases (SCP50 and SCP393; http://singlecell.broadinstitute.org). Cluster analysis was performed, and SOX10 was identified as a distinguished biomarker to explore the prognostic value and association with the glioma immune microenvironment.

Materials and Methods

Data collection and preprocessing

1685 samples of diffuse glioma related data and complete clinicopathological annotations were obtained from two datasets: TCGA (https://xenabrowser.net/) and CGGA (http://www.cgga.org.cn/). 672 samples in TCGA were used as the training set, while 1013 samples in CGGA were used as the validation set. We excluded samples with insufficient OS. The RNA-sequencing data, SCP50 and SCP393, was collected form Single Cell Portal platform (http://singlecell.broadinstitute.org). To possessing a similar signal intensity with the RMA- processed values, the fragments per kilobase million (FPKM) values were transformed into transcripts per kilobase million (TPM) values.

Genomic alteration

We obtained the somatic mutation and copy number variant (CNV) profiles from TCGA dataset. We used GISTIC 2.0 analysis (https://cloud.genepattern.org) to assess the landscape of CNV, including the frequency of function mutation gain or loss at the amplified or deleted peaks.

Unsupervised consensus clustering for the SOX family and the selection of SOX10

Using the ConsensusClusterPlus R package, we determined the optimal cluster number and their constancy and authenticity in TCGA cohort and meta-cohort. We performed principal component analysis (PCA) to ensure the clustering tendency. The LASSO-LR algorithm, Pamr algorithm, random forest algorithm, XGboost algorithm, and Boruta algorithm were used to screen out the most characteristic genes, SOX10.

TME immunological characteristic analysis

The Estimation of STromal and Immune cells in MAlignant Tumors using Expression data (ESTIMATE) algorithm was used

to estimate the stromal score, immune score, and estimate score of the infiltrating immune cells in the TME. The Tumor Immune Estimation Resource 2.0 (TIMER2.0; http://timer.cistrome.org/) web server was used to thoroughly evaluate the degree of immune infiltrating cells in gliomas. We used the xCell algorithm to ascertain the enrichment levels of 64 types of immune cells. The proportions of 22 types of TME cells in tumor tissues were evaluated by the CIBERSORT algorithm. The R gene set variation analysis (GSVA) package was implemented to calculate enrichment scores by single-sample gene set enrichment analysis (ssGSEA). Besides, the EPIC algorithm, MCPcounter algorithm, and QuantSeq algorithm were also executed to estimate the immune infiltrating cell abundance. Using ssGSEA, we evaluated the seven steps of cancer immune cascades. This immunity cycle determined the destination of tumor cells and reflected the immune response of the anticancer. The subMap algorithm was used to evaluate the response to therapies of anti-CTLA4 and anti-PD1. Also, the Tumor Immune Dysfunction and Exclusion (TIDE) (http://tide.dfci.harvard.edu/setquery/) and Tumor Immune Syngeneic MOuse (TISMO) (http://tismo.cistrome.org) algorithm was utilized for deducing the immune checkpoint blockade immunotherapy responses in gliomas.

Single-cell sequencing

R package Seurat was employed to process the single-cell data expression matrix. We used "NormalizeData" to renormalize data. Then, 2,000 highly changeable genes were identified by "FindIntegrationAnchors." "FindIntegrationAnchors" and the "Intergratedata" were used to merge GBM sample data sets. "RunPCA" and "FindNeighbors" was used to perform PCA. Afterwards, to alternately combine cells together, we used the "FindClusters" function. Finally, to visualize the analyses, "UMAP" was performed.

Multiplex immunofluorescence staining

We purchased the glioma tissue array from Wuhan Tanda Scientific Co., Ltd. (NGL1021), with ethics approvement. SOX10 (Mouse, 1:100, Proteintech, China), CD163 (Rabbit, 1:3,000, Proteintech, China), and CD68 (Rabbit, 1:3,000, Servicebio, China) were the primary Abs. Horseradish peroxidase-conjugated secondary antibody incubation (GB23301, GB23303, Servicebio, China) was the secondary antibody. The tyramide signal was amplified into TSA [FITC-TSA, CY3-TSA, 594-TSA, and 647-TSA (Servicebio, China)]. The stained slides were scanned using the TissueFAXS platform (TissueGnostics, Vienna, Austria). The spatial analysis of the stained cells was performed using the StrataQuest software (TissueGnostics, Vienna, Austria).

Drug response prognostication

All pharmacogenomic data were downloaded from the Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/). The semi-inhibition rate (IC50) reckoned by the pRRophetic R package was utilized to predict the drug susceptibilities and responses.

Statistical analysis

The overall survival of divergent groups was assessed by Kaplan–Meier curves (KM curves) with the log-rank test. All OS curves were produced by the survminer R package. Mutation landscape OncoPrint was executed by the maftools R package. Heatmaps were pictured found on the R package complexHeatmap. Student's t-test was conducted to analysis normally distributed variables between the two groups while one-way analysis of variance (ANOVA) was conducted to analysis normally distributed variables between multiple groups. The Wilcoxon test was applied to analysis non-normally distributed variables between the two groups while Kruskal-Wallis test was applied to analysis normally distributed variables between multiple groups. R 3.6.3 was used to conduct all statistical analyses. Statistics were considered significant when p-value< 0.05.

Results

Two distinct clusters of the SOX family

Firstly, we evaluated the clustering capabilities of the SOX family and visualized it (Figure 1A). To choose the ideal cluster number, the stability of clustering was appraised by the ConsensusClusterPlus package in TCGA (Figures 1B, C). It was found that k = 2, with the flattest CDF curve, is the optimal choice (Figure 1B). Then, clustering tendency was evaluated by principal component analysis (PCA). We used blue dots to represent cluster1, while red dots represent cluster2. SOX clusters were separated significantly, indicating a high-quality consensus cluster result (Figure 1D).

We further explored the overall survival of glioma patients in cluster1 and cluster2, $p < 0.001$. The Kaplan–Meier curves firmly demonstrated that cluster2 had higher and more prolonged survival than cluster1 (Figure 1E). Besides, Figures 1F, G show the global view of mutational distribution in cluster1 and cluster2, respectively. As a biomarker related to the malignancy of gliomas, IDH1 mutation took up 69% of the general in cluster2, much higher than that of cluster1, 44% (19, 20). Cellular tumor antigen p53 (TP53) alteration was presented similarly in cluster1 (37%) and cluster2 (45%). In cluster1, the

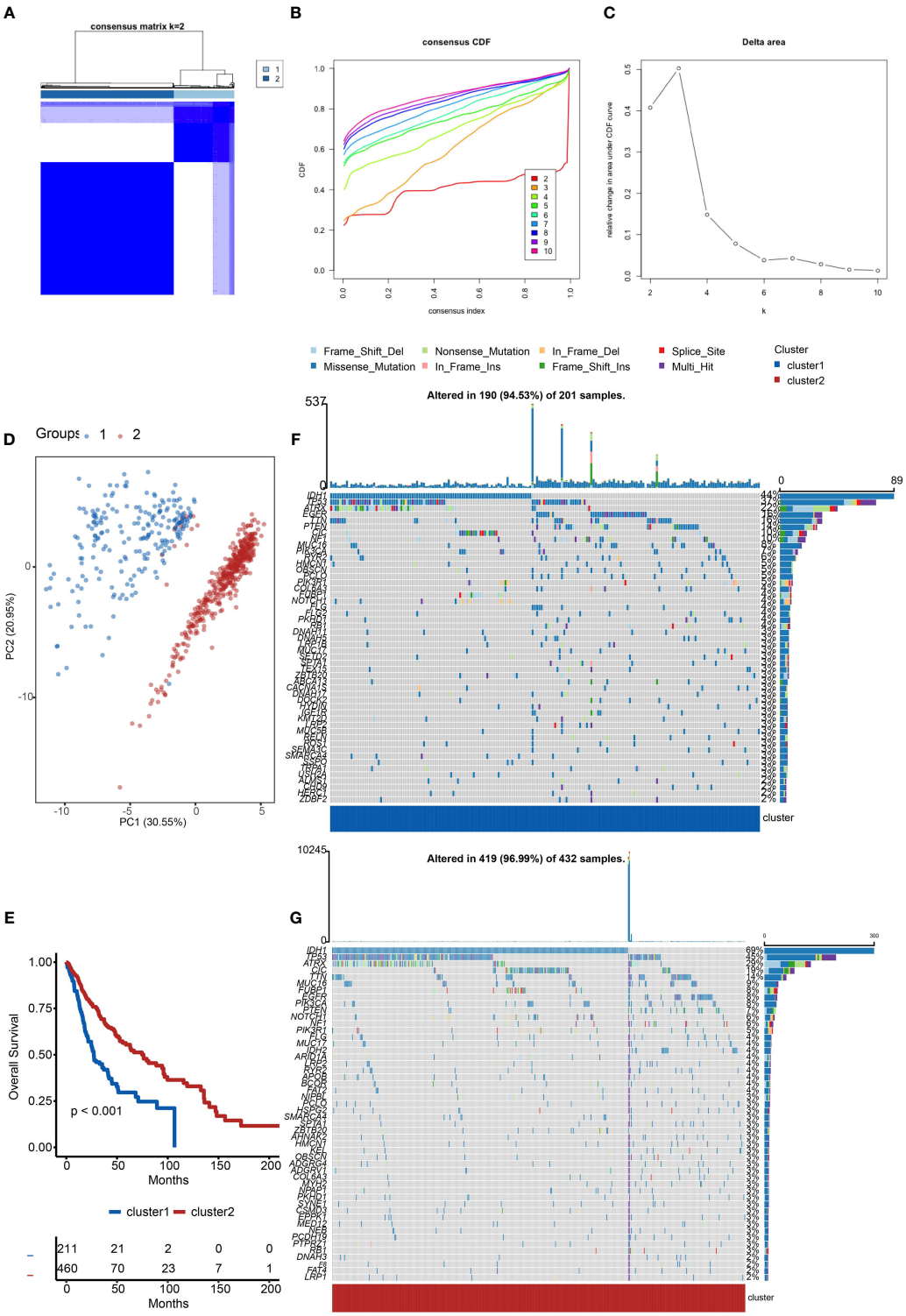


FIGURE 1
Cluster analysis of the SOX family. **(A)** Visualization of cluster analysis. **(B, C)** Determination of the number of clusters. **(D)** Significant separation of SOX clusters by PCA. Blue dots represent cluster1, while red dots represent cluster2. **(E)** Kaplan–Meier overall survival curves comparing cluster1 and cluster2 in gliomas. **(F)** Detection of the genes with the highest mutation frequency in cluster1. **(G)** Detection of the genes with the highest mutation frequency in cluster2.

following three genes ranked by frequency were alpha-thalassemia/mental retardation syndrome x-linked chromatin remodeler (ATRX) (22%), epidermal growth factor receptor (EGFR) (16%), and titin (TTN) (16%), while those in cluster2 were ATRX (29%), CIC (19%), and TTN (14%). In conclusion, the SOX family has a close correlation with the prognosis and proliferation of gliomas.

Immune characteristics of two clusters

We investigated the TME characteristics of cluster1 and cluster2. We evaluated the ESTIMATEScore, ImmuneScore, and StromalScore of the two clusters in TCGA dataset (Figures 2A–C). Among these three evaluations, scores of cluster1 were all higher than those of cluster2. Moreover, they could be thought to have a significant difference, on account of $p < 0.001$. Then, we calculated the proportion of five immune subtypes in the two clusters (Figure 2D). Immunologically Quiet was generally presented in cluster2 (more than 50%) and was partially observed in cluster1 (less than 50%). On the contrary, Lymphocyte Depleted was the frequentist immune subtype in cluster1, which took up over 50%. Moreover, we found that the TME indicator scores of cluster2 seemed to be lower than those of cluster1, which indicated a weaker immune response (Figure 2E). It also revealed the difference between the two clusters.

We calculated relating levels of 64 cell types by the xCell algorithm and clusters in TCGA (Figure 2F). We defined four subtypes of glioma: pro-neural (PN), classical (CL), neural (NE), and mesenchymal (ME), among which CL and ME are more severe (21). It is found that some types of cells are different in the two clusters with statistical significance. Plasma cells and neurons are more positively related to cluster2, while macrophages, macrophages M1, and macrophages M2 are more positively related to cluster1. Additionally, we used box plots to present the proportions of 22 TME cell types in tumor tissues with cluster1 and cluster2 (Figure 2G). Only four cell types had significant differences: B cells memory, macrophages M0, M1, and neutrophils. B cells memory in cluster2 were higher than in cluster1. Meanwhile, macrophages M1 and neutrophils in cluster2 were lower than in cluster1.

Distinct genomic profiles of the two clusters

Considering the apparent differences in overall survival and immune characteristics in cluster1 and cluster2, genomic profiles of the two clusters were supposed to be distinct. To validate it, we analyzed the co-occurrence/mutual exclusivity of the 25 most altered genes in cluster1 (Figure 3A) and cluster2 (Figure 3B). The strongest co-occurrent couples of gene mutation in cluster1 and cluster2 were IDH1 and ATRX, IDH1 and CIC, IDH1 and FUBP1, ATRX and TP53. IDH1 and EGFR were mutually exclusive pairs in cluster1 and cluster2. Higher co-occurrence is usually functionally linked to the proliferation of gliomas (22, 23). Then, we used a forest plot to list the 11 most variated genes between the two clusters (Figure 3C). Except for IDH1 and CIC, the other nine genes were more likely to mutate in cluster1.

Furthermore, we compared the frequency of different somatic mutations between the two clusters, including the single-nucleotide polymorphism (SNP), single nucleotide variant (SNV), deletion, insertion, and intergenic region (IGR). The frequency of insertion and deletion seemed to be non-statistically different, while SNPs were a little more common in cluster1 (Figure 3D). Among the identified SNVs, C was more presumably to mutate to T, which was also the most common mutation in cluster1 (Figure 3E). Transformation of splice region and missense were more common in cluster1 than in cluster2 (Figure 3F) (24). Amplifications and deletions of chr7 and chr10 have statistically differences in cluster1 and cluster2 (Figure 3G).

Identification of SOX10 as a prognostic gene

To distinguish the two clusters more accurately and precisely, we executed machine learning and prediction on the two populations, screening out the most characteristic genes. Using the LASSO-LR, XGboost, Boruta, Pamr, and RandomForest machine learning algorithms, we filtrated 15, 5, 11, 5, and 4 genes, correspondingly (Figures 4A–E). We used a Venn diagram to take the intersection of the five algorithms (Figure 4F). These two characteristic genes in the intersection corner, SOX10 and sex determining region Y (SRY), were the most potential to best classify the two clusters (25). Considering that SRY mainly depends male sex, we identified SOX10 as a biomarker of glioma prognosis.

The prognostic potential of SOX10

We performed a survival analysis of different SOX10 expressions in pan-glioma, LGG, and GBM based on TCGA and CGGA datasets (Figures 5A, B). The Kaplan–Meier curves more securely demonstrated that grievous survival mischief in glioma patients with high SOX10. However, the GBM Kaplan–Meier curves in TCGA were contrary to those in CGGA, which could account for the small number of samples of GBM patients in TCGA. Moreover, we predicted the value of SOX10, IDH, and subtype measured by receiver operating characteristic (ROC) curves in TCGA dataset (Figure 5C). The results firmly proved that SOX10 was a predictor of IDH and subtype. The ROC curves exhibited high sensitivity and specificity, with all areas under the curves (AUC) bigger than 0.7 and 0.9.

Additionally, to probe the latent pathological function of SOX10, the KEGG and GO enrichment analyses were performed. Figure 5D depicts 20 related pathways in the two selected pathways.

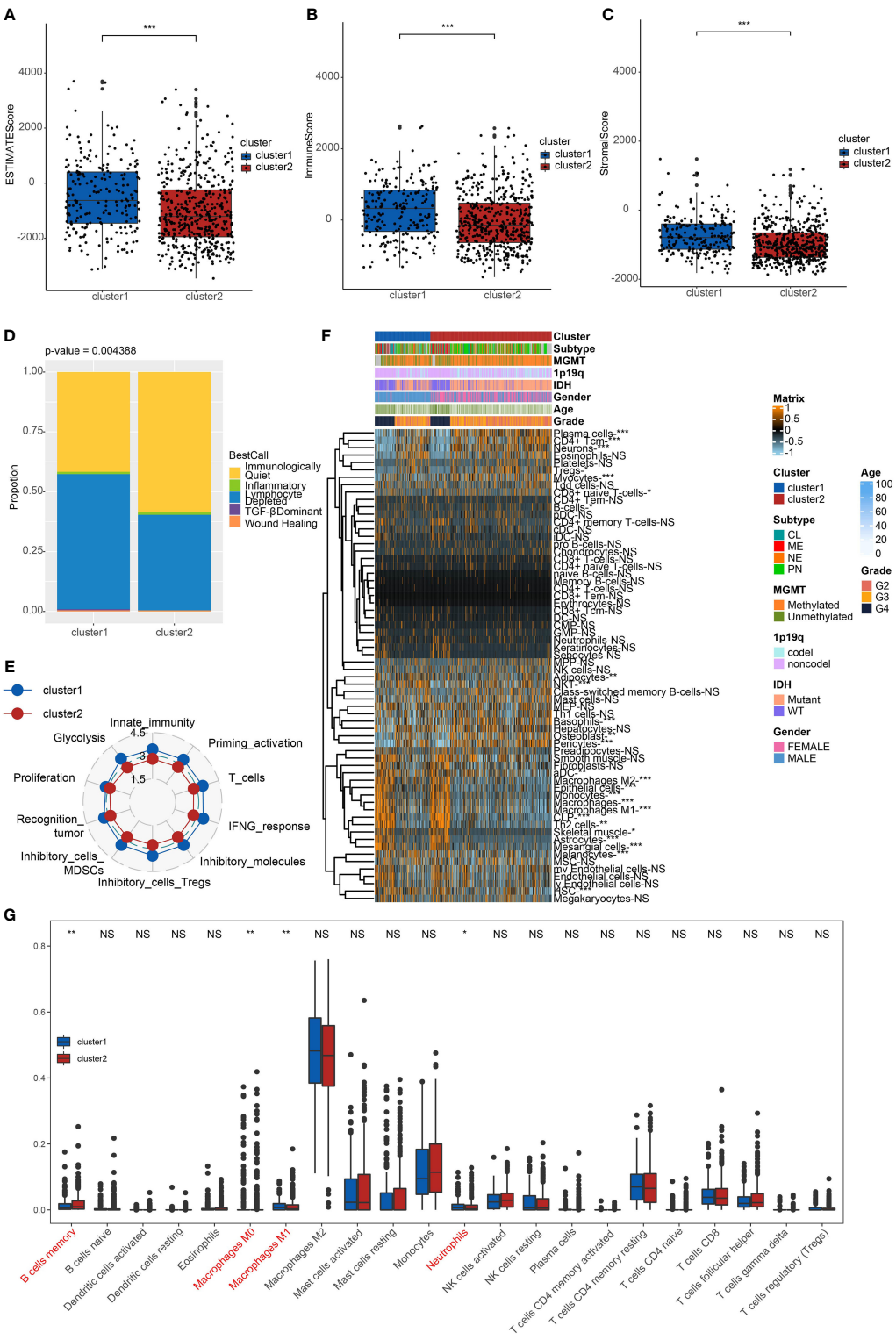


FIGURE 2
Immune characteristics of the clusters. **(A)** ESTIMATEScore **(B)** ImmuneScore. **(C)** StromalScore of cluster1 and cluster2 in TCGA. ***p < 0.001. **(D)** The proportion of five immune subtypes in cluster1 and cluster2. p = 0.004388. **(E)** The tendency of the difference between two clusters based on TME indicator scores. **(F)** Dendrogram corresponding to the 64-cell type level calculated by xCell and clusters in TCGA. *p < 0.05, **p < 0.01, ***p < 0.001. **(G)** Box plots of the proportions of 22 TME cell types in tumor tissues with cluster1 and cluster2. The dispersed dots represent values of TME cell expression in each cluster. **p < 0.01, *p < 0.05, NS, no significance.

10.3389/fimmu.2022.1007461

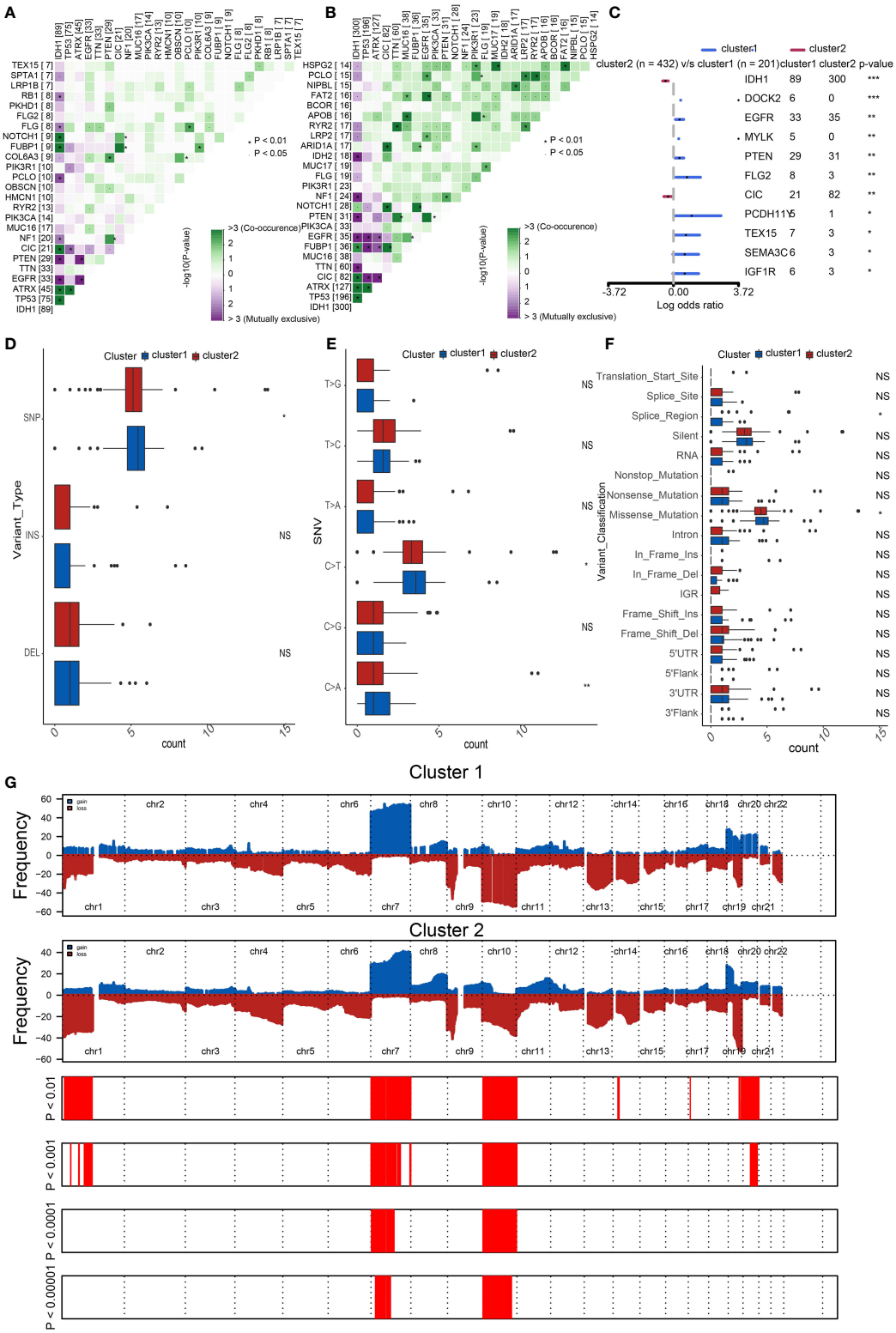


FIGURE 3
Distinct genomic profiles of the two clusters. The co-occurrence or mutual exclusivity of the top 25 most mutated genes in **(A)** cluster1 and **(B)** cluster2. *p < 0.01, `p < 0.05. **(C)** Demonstration of the 11 most altered genes between the clusters by the forest plot. Frequency comparison according to **(D)** variant type, **(E)** SNV, and **(F)** variant classification between the two clusters. **(G)** Amplifications and deletions in two clusters of SOX family by GISTIC 2.0. ***p < 0.001, **p < 0.01, *p < 0.05, NS, no significance.

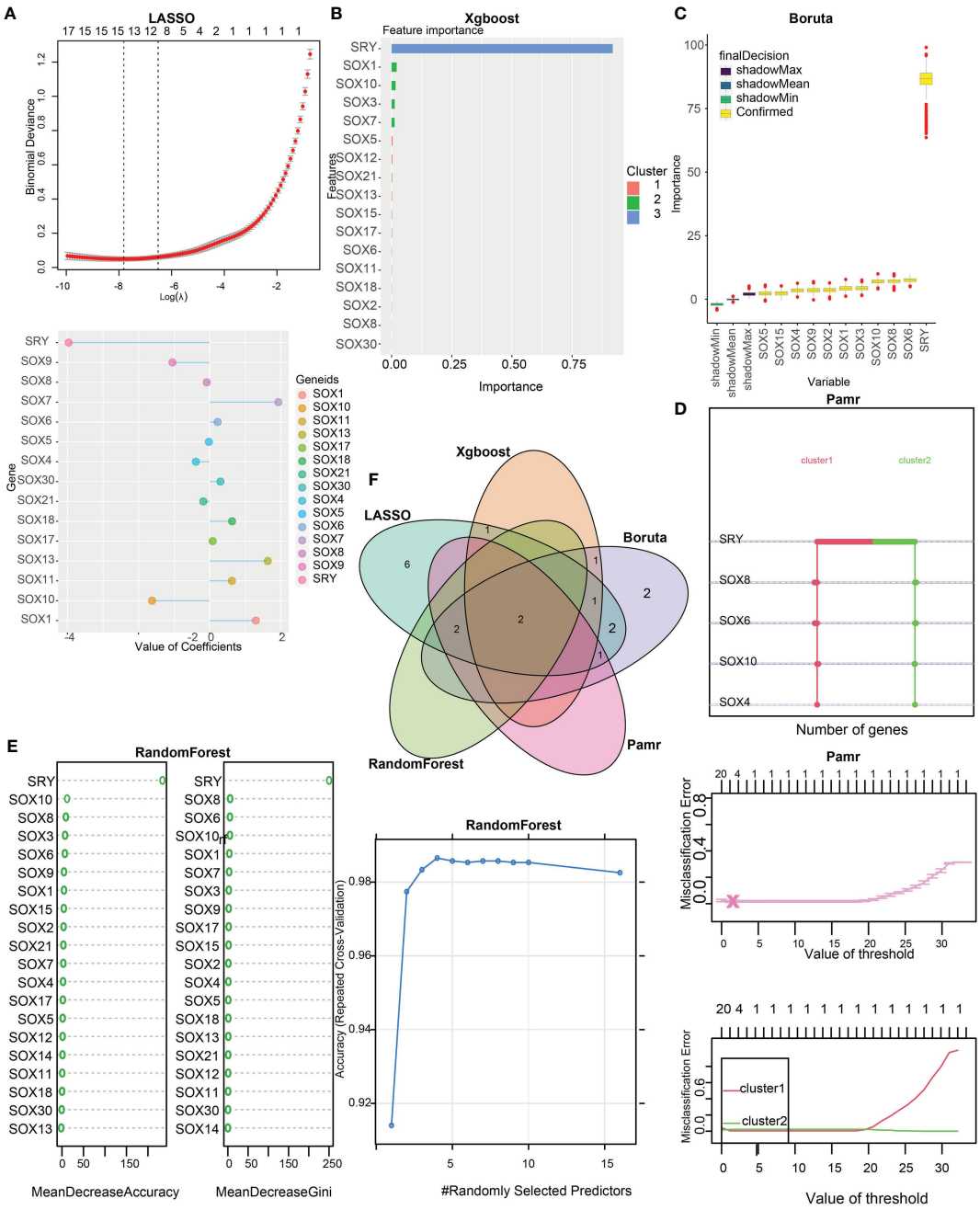


FIGURE 4
Identification of SOX10 as a prognostic gene by machine learning. **(A)** The assessment of the weighted importance of genes between two clusters by the LASSO-LR algorithm. **(B)** The evaluation of feature importance of genes between two clusters by the XGboost algorithm. **(C)** The selection of all relevant features of genes between two clusters by the Bruta algorithm. **(D)** The assessment of genes between two clusters by the Pamr algorithm. **(E)** The evaluation of genes between two clusters by the random forest algorithm. **(F)** Validation of the intersection of glioma prognostic genes from LASSO, Xgboost, Boruta, Pamr, and Random Forest.

The high expression of SOX10 seems to be correlated with the negative regulation of regulatory T-cell differentiation, DNA replication, and mismatch repair. Besides, Figure 5E demonstrates the abundance of infiltrating immune cell groups with divergent SOX10 expressions identified by the CIBERSORT, ESTIMATE, MCP, and TIMER algorithms of TCGA datasets. With the increasing expression of SOX10, the proportion of B cells, T cell CD4, and macrophages increased.

10.3389/fimmu.2022.1007461

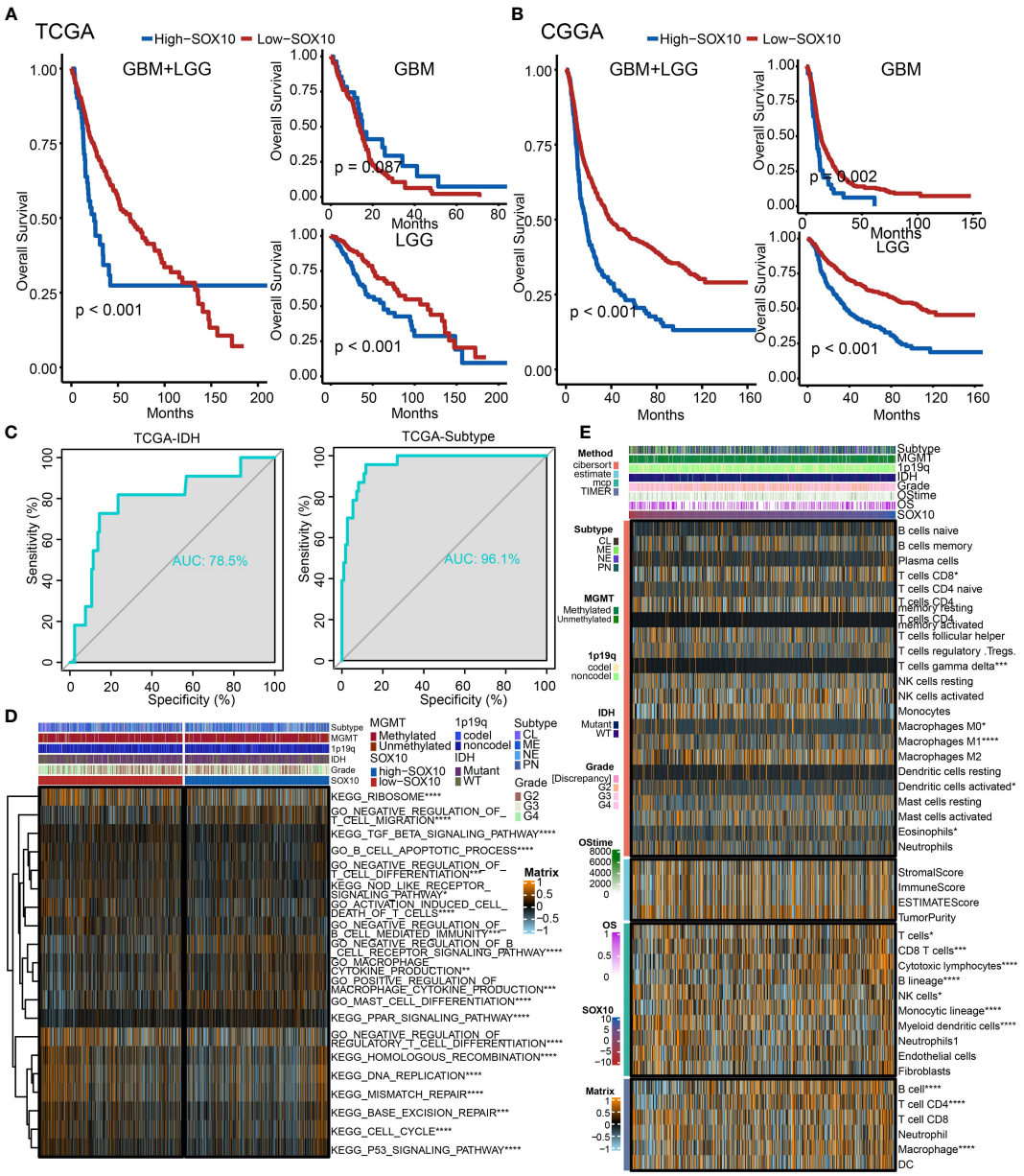


FIGURE 5
The prognostic potential of SOX10. Kaplan–Meier overall survival analysis of SOX10 in GBM, LGG, and pan-glioma based upon the **(A)** TCGA and **(B)** CGGA datasets. **(C)** Predictive value of SOX10, IDH, and subtype measured by ROC curves in TCGA dataset. **(D)** The heatmap for gene set variation analysis of SOX10 from TCGA. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. **(E)** Heatmap visualized the abundance of infiltrating immune cell groups with divergent SOX10 degree. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

SOX10 is associated with immunotherapy response

As manifested in the heatmap (Figure 6A), SOX10 was negatively correlated with T-cell dysfunction, implying its potential impact on immunotherapy. Specifically, SOX10 positively correlated with the normalized Z score from selection log2FC in the CRISPR screen dataset and normalized expression value from immune-suppressive cell types.

Metabolism has been considered a vital determining factor in the survivability and potency of immune cells (26). We explored correlations between SOX10 and enrichment scores of metabolism-pertinent pathways and cancer-immune cascade steps by GSVA. Figure 6B concludes that SOX10 was negatively

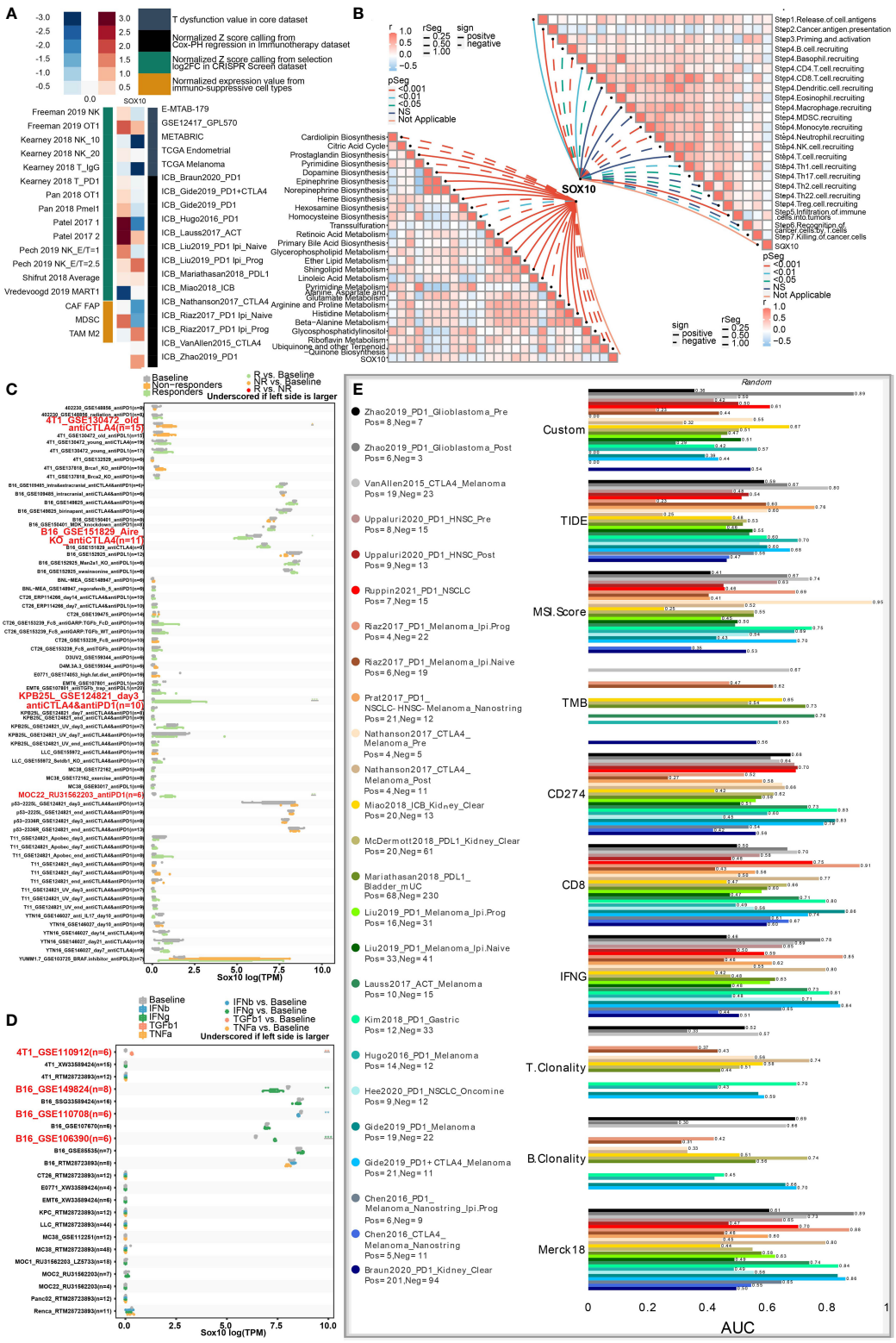


FIGURE 6
Roles of SOX10 in immunotherapy response, metabolism phenotypes, and biomarker relevance. **(A)** Heatmap showing the correlation with T-cell dysfunction, normalized Z score, and normalized expression. **(B)** Correlations between SOX10 and enrichment outcomes of metabolism-relevant pathways together with cancer-immune cascade steps. **(C)** Immunotherapy response of SOX10 in immunotherapy cohorts. **(D)** Immune effector molecule relevance of SOX10 in immunotherapy cohorts. **(E)** Biomarker relevance of SOX10 in immunotherapy cohorts.

associated with cardiolipin biosynthesis, citric acid cycle, transsulfuration, pyrimidine metabolism, ubiquinone, and another terpenoid. Notably, SOX10 was observed to be correlated with most steps of the immune cascade.

Then, to thoroughly analyze the prospective merit of SOX10 as a new immune target in pan-cancer, sensitive drugs predicated on SOX10 expression were predicted (Figure S1A). We also explored the semi-inhibition rates of gefitinib and nilotinib. The results showed that the estimated IC50 was higher in low-expression SOX10 than in high-expression (Figure S1B, C). Another noteworthy observation was that SOX10 could significantly predict immunotherapy response, whose responders were correlated with SOX10 levels (Figure 6C). Besides, SOX10 could significantly predict the cytokine treatment of immune effector molecules in four immunotherapy cohorts (Figure 6D). We also computed the biomarker pertinence of SOX10 by comparing it with normalized biomarkers based off of their prognosticative ability of response outcomes and OS of human immunotherapy cohorts. Fascinatingly, it was found that SOX10 gave an AUC of more than 0.5 in eight out of the 25 immunotherapy cohorts (Figure 6E). SOX10 presented a better predictive value than B clonality, with AUC values over 0.5 in 8 immunotherapy cohorts. However, the prognosticative ability of SOX10 was lower than that of the TIDE (AUC > 0.5 in 18 immunotherapy cohorts), MSI score (AUC > 0.5 in 13 immunotherapy cohorts), TMB (AUC > 0.5 in 8 immunotherapy cohorts), CD274 (AUC > 0.5 in 21 immunotherapy cohorts), CD8 (AUC > 0.5 in 18 immunotherapy cohorts), IFNG (AUC > 0.5 in 17 immunotherapy cohorts), T clonality (AUC > 0.5 in 9 immunotherapy cohorts), and Merk 18 (AUC > 0.5 in 18 immunotherapy cohorts).

Single-cell sequencing and SOX10 co-expression on glioma cells

Finally, we utilized single-cell sequencing to analyze the circumstances of stratification, identification, and SOX10 co-expression on glioma cells. UMAP determined by Copynumber Karyotyping of Tumors analysis stratified cells into diploid (average) status and aneuploid (malignant) status (Figure 7A). At the same time, we identified cell types and used UMAP to make it intuitionistic, which demonstrated 13 cell clusters (Figure 7B). Similarly, the co-expression status of different types of cells is shown in Figure 7C. Figure 7D shows the division of cell clusters into two groups, based upon the high and low expression levels of SOX10. In the high-expression cluster of SOX10, OPC was observed to take more than 50% of all TME cells, followed by neuroprogenitor cells (NPC), mesenchyme (MES), and astrocyte (AC) (Figure 7E). In descending order, the proportion of subtypes in a low-expression cluster of SOX10 was NPC (less than 50%), MES, AC, and OPC. The expression level of SOX10 in different subtypes is shown in the violin plot (Figure 7F). It was found that SOX10 had a high expression in neoplastic cells, astrocytes, neurons, oligodendrocytes and oligodendrocyte progenitor cells.

Subsequently, we performed enrichment analysis to determine the correlation between different immune regulatory processes and SOX10 expressions. High-expression SOX10 was significantly positively associated with the Notch signaling pathway and migration and activation regulation of the macrophage. In contrast, low-expression SOX10 was positively associated with the activation of MAPK activity, MAPK signaling pathway, and regulation of T-cell migration (Figure 7G). Moreover, we drew correlation circles for positively and negatively correlated checkpoint genes, growth factors, cytokines, and other genes in the SOX10 high-expression group (Figures 7H–K). For checkpoint genes, we could see a strong correlation in LGALSS of M1 macrophages and HAVCR2 of microglial cells, macrophages, M2 macrophages, and M1 macrophages themselves (Figure 7H). Also, HAVCR2 of M1 and M2 macrophages seemed to be regulated by LGALS9 of many cells, such as neoplastic cells, microglial cells, and neural stem cells. As for the growth factors, IGF2 of unknown cells showed robust correlations with IGF1R, IGF2R, and INSER of the other six cell types (Figure 7I). Macrophages may interact with microgial cells *via* CCL4, and interact with T cells and oligodendrocyte progentior cells *via* IL 1B (Figure 7J). The correlation with other genes could be found in Figure 7K by the same means.

Differences in cells neighboring SOX10-expressed cells

We performed multiplex immunofluorescence in the controlled group and different grades of glioma groups to further characterize the relationship between SOX10-expressed cells and neighboring CD68+CD163+ cells, and CD8+ cells. The results revealed that SOX10 expression is elevated with the increase in glioma grades (Figures 8A, B). Besides, with the increase in SOX10 expression, neighboring CD68+ cells, CD163+ cells, and CD8+ cells are also increased (Figures 8C, D). The quantity of CD8+ cells at the distance of 0–25 μm and 25–50 μm neighboring SOX10-expressed cells exploded in the Glioma WHO IV group, while the amount of CD68+CD163+ cells also increased. Hence, we concluded that CD68+CD163+ M2 macrophages, and CD8+ T cells, were the prepotent infiltrated immune cell types in glioma. Meanwhile, SOX10 expression is a regulator of neighboring immune cells.

Discussion

In recent years, studies revolving around gliomas are mostly concentrating on the TME, which is suggested to be one of the main obstacles to improving the prognosis and OS of HGG patients (27). To explore and clarify the mechanism of how

10.3389/fimmu.2022.1007461

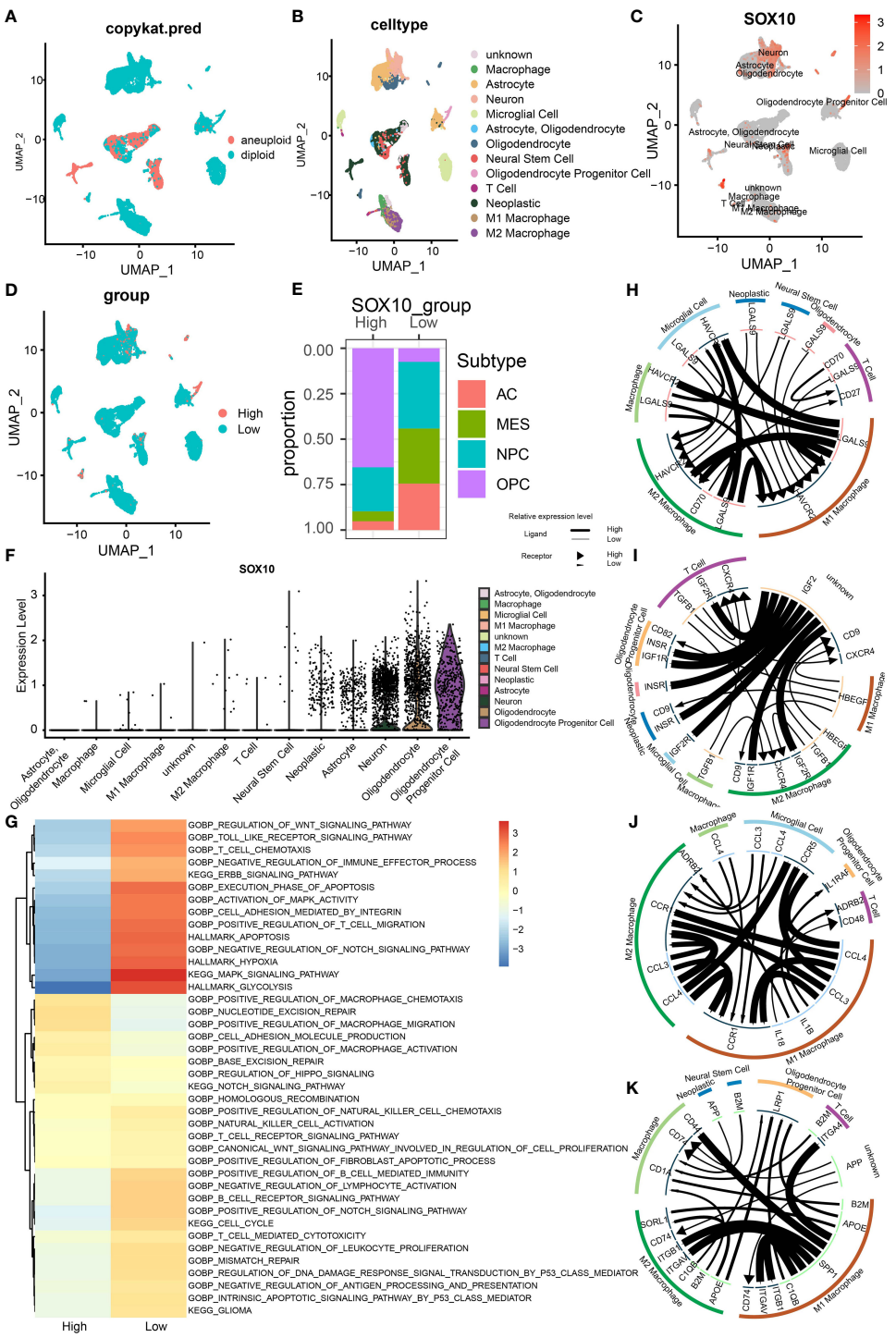


FIGURE 7
Stratification, identification, and SOX10 co-expression on glioma cells by single-cell sequencing analysis. **(A)** UMAP projection determined by CopyKat analysis. **(B)** UMAP projections of different cells, color-coded by cell types. **(C)** Annotation of different cell types and the expression of SOX10 in each cell type. **(D)** UMAP projections of two cell groups based on the expression of SOX10. **(E)** The proportion of glioma subtypes in the high and low SOX10 expression group. Astrocytes (AC), neural progenitor cells (NPC), mesenchyme (MES), and oligodendrocyte precursor cells (OPC). **(F)** Violin plot of SOX10 expression distribution of divergent cell clusters. **(G)** Enrichment analysis correlating divergent immune regulatory processes with high and low SOX10 expressions. **(H)** Correlation circles for positively and negatively correlated checkpoint genes in the high expression group of SOX10. **(I)** Correlation circles for positively and negatively correlated growth factors in the high expression group of SOX10. **(J)** Correlation circles for positively and negatively correlated cytokines in the high expression group of SOX10. **(K)** Correlation circles for positively and negatively correlated other genes in the high expression group of SOX10.

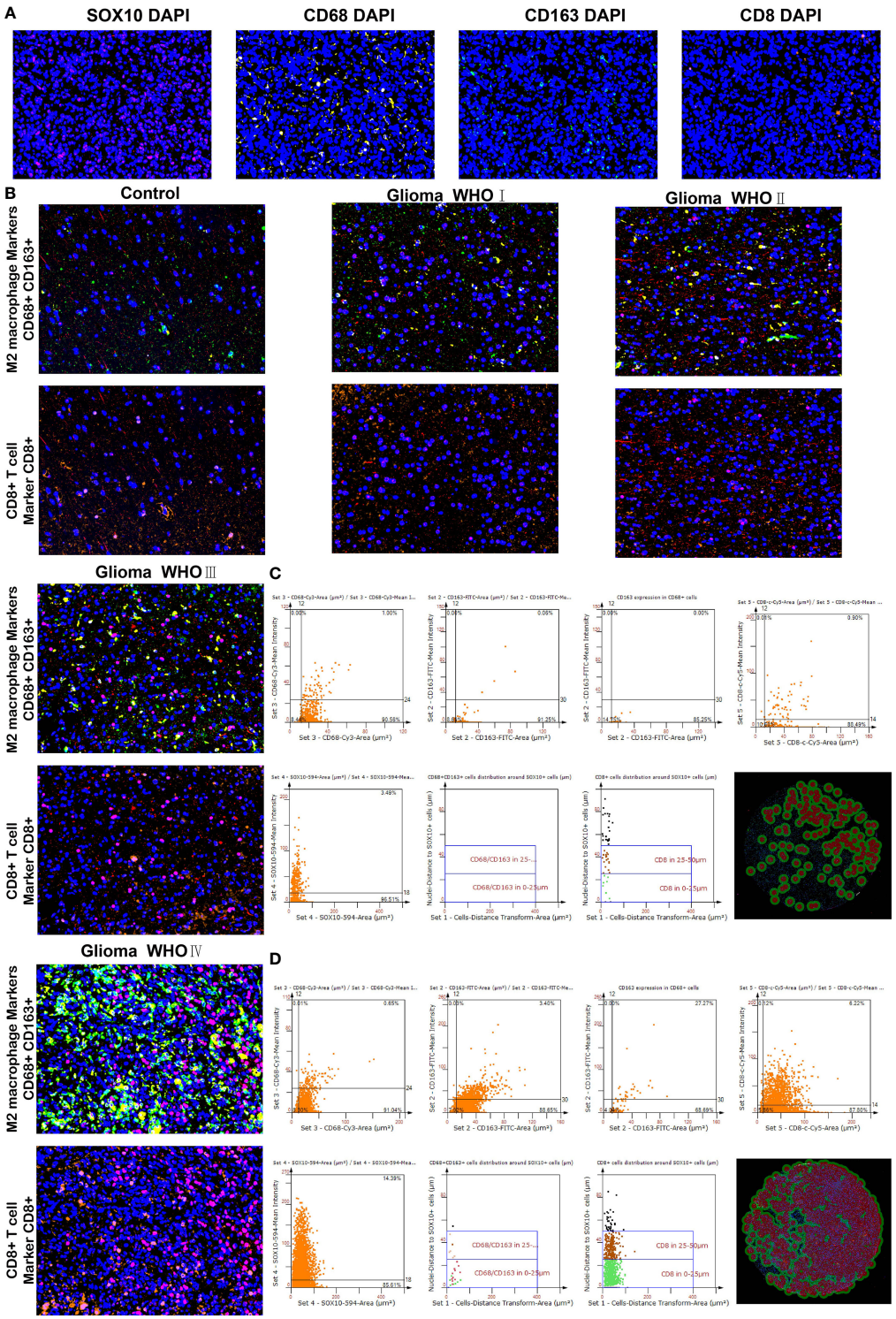


FIGURE 8
Differences in CD68, CD163 macrophages, and CD8 T cells neighboring SOX10-expressed cells. **(A)** Multiplex immunofluorescence staining of SOX10 (pink), CD68 macrophages (yellow), CD163 macrophages (green), CD8 T cells (orange), and DAPI (blue). **(B)** Multiplex immunofluorescence images of M2 macrophage markers CD68+ and CD163+, and CD8+ T cell marker CD8+ in control, Glioma WHO I, Glioma WHO II, Glioma WHO III, and Glioma WHO IV groups, respectively. The scatter diagrams display the quantity of CD68+ cells, CD163+ cells, CD68+CD163+ cells, CD8+ cells, and SOX10 expressed cells and the quantity of $CD68^{+}CD163^{+}$ cells along with CD8+ cells at different locations neighboring SOX10 expressed cells at 0–25 and 25–50 μm. Images of automatically identified staining by the TissueFAXS and StrataQuest software in **(C)** Glioma WHO III group and **(D)** Glioma WHO IV group.

infiltrating immune cells in the glioma TME influence the prognosis and OS, much large-scale bioinformatic analyses have been performed, and several biomarkers have been found (5, 28, 29). However, as a typical transcription factor family, the expression of the SOX family in gliomas has not been fully discussed before. We are the first to evaluate the cluster ability thoroughly and other characteristics of the SOX family and analyze SOX10 expression profiles in gliomas in prognostic potential, immune response, and co-expression in single-cell sequencing. Significantly, our results suggest that the SOX family has two distinct clusters regarding gliomas. Compared with cluster1, cluster2 seems to have a higher OS but weaker immune response. Correspondingly, the genomic profiles of the two clusters are dissimilar. IDH has co-occurrent relations with many other genes, such as ATRX, CIC, and FUBP1, which are potent regulators of cell growth (30). IDH1 is more likely to mutate in cluster2. More importantly, our analysis of SOX10 expression files in gliomas implies its predictive ability. Moreover, overexpression of SOX10 indicates a worse OS and prognosis. Also, SOX10 has the potential to predict immunotherapy response and immune effector molecules.

A complex range of genomic alterations also has clinical implications for glioma classification and prognosis. In SNV analysis, several frequent somatic mutations in gliomas, including IDH1, TP53, and ATRX, have been found to present more in cluster2 than in cluster1 (31, 32). Besides, as mentioned before, the TME has been reported to influence the gene expression of gliomas and the infiltration circumstance of stromal and immune cells, which are significant indicators of predicting prognosis (33). Data on the ESTIMATE algorithm show that stromal, immune, and ESTIMATE scores are higher in cluster1. The results indicate a worse prognosis and shorter OS in cluster1. Consequently, the SOX family is thought to have the cluster ability in gliomas to predict malignancy.

The SOX family has been confirmed to be closely associated with the immune features of the TME. In gliomas, the copious SOX family has played a crucial role in cell differentiation. Also, the SOX family and their mRNA expression levels have been associated with glioma patients' prognosis (13). In our study, the high expression of SOX10 is related to shorter OS in glioma.

An epigenome profiling of GBM indicates that SOX10, an oligodendrocyte forerunner marker and chromatin modifier, is a dominant regulator in RTK I-subtype tumors (34). It also affects the glioma TME. This is consistent with our results. Numerous types of immune cells are enriched in high-SOX10-expression patients in our analysis of infiltrating immune cell populations. Our results suggest that an increased expression of SOX10 is associated with the DNA replication, mismatch pair, and regulation of negative regulatory T-cell differentiation. With increasing SOX10 expression, B cells, T cell CD4, and macrophages are elevated. We can infer that SOX10 is correlated with T-cell dysfunction with the heatmap. As a consequence, we draw a conclusion that SOX10 is a significant regulator in the glioma TME.

Based on the types of function and activation, macrophages can be divided into two types: M1 macrophages and M2 macrophages (35). M1 macrophages are induced by LPS, INF-γ, and TNF-α and mainly release TNF-α, CXCL9, and CCL4. M2 macrophages mainly release TGF-β and CCL1 (36). Macrophages, especially M2 macrophages, are negatively associated with the survival of glioma patients (37). In our study, macrophages elevate in the SOX10 high expression group. Especially for CD68+CD163+ M2 macrophages neighboring SOX10-expressed cells, an increased number of these macrophages are found with the elevation of SOX10 expression. Besides, CD8+ T cells are also found to explode at the distance of 0–25- and 25–50-μm neighboring SOX10 high-expression cells. The results indicate that SOX10 regulates the types and quantity of glioma infiltrated immune cells.

GSVA shows a negative association with SOX10 and cardiolipin biosynthesis, citric acid cycle, trans-sulfuration, pyrimidine metabolism, and ubiquinone. SOX10 also has pleiotropic effects in cancer-immune cascade steps. Considering metabolism is a vital determining factor in the survivability and potency of immune cells, SOX10 is supposed to be a more remarkable biomarker in immunotherapy response than B clonality (26).

The SOX family has also been found to have the ability to regulate stem and progenitor cells in adult tissues (38). Our single-cell sequencing results reveal that SOX10 is highly expressed in OPC and NPC, indicating a regulatory function. An immune checkpoint, manifesting the capability of inhibiting T-cell function, refers to specified molecular interactions at the interface between antigen-presenting cells and T cells (39). In melanoma, regulated by fat mass and obesity-associated protein, enrichment of SOX10 decreases the effect on anti-PD-1 blockade immunotherapy (40). Similarly, our data imply that SOX10 can predict anti-PD1 and anti-CTLA4 immune therapy responses. Besides, we have found correlations between SOX10 and HAVCR2, LGALS9, and CD70. These results suggest a coordinated role with SOX10 and those immune checkpoints in glioma development.

Glioma invasion is driven by autocrine signaling transmitted by secretory factors that signal through receptors on the tumor, including growth factors and cytokines (41). We have found that IGF2R, INSR, and IGF1R have a tight relationship with SOX10 in gliomas. Besides, EGFR amplification and PTEN inactivation in GBM have recently been shown to regulate the activity of the DNA repair (42). Overexpression of EGFR drives GBM cell invasion. Gefitinib is a tyrosine kinase inhibitor targeting EGFR (43). The semi-inhibition rate demonstrates that the estimated IC50 is lower in the high-expression SOX10 group than in the counterpart, which suggests that high-expression SOX10 has higher sensitivity to gefitinib; in other words, gefitinib is more effective in gliomas overexpressing SOX10 (44). Our data indicate that gefitinib might be a molecularly targeted agent for treating patients with highly expressed SOX10.

Notwithstanding, the complete comprehensive information, specific functions, and clarified mechanisms of these SOX families in gliomas and many other diseases have not been fully explored and explained. It is reported that as an oncogene, more than 50% of the cancer patients present NOTCH activation mutations (45). The activation of NOTCH significantly favors tumor progression (46). It is accordant with our data. We have discovered through the enrichment analysis that the high expression of SOX10 has a positive correlation with the NOTCH signaling pathway. Tumor-associated macrophages have a complex interaction with glioma progression (47). In our study, positive regulation of macrophage chemotaxis and activation are also related to the high expression of SOX10, which may be the reason for the elevation of macrophages in the high expression SOX10 group. Therefore, it can be surmised that the overexpression of SOX10 may activate macrophages and then elevate the number of CD68+CD163+ macrophages, which are important components of the immune microenvironment. Then, macrophages release cytokines to regulate the signaling pathway, such as NOTCH, to affect glioma progression. Consequently, we infer that the overexpression of SOX10 can promote glioma progression.

In conclusion, our study demonstrates the outstanding cluster ability of the SOX family. Cluster2 has a better prognosis and longer OS than cluster1. Concentrating on SOX10, multiple results imply that it has a multifaceted prognostic value in gliomas. In gliomas, SOX10 overexpression corresponds to immune infiltration and bleak prognosis. However, Gefitinib and Nilotinib have more utility in patients with highly expressed SOX10. Except for PD1 and EGFR, our results suggest that the high expression of SOX10 may also correlate with other potential immune checkpoints. Given that, SOX10 has the potential to be an auspicious target for glioma immunotherapy.

Data availability statement

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

Author contributions

Conception and design: GX and QC; Foundation support: QC, GX, and RP; Acquisition and analysis of data: GX, KW, ZW, ZD, XL, WY, and ZL; Interpretation of data: PL and JZ; Drafting the manuscript and revising for submission quality: GX, KW, and QC; All authors; Study supervision: QC and RP. All authors contributed to the article and approved the submitted version.

Funding

Financial support was provided by the National Natural Science Foundation of China (Nos. 82171347, 82073893, and 81901268), Hunan Provincial Natural Science Foundation of China (Nos. 2022JJ30971 and 2022JJ20095), Hunan Provincial Health Committee Foundation of China (No. 202204044869), and Xiangya Hospital Central South University postdoctoral foundation.

Acknowledgments

The author express gratitude to the public databases, websites, and softwares used in the paper. We are grateful to the High Performance Computing Center of Central South University for partial support of this work.

Conflict of interest

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

The reviewer DW declared a shared parent affiliation with the authors PL and JZ to the handling editor at the time of the review.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1007461/full#supplementary-material

SUPPLEMENTARY FIGURE 1
(A) Correlation between SOX10 and various immunotherapy drugs. Semi-inhibition rate of different SOX10 expressions in **(B)** Gefitinib and **(C)** Nilotinib.

References

1. Ostrom QT, Patil N, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. Cbtrus statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2013-2017. *Neuro Oncol* (2020) 22(12 Suppl 2): iv1–iv96. doi: 10.1093/neuonc/noaa200

2. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol* (2016) 131(6):803–20. doi: 10.1007/s00401-016-1545-1

3. Aldape K, Brindle KM, Chesler L, Chopra R, Gajjar A, Gilbert MR, et al. Challenges to curing primary brain tumours. *Nat Rev Clin Oncol* (2019) 16(8):509–20. doi: 10.1038/s41571-019-0177-5

4. Klemm F, Maas RR, Bowman RL, Kornete M, Soukup K, Nassiri S, et al. Interrogation of the microenvironmental landscape in brain tumors reveals disease-specific alterations of immune cells. *Cell* (2020) 181(7):1643–60:e17. doi: 10.1016/j.cell.2020.05.007

5. Zhang H, He J, Dai Z, Wang Z, Liang X, He F, et al. Pdia5 is correlated with immune infiltration and predicts poor prognosis in gliomas. *Front Immunol* (2021) 12:628966. doi: 10.3389/fimmu.2021.628966

6. Xu S, Tang L, Liu Z, Luo C, Cheng Q. Hypoxia-related lncrna correlates with prognosis and immune microenvironment in lower-grade glioma. *Front Immunol* (2021) 12:731048. doi: 10.3389/fimmu.2021.731048

7. Kamachi Y, Kondoh H. Sox proteins: Regulators of cell fate specification and differentiation. *Development* (2013) 140(20):4129–44. doi: 10.1242/dev.091793

8. Grimm D, Bauer J, Wise P, Kruger M, Simonsen U, Wehland M, et al. The role of sox family members in solid tumours and metastasis. *Semin Cancer Biol* (2020) 67(Pt 1):122–53. doi: 10.1016/j.semcancer.2019.03.004

9. Garcia I, Aldaregia J, Marjanovic Vicentic J, Aldaz P, Moreno-Cugnon L, Torres-Bayona S, et al. Oncogenic activity of Sox1 in glioblastoma. *Sci Rep* (2017) 7:46575. doi: 10.1038/srep46575

10. Ge Y, Zhou F, Chen H, Cui C, Liu D, Li Q, et al. Sox2 is translationally activated by eukaryotic initiation factor 4e in human glioma-initiating cells. *Biochem Biophys Res Commun* (2010) 397(4):711–7. doi: 10.1016/j.bbrc.2010.06.015

11. Zhao T, Yang H, Tian Y, Xie Q, Lu Y, Wang Y, et al. Sox7 is associated with the suppression of human glioma by hmg-box dependent regulation of Wnt/Beta-catenin signaling. *Cancer Lett* (2016) 375(1):100–7. doi: 10.1016/j.canlet.2016.02.044

12. Glasgow SM, Zhu W, Stolt CC, Huang TW, Chen F, LoTurco JJ, et al. Mutual antagonism between Sox10 and nfia regulates diversification of glial lineages and glioma subtypes. *Nat Neurosci* (2014) 17(10):1322–9. doi: 10.1038/nn.3790

13. Stevanovic M, Kovacevic-Grujicic N, Mojsin M, Milivojevic M, Drakulic D. Sox transcription factors and glioma stem cells: Choosing between stemness and differentiation. *World J Stem Cells* (2021) 13(10):1417–45. doi: 10.4252/wjsc.v13.i10.1417

14. Hide T, Takezaki T, Nakatani Y, Nakamura H, Kuratsu J, Kondo T. Sox11 prevents tumorigenesis of glioma-initiating cells by inducing neuronal differentiation. *Cancer Res* (2009) 69(20):7953–9. doi: 10.1158/0008-5472.CAN-09-2006

15. Mou W, Xu Y, Ye Y, Chen S, Li X, Gong K, et al. Expression of Sox2 in breast cancer cells promotes the recruitment of M2 macrophages to tumor microenvironment. *Cancer Lett* (2015) 358(2):115–23. doi: 10.1016/j.canlet.2014.11.004

16. Liu Y, Guo W. Sox factors as cell-state regulators in the mammary gland and breast cancer. *Semin Cell Dev Biol* (2021) 114:126–33. doi: 10.1016/j.semcdb.2021.01.002

17. Malladi S, Macalinao DG, Jin X, He L, Basnet H, Zou Y, et al. Metastatic latency and immune evasion through autocrine inhibition of wnt. *Cell* (2016) 165 (1):45–60. doi: 10.1016/j.cell.2016.02.025

18. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallee VP, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat Med* (2020) 26(2):259–69. doi: 10.1038/s41591-019-0750-6

19. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, et al. Glioma groups based on 1p/19q, idh, and tert promoter mutations in tumors. *N Engl J Med* (2015) 372(26):2499–508. doi: 10.1056/NEJMoa1407279

20. Jiang T, Mao Y, Ma W, Mao Q, You Y, Yang X, et al. Cgcg clinical practice guidelines for the management of adult diffuse gliomas. *Cancer Lett* (2016) 375 (2):263–73. doi: 10.1016/j.canlet.2016.01.024

21. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* (2006) 9 (3):157–73. doi: 10.1016/j.ccr.2006.02.019

22. Cancer Genome Atlas Research N, Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* (2015) 372(26):2481–98. doi: 10.1056/NEJMoa1402121

23. Xie Y, Tan Y, Yang C, Zhang X, Xu C, Qiao X, et al. Omics-based integrated analysis identified atrx as a biomarker associated with glioma diagnosis and prognosis. *Cancer Biol Med* (2019) 16(4):784–96. doi: 10.20892/j.issn.2095-3941.2019.0143

24. Zhang S, Samocha KE, Rivas MA, Karczewski KJ, Daly E, Schmandt B, et al. Base-specific mutational intolerance near splice sites clarifies the role of nonessential splice nucleotides. *Genome Res* (2018) 28(7):968–74. doi: 10.1101/gr.231902.117

25. Racca JD, Chen YS, Maloy JD, Wickramasinghe N, Phillips NB, Weiss MA. Structure-function relationships in human testis-determining factor sry: An aromatic buttress underlies the specific DNA-bending surface of a high mobility group (Hmg) box. *J Biol Chem* (2014) 289(47):32410–29. doi: 10.1074/jbc.M114.597526

26. Pearce EL, Pearce EJ. Metabolic pathways in immune cell activation and quiescence. *Immunity* (2013) 38(4):633–43. doi: 10.1016/j.immuni.2013.04.005

27. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* (2017) 32(1):42–56:e6. doi: 10.1016/j.ccell.2017.06.003

28. Zhang H, Zhang N, Wu W, Zhou R, Li S, Wang Z, et al. Machine learning-based tumor-infiltrating immune cell-associated lncrnas for predicting prognosis and immunotherapy response in patients with glioblastoma. *Brief Bioinform* (2022). doi: 10.1093/bib/bbac386

29. Zhang N, Zhang H, Wu W, Zhou R, Li S, Wang Z, et al. Machine learning-based identification of tumor-infiltrating immune cell-associated lncrnas for improving outcomes and immunotherapy responses in patients with low-grade glioma. *Theranostics* (2022) 12(13):5931–48. doi: 10.7150/thno.74281

30. Jiao Y, Killela PJ, Reitman ZJ, Rasheed AB, Heaphy CM, de Wilde RF, et al. Frequent atrx, cic, Fubp1 and Idh1 mutations refine the classification of malignant gliomas. *Oncotarget* (2012) 3(7):709–22. doi: 10.18632/oncotarget.588

31. Segura-Collar B, Gargini R, Tovar-Ambel E, Hernandez-SanMiguel E, Epifano C, Perez de Castro I, et al. The egfr-Tmem167a-P53 axis defines the aggressiveness of gliomas. *Cancers (Basel)* (2020) 12(1):208. doi: 10.3390/cancers12010208

32. Koschmann C, Calinescu AA, Nunez FJ, Mackay A, Fazal-Salom J, Thomas D, et al. Atrx loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Sci Transl Med* (2016) 8(328):328ra28. doi: 10.1126/scitranslmed.aac8228

33. Huang S, Song Z, Zhang T, He X, Huang K, Zhang Q, et al. Identification of immune cell infiltration and immune-related genes in the tumor microenvironment of glioblastomas. *Front Immunol* (2020) 11:585034. doi: 10.3389/fimmu.2020.585034

34. Wu Y, Fletcher M, Gu Z, Wang Q, Costa B, Bertoni A, et al. Glioblastoma epigenome profiling identifies Sox10 as a master regulator of molecular tumour subtype. *Nat Commun* (2020) 11(1):6434. doi: 10.1038/s41467-020-20225-w

35. Mills CD, Kincaid K, Alt JM, Heilman MJ. Hill AM. m-1/M-2 macrophages and the Th1/Th2 paradigm. *J Immunol* (2000) 164(12):6166–73. doi: 10.4049/jimmunol.164.12.6166

36. Huang X, Li Y, Fu M, Xin HB. Polarizing macrophages in vitro. *Methods Mol Biol* (2018) 1784:119–26. doi: 10.1007/978-1-4939-7837-3_12

37. Zhang H, Luo YB, Wu W, Zhang L, Wang Z, Dai Z, et al. The molecular feature of macrophages in tumor immune microenvironment of glioma patients. *Comput Struct Biotechnol J* (2021) 19:4603–18. doi: 10.1016/j.csbj.2021.08.019

38. Sarkar A, Hochedlinger K. The sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* (2013) 12(1):15–30. doi: 10.1016/j.stem.2012.12.007

39. Wang Y, Zhang H, Liu C, Wang Z, Wu W, Zhang N, et al. Immune checkpoint modulators in cancer immunotherapy: recent advances and emerging concepts. *J Hematol Oncol* (2022) 15(1):111. doi: 10.1186/s13045-022-01325-0

40. Yang S, Wei J, Cui YH, Park G, Shah P, Deng Y, et al. M(6)a mrna demethylase fto regulates melanoma tumorigenicity and response to anti-Pd-1 blockade. *Nat Commun* (2019) 10(1):2782. doi: 10.1038/s41467-019-10669-0

41. Hoelzinger DB, Demuth T, Berens ME. Autocrine factors that sustain glioma invasion and paracrine biology in the brain microenvironment. *J Natl Cancer Inst* (2007) 99(21):1583–93. doi: 10.1093/jnci/djm187

42. Squatrito M, Holland EC. DNA Damage response and growth factor signaling pathways in gliomagenesis and therapeutic resistance. *Cancer Res* (2011) 71(18):5945–9. doi: 10.1158/0008-5472.CAN-11-1245

43. Blandin AF, Cruz Da Silva E, Mercier MC, Glushonkov O, Didier P, Dedieu S, et al. Gefitinib induces egfr and Alpha5beta1 integrin Co-endocytosis in glioblastoma cells. *Cell Mol Life Sci* (2021) 78(6):2949–62. doi: 10.1007/s00018-020-03686-6

44. Meco D, Servidei T, Riccardi A, Ferlini C, Cusano G, Zannoni GF, et al. Antitumor effect in medulloblastoma cells by gefitinib: Ectopic Her2 overexpression enhances gefitinib effects in vivo. *Neuro Oncol* (2009) 11(3):250–9. doi: 10.1215/15228517-2008-095

45. Parmigiani E, Taylor V, Giachino C. Oncogenic and tumor-suppressive functions of notch signaling in glioma. *Cells* (2020) 9(10):2304. doi: 10.3390/cells9102304

46. Radtke F, Fasnacht N, Macdonald HR. Notch signaling in the immune system. *Immunity* (2010) 32(1):14–27. doi: 10.1016/j.immuni.2010.01.004

47. Hambardzumyan D, Gutmann DH, Kettenmann H. The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci* (2016) 19 (1):20–7. doi: 10.1038/nn.4185

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 05 December 2022
DOI 10.3389/fonc.2022.1045477

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Dan Zhang,
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS), China
Lingnan He,
Tongji University, China

*CORRESPONDENCE
Hongwei Yang
18831099969@163.com

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

RECEIVED 15 September 2022
ACCEPTED 09 November 2022
PUBLISHED 05 December 2022

CITATION
Zhao G, Wang C, Jiao J, Zhang W and Yang H (2022) The novel subclusters based on cancer-associated fibroblast for pancreatic adenocarcinoma. *Front. Oncol.* 12:1045477. doi: 10.3389/fonc.2022.1045477

The novel subclusters based on cancer-associated fibroblast for pancreatic adenocarcinoma

Guojie Zhao[1], Changjing Wang[2], Jian Jiao[1], Wei Zhang[1] and Hongwei Yang[3]*

[1]The Seventh Department of General Surgery, HanDan Central Hospital, Handan, Hebei, China, [2]The Department of Gastrointestinal surgery, The Third Hospital of Hebei Medical University, Shijiazhuang, Hebei, China, [3]The First Department of Oncology, HanDan Central Hospital, Handan, Hebei, China

Introduction: Pancreatic adenocarcinoma (PAAD) is a fatal disease characterized by promoting connective tissue proliferation in the stroma. Activated cancer-associated fibroblasts (CAFs) play a key role in fibrogenesis in PAAD. CAF-based tumor typing of PAAD has not been explored.

Methods: We extracted single-cell sequence transcriptomic data from GSE154778 and CRA001160 datasets from Gene Expression Omnibus or Tumor Immune Single-cell Hub to collect CAFs in PAAD. On the basis of Seurat packages and new algorithms in machine learning, CAF-related subtypes and their top genes for PAAD were analyzed and visualized. We used CellChat package to perform cell–cell communication analysis. In addition, we carried out functional enrichment analysis based on clusterProfiler package. Finally, we explored the prognostic and immunotherapeutic value of these CAF-related subtypes for PAAD.

Results: CAFs were divided into five new subclusters (CAF-C0, CAF-C1, CAF-C2, CAF-C3, and CAF-C4) based on their marker genes. The five CAF subclusters exhibited distinct signaling patterns, immune status, metabolism features, and enrichment pathways and validated in the pan-cancer datasets. In addition, we found that both CAF-C2 and CAF-C4 subgroups were negatively correlated with prognosis. With their top genes of each subclusters, the sub-CAF2 had significantly relations to immunotherapy response in the patients with pan-cancer and immunotherapy.

Discussion: We explored the heterogeneity of five subclusters based on CAF in signaling patterns, immune status, metabolism features, enrichment pathways, and prognosis for PAAD.

KEYWORDS

pancreatic adenocarcinoma, immune features, machining learning, prognosis, immunotherapy, subclusters

Introduction

Pancreatic adenocarcinoma (PAAD) is a serious threat to people's life and health due to its high degree of malignancy and poor prognosis. According to the latest epidemiological data, pancreatic cancer is the 12th most common tumor in the world but the fourth most deadly cancer worldwide (1, 2). Pancreatic ductal adenocarcinoma (PDAC) is the most common histologic type of PAAD. PDAC has low resection rate, insensitive radio chemotherapy, and poor prognosis, and the 5-year survival rate is less than 7% (1, 3). PAAD develops gradually from genetic abnormality to abnormal cell proliferation and precancerous lesions and then to minimal early carcinoma, which takes a very long time, about 5–20 years. However, it only takes 6 to 20 months to develop from a small tumor to a significant mass and then to the advanced stage. In addition, because of the painless and insidious growth of pancreatic masses, most patients with pancreatic cancer are already diagnosed in advanced stages. Therefore, the study of the pathogenesis and progression of PAAD and the search for suitable bimolecular targets are of great significance to enrich the treatment strategies of pancreatic cancer and improve the prognosis of patients.

The occurrence and development of tumors are closely related to their living environment, and the internal environment composed of tumor cells, mesenchymal cells, immune cells, vascular endothelial cells, and extracellular matrix (ECM) is called tumor microenvironment (TME) (4). During the development of PAAD, a microenvironment is formed, which is favorable for the survival, proliferation, and distant metastasis of PAAD cells (5). The poor prognosis of pancreatic cancer may be associated with specific biological characteristics, such as significant interstitial fibrosis (6). In recent years, researchers have paid more and more attention to the stroma of PAAD (7, 8). Dense fibrous tissue surrounding tumor cells is an important histologic feature of PDAC (9–11). The main components of interstitium include ECM, immune cells, endothelial cells, and cancer-associated fibroblast (CAF) (12), and stroma microenvironment cells interact with tumor cells in a complex way (13). TME can determine the biological behavior of the tumor, which, in turn, affects patient prognosis. Therefore, understanding the biological characteristics of TME is crucial for understanding the biological behavior of PAAD (14).

Tumor stroma cells are complex, and interstitial cells interact with each other (7, 15). The relatively abundant cell components in the stroma are CAFs, and CAFs are closely related to the significant proliferation of connective tissue of PAAD cells. CAFs are considered to be fibroblasts that produce ECM, cytokines, chemokines, and growth factors, with the primary function of promoting tumor progression (16). However, some targeted therapy studies on CAFs suggested that removal of CAFs can promote tumor progression or metastasis (17, 18), suggesting significant heterogeneity of CAFs within tumors (19), that is, some CAF subgroups may play a role in inhibiting tumor progression. A large number of single-cell transcriptome sequencing studies have further clarified the significant heterogeneity of CAFs within and between tumors, as well as the functional classification of CAFs (20, 21). Currently, commonly accepted cancer-associated fibroblast (CAF) are categorized as myofibroblastic CAFs (myoCAFs) and inflammatory CAFs (iCAFs). myoCAFs are mainly distributed around tumor cells and are mainly related to the generation of ECM. Some reports suggested that some subgroups of myoCAFs may be involved in immune regulation (22). iCAFs mainly secrete cytokines and chemokines to act on tumor cells. In addition, other small CAF subsets, such as apCAF (20) and LRRC15 (+) myoCAFs (21), were identified. Of course, different subsets of cells perform different functions, and as single-cell sequencing technology continues to mature, more functional subsets of CAFs may be discovered. At present, single-cell sequencing studies suggested that representative markers of myoCAFs were Alpha-smooth muscle actin (α-SMA), periostin, and matrix metallopeptidase-11 (MMP-11); representative markers of iCAFs were Interleukin-6 (IL-6), C-X-C Motif Chemokine Ligand 12 (CXCL12) stromal cell-derived factor-1 (SDF-1), and Platelet-derived growth factor receptors-beta (PDGFR-β); and fibroblast activation protein-alpha (FAP-α) was a co-expression marker of two types of CAFs (20). However, the origin, function, and biological characteristics of CAFs need to be further studied.

In the current study, we gained single-cell sequence transcriptomic data from public databases. We carried out comprehensive analysis to generate five CAF subclusters and explore the differences among them. This will provide new insights into the treatment of PAAD.

Abbreviations: PAAD, pancreatic adenocarcinoma; CAF, cancer-associated fibroblast; PDAC, pancreatic ductal adenocarcinoma; TME, tumor microenvironment; ECM, extracellular matrix; GEO, Gene Expression Omnibus; TISCH, Tumor Immune Single-cell Hub; TCGA, The Cancer Genome Atlas; ICB, immune checkpoint blockade; CCA, canonical correlation analysis; DEG, differentially expressed gene; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; GSVA, gene set variation analysis; K-M, Kaplan–Meier; OS, overall survival; RFS, relapse-free survival; ROC, receiver operating characteristic.

Materials and methods

Study design and data collection

The flowchart of present study is shown in Figure 1. Single-cell sequence transcriptomic data from the GSE154778 and CRA001160 datasets were collected to analyze the fibroblast cells (23, 24). Full details can be downloaded from Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo) and Tumor Immune

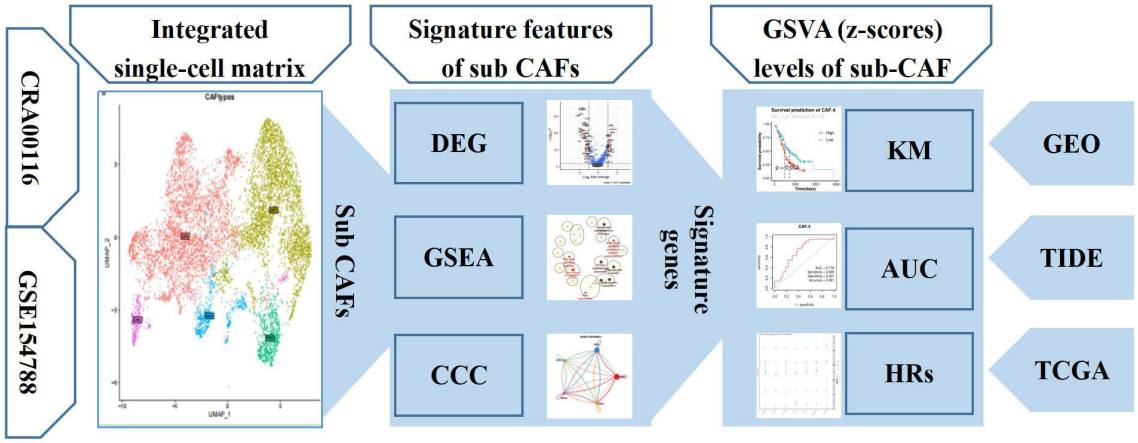


FIGURE 1
The workflow diagram depicting collection of data and processing of the analysis to show the framework of our study.

Single-cell Hub (TISCH; http://tisch.comp-genomics.org/) databases (25). Among them, we extracted CAF cells to analyze the features. In addition, seven bulk-sequence data for PDAC—TCGA (n = 146), ICGC-PACA-AU (n = 267), GSE71729 (n = 125), GSE62452 (n = 66), GSE57495 (n = 63), ICGC-PACA-CA (n = 182), and E_MTAB_6134 (n = 50)—were enrolled from GEO and The Cancer Genome Atlas (TCGA) databases based on previous data (26). Pan-cancer dataset with 31 cancer types was also collected to verify the features of the single-cell subsets. All data generated or analyzed during this study are freely available in the previous publications. Last, to get the immune features of the subset of single-cells, 10 cohorts with different tumors before or after immunotherapy [immune checkpoint blockade (ICB)] were collected in Tumor Immune Dysfunction and Exclusion (TIDE) database to further analysis (27).

Subset for fibroblast cells

The Seurat R package was used to visualize the CAF cells from two cohorts (28), including the 1,656 CAFs in GSE154778 and 6,228 CAFs in CRA001160. Top 2,000 genes were integrated by the method of canonical correlation analysis (CCA) to integrate CAFs for dimensionality reduction cluster analysis (Resolution = 0.1, N = 5) (29). We also performed ScaleData, RunPCA, DimPlot, and t-SNE (t-distributed stochastic neighbor embedding) based on R packages to analyze and visualize results.

Cell–cell communication analysis

The CellChat R package with full of ligand–receptor interactions can analyzed the intercellular communication networks between different cell clusters in the single-cell dataset (30). To access the major signaling inputs and outputs among subsets and other TME cells, the CellChatDB.human, netVisual_circle and netVisual_bubble functions were used to show the strength or weakness of cell–cell communication networks from the CAF subclusters to other different cell clusters in single-cell dataset.

Identification of the marker genes of CAF cell subtypes

FindAllMarkers function was used to list the markers of subclusters of CAF (31). The min.pct and logfc.threshold functions were all set as 0.25. The AddModuleScore function could obtain the signature scores based on differentially expressed genes (DEGs) (32). The dot plot function was performed to show the top highest gene expressions in subcluster (33). The FeaturePlot function was used to show the distribution of specific signatures of subcluster scores. The volcano plot based on the marker genes among different subsets of CAF was displayed.

Functional enrichment analysis for CAF subsets

The significant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) functions were detected by the clusterProfiler R package (34) based on marker genes among different subsets of CAF. To cluster the special pathways, the Cytoscape enrichment map function was performed in the Cytoscape software (35).

Prognosis analysis and prediction analysis of subsets

We first performed the gene set variation analysis (GSVA) (36) based on the subset signatures of CAF subsets to get the enrichment scores for these subclusters of CAF in the PDAC bulk sequence. On the basis of their prognostic information, we analyzed the prognosis features of subsets of CAF enrichment score in the cohorts from TCGA and GEO. The cutoff values of different NMF cell signatures in the different public datasets were determined by the survminer R package (37) used to plot Kaplan–Meier (K-M) curves. The prediction value of subsets of CAF for immunotherapy also was performed by the receiver operating characteristic (ROC) analysis. The ComplexHeatmap (38) or pheatmap (39) packages in R visualize the pooled values of CAF in these cohorts.

Gene expression detecting using quantitative real-time PCR assays

The human pancreatic CAF-stellate cell named CAF118 was supplied by Neuromics (Edina, USA) and was cultured using Stem Cell Complete Low Serum Media (Neuromics, Edina, USA). The human pancreatic cell HPC-Y5 was purchased from National Collection of Authenticated Cell Cultures and was cultured in 90% MEM Eagles with Earle's Balanced Salts (EME-EBSS) with 10% FBS (fetal bovine serum). The human pancreatic cancer cell line SW1990 was purchased from Procell (Wuhan, China) and cultured in 90% LEIBOVITZ (L-15) with 10% FBS. After extracting the total RNA of the cell lines by the RNAsimple Total RNA Kit (Tiangen, China), we reverse-transcribed RNA to acquire cDNA using the PrimeScript RT reagent Kit (Takara, Otsu, Japan). Finally, on the basis of the premixed system of 2 μl of cDNA with SYBR Premix Ex Taq (Takara, Otsu, Japan) and primers, we detected the expression values of related genes in cell lines by an Applied Biosystems StepOne Plus Real-Time PCR system (Life Technologies, Grand Island, NY, USA). The primers of the target gene were supplied by Sangon Biotech (Shanghai, China). The sequences of the primers used are listed in Table 1.

TABLE 1 The primer sequences in PCR analysis.

Symbol	Sequences (5′-3′)
ADM-F	CTGATGTACCTGGGTTCGCT
ADM-R	ATGTCCTGGGGCCGAATAAG
Eno2-F	CTCTGTGGTGGAGCAAGAGA
Eno2-R	ATTGATCACGTTGAAGGCCG
ERO1A-F	TTGGATCTGCTGGTGGTCAT
ERO1A-R	TCCCTTGACCAGAAGCCAAA
BNIP3-F	CGCAGACACCACAAGATACC
BNIP3-R	GCGCTTCGGGTGTTTAAAGA
UPP1-F	TTGACTGCCCAGGTAGAGAC
UPP1-R	TGCCTGCTCTGTTATGACCA
Actin-F	ACTTCGAGCAAGAGATGGCC
Actin-R	GCTGATCCACATCTGCTGGA

Statistical analysis

Routine statistical analyses of the present study were performed in R 4.0 software. The relationships of sub-CAF with other special genes were calculated by the Spearman's rank correlation. The K-M method, log-rank test, and Cox regression analysis were performed to detect the prognosis of subset of CAF in the OS (overall survival) and RFS (relapse-free survival) in patients with Pancreatic ductal adenocarcinoma (PDAC) and other tumor. The area under the ROC curve was used to estimate the diagnostic value of GSVA score of subset of CAF. A two-sided p-value below 0.05 was considered statistically significant.

Results

Identification of five CAF-related subtypes for PDAC

Recent SCNA-SEQ studies of human PDAC have shown that intra-tumor heterogeneity of PDAC is key to the analysis of tumor-related mechanisms. Extensive fibrous proliferation caused by CAFs is common in PDAC. In clinical practice, we often encounter PDAC tumors with unique histological characteristics. To characterize the CAF subpopulations in PDAC, we performed unsupervised clustering analysis (Figure 2A). The all-positive expressed markers (log2FC > 1) are shown in Figure 2B and Supplementary Table S1. The results showed that CAFs were divided into five subclusters based on their marker genes (Figure 2C): CAF-C0 (by marker genes C7 and PTGDS), CAF-C1 (by marker genes COL11A1 and COL10A1), and CAF-C2 (by marker genes EPB41L4A-AS1 and ENO2). Proportions of sub-CAF in each patient is different, and that verifies the features for the single-cell subsets (Figure 2D).

CAF subclusters exhibited distinct signaling patterns

The major signaling inputs and outputs among subclusters were different. The characteristics of signaling patterns within each CAF subgroup were different. The results showed that

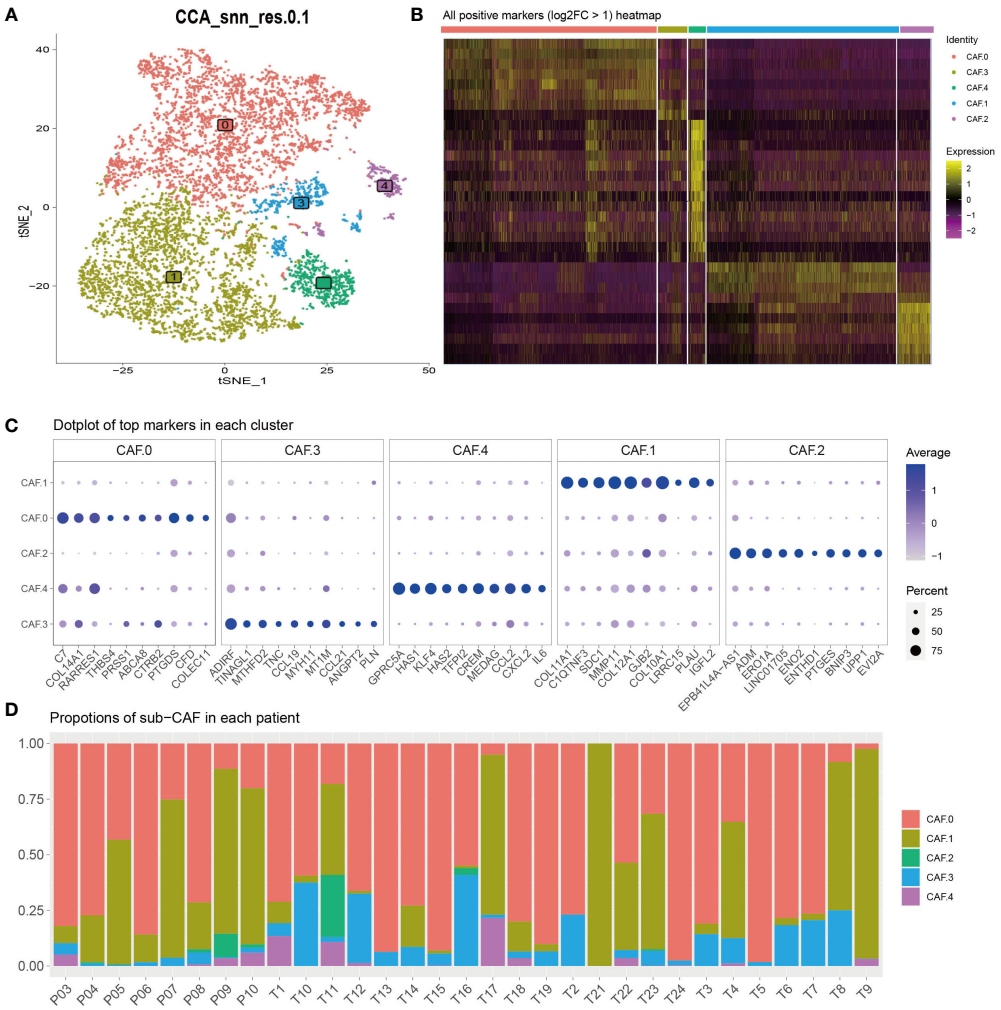


FIGURE 2
(A) To characterize the CAF subpopulations in PDAC, we performed unsupervised clustering analysis and showed that CAFs, which were categorized into five subclusters (C0, C1, C2, C3, and C4). **(B)** All-positive markers (log2FC > 1) heatmap of CAF cell subtypes. The colors of the top bar represent the different subclusters. Yellow indicates higher expression, and purple indicates lower expression. **(C)** Dot plot of top 10 markers in each cluster. The color represents the average expression. The size of the circle represents the percent. **(D)** Proportions of sub-CAF in each patient. The axis represents the ratio of different subclusters for each patient. The colors of the bar represent the different subclusters.

subcluster CAF-C0 was related to CD99, MK, PDGF, NEGR, NCAM, BMP, and CD46; CAF-C1 was related to FN1, CD99, MK, PDGF, NEGR, NCAM, BMP, and CD; CAF-C2 was related to TGB2; whereas CAF-C4 was related to ITGB2 (Figure 3A). The cross-linking between CAF and 14 kinds of main TME cells in each subcluster was also different (Figures 3B, C). CAF-C0 was closely related to adenocyte, epithelial-to-mesenchymal transition (EMT), endocrine, epithelial, and malignant, whereas CAF-C1 was closely related to adenocyte, EMT, endocrine, epithelial, and malignant. CAF-C2 was closely related to adenocyte, EMT, endocrine, epithelial, and malignant, whereas CAF-C3 was closely related to adenocyte, EMT, endocrine, epithelial, and malignant. CAF-C4 was associated with adenocyte, EMT, endocrine, epithelial, malignant, endothelial, and plasma.

CAF subclusters exhibited distinct immune and metabolism features

Some subsets based on CAFs were significantly correlated with immune gene sets, such as CAF-C2 and immune modulators, other cytokines, C3 and co-inhibitors, immune checkpoint, MHC class I, and C4 and MHC non-class (Figures 4A–D). The expression of metabolism-related genes of CAF in each subclusters was different (Figure 4E). The genes

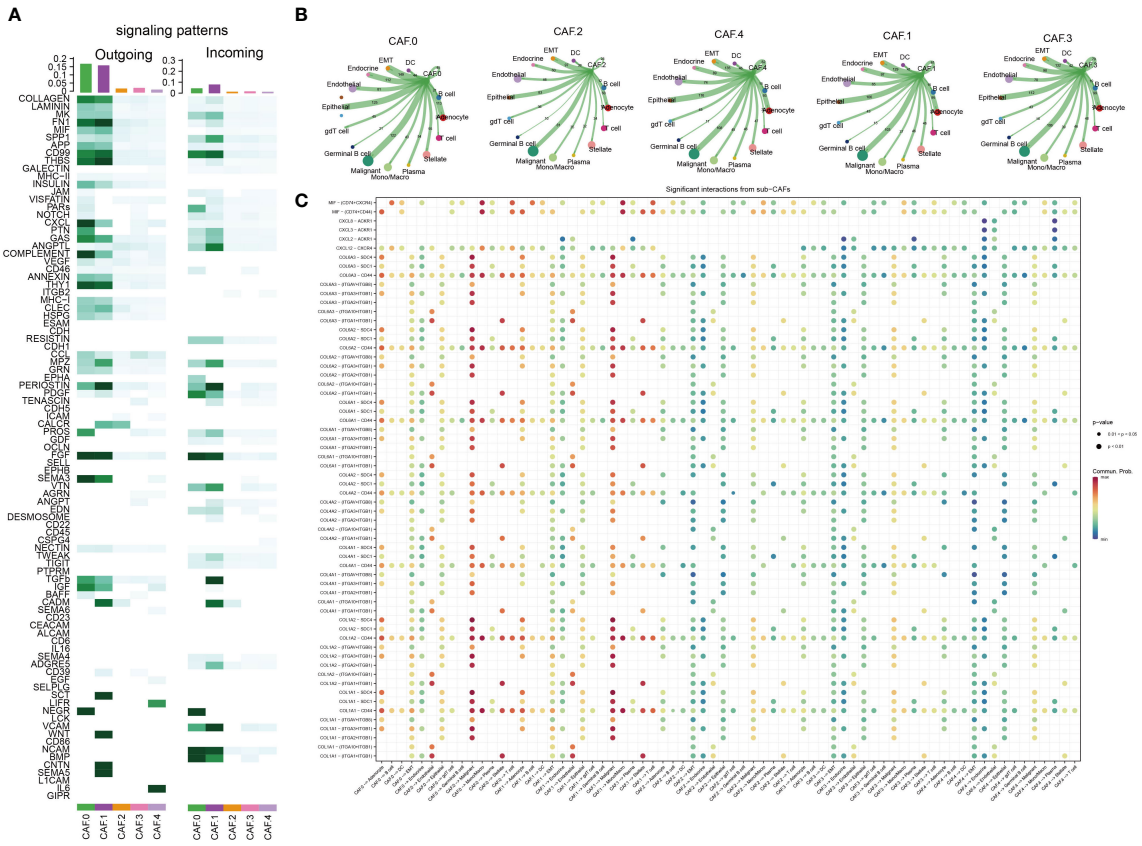


FIGURE 3
Cell–cell communication analysis. **(A)** The major signaling inputs and outputs among subsets. **(B)** Detailed view of the ligand–receptor expressed by each cell type and the other cell types for each CAF subclusters. The thicker the lines, the greater the number/intensity of ligand receptor. The size of the dots represents the number of cells in the subpopulation. **(C)** Bubble plot showing the ligand–receptor interactions between CAF clusters and cells. P-values are indicated by circle size. Communication proportion is indicated by color. The redder the color, the more important the interaction.

related to arachidonic acid metabolism (HSD11B1, PDK4, ALDH1A1, GPX3, PTGDS, GGT5, RBP1, PNLIP, CYP1B1, ADH1B, PTGIS, and INMT), arginine and proline metabolism (PLA2G1B, AMY2A, PLA2G2A, ALDH2, MGST1, PLPP3, CDO1, FMO1, and LTC4S), Cyclooxygenase arachidonic acid metabolism (FMO3 and DHRS3), and drug metabolism by cytochrome P450 (GLUL, LAP3, ALDH1A3, STRA6, CHST1, and CH25H) were highly expressed in CAF-C0. The genes related to purine metabolism (HSD17B6, ALDH1B1, PLOD1, ALOX15B, and PYCR1) and pyrimidine metabolism (ENPP1 and SCD) were highly expressed in CAF-C1. Genes associated with sugar synthesis and metabolism, such as N-glycan biosynthesis (ENO1), oxidative phosphorylation (HMOX1), primary bile acid biosynthesis (PKM), retinol metabolism (ENO2, PTGES, UPP1, and CA12), starch and sucrose metabolism (PSAT1), and steroid hormone biosynthesis (PHGDH, GSTA1, and CA9), were highly expressed in CAF-C2. The genes related to glycerolipid metabolism (CMPK2, TYMS, and AKR1C1), sphingolipid metabolism (GK), taurine and hypotaurine metabolism (NDUFA4L2), and testosterone biosynthesis (MGLL) were highly expressed in CAF-C3. The metabolic genes related to lipid and amino acid [such as ether lipid metabolism (GAPDH and GDA); fatty acid degradation (CP and VNN2); gluconeogenesis (ENPP2, TPI1, NAMPT, CA2, and ST6GALNAC5); glycine, serine, and threonine metabolism (SAT1 and UAP1); glycosaminoglycan biosynthesis (RDH10 and CRABP2); hexosamine biosynthesis (PTGS2), lysine degradation (B4GALT1 and NME1); and nicotinate and nicotinamide metabolism (ODC1, ANXA1, HSPA5, and SRM)] were highly expressed in CAF-C4.

CAF subclusters exhibited distinct enrichment pathways

GO and KEGG analysis suggested differences in their biological functions of the five subclusters (Figures 5A, B) and Supplementary Table S2. Interestingly, all five subtypes were

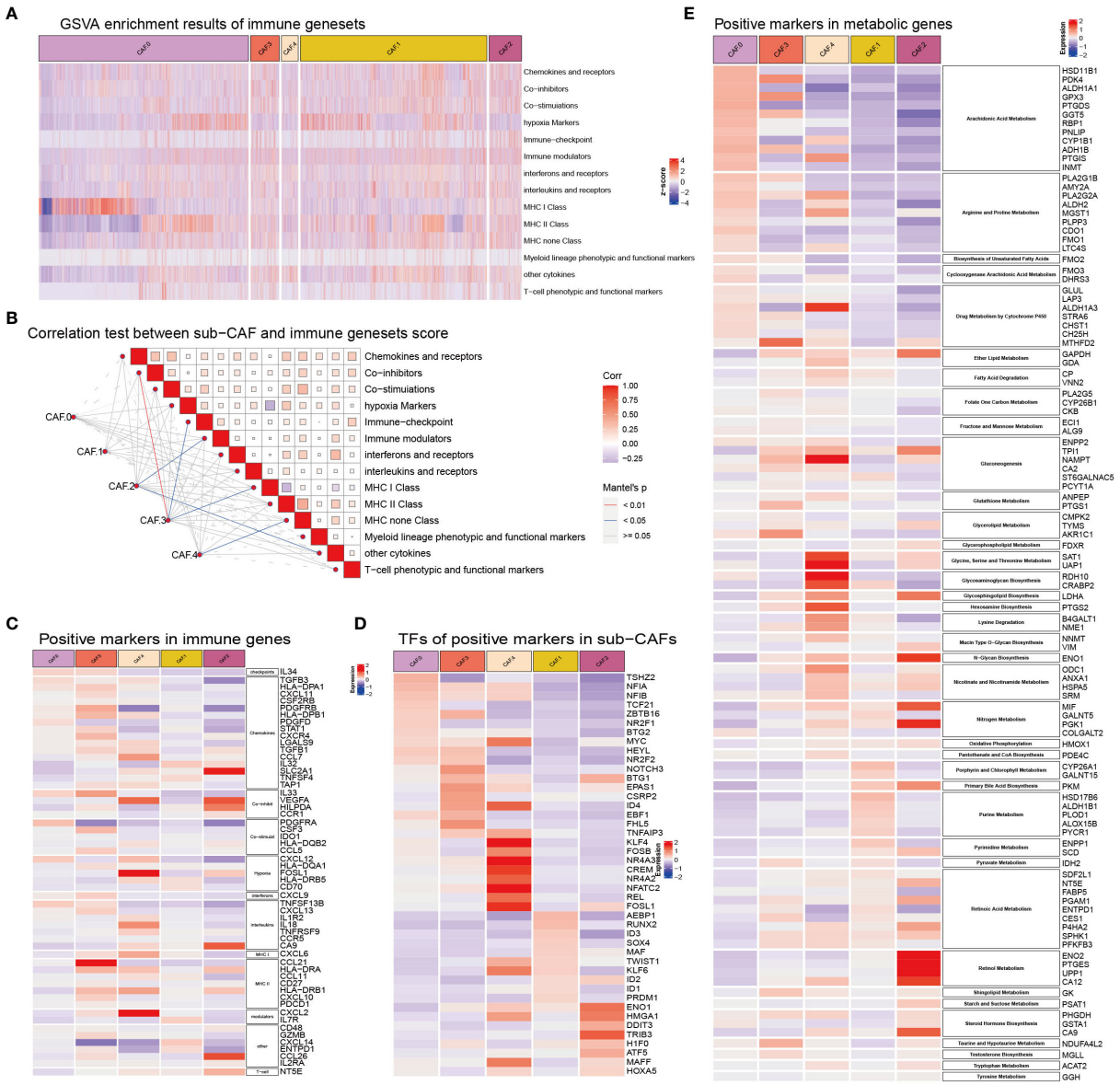


FIGURE 4
(A) GSVA enrichment results of immune gene sets in sub-CAFs. The z-score represents the rating. **(B)** Correlation test between sub-CAF and immune gene sets score. The color of the lines indicates the Mantel's p-value. The color of the box represents the correlation. **(C)** Positive markers in immune genes in sub-CAFs. The redder the color, the higher the expression of the gene in the CAF cluster. The bluer the color, the lower the expression of the gene in the CAF cluster. **(D)** TFs of positive markers in sub-CAFs. The bluer the color, the lower the expression of the gene in the CAF cluster. **(E)** Positive markers in metabolic genes in sub-CAFs. The bluer the color, the lower the expression of the gene in the CAF cluster.

enriched in four pathways: complement and coagulation cascades, ECM–receptor interaction, proteoglycans in cancer, and AGE-RAGE signaling pathway in diabetic complications (Figure 5A). As for CAF-0, there were highly expressed genes involved in T cell activation, ATP generation from ADP, tumor necrosis factor production, vasoconstriction, cellular response to ketone, biosynthesis of amino acids, and so on. For CAF-C1, there were highly expressed genes involved in regulation of peptide secretion, positive regulation of apoptotic signaling pathway, negative regulation of cell morphogenesis involved in differentiation, Wnt signaling pathway, and signaling pathways regulating pluripotency of stem cells. For CAF-C2, there were highly expressed genes involved in cellular response to extracellular stimulus, neutrophil activation involved in immune response, negative regulation of cell activation, HIF-1 signaling pathway, and arachidonic acid metabolism. For CAF-

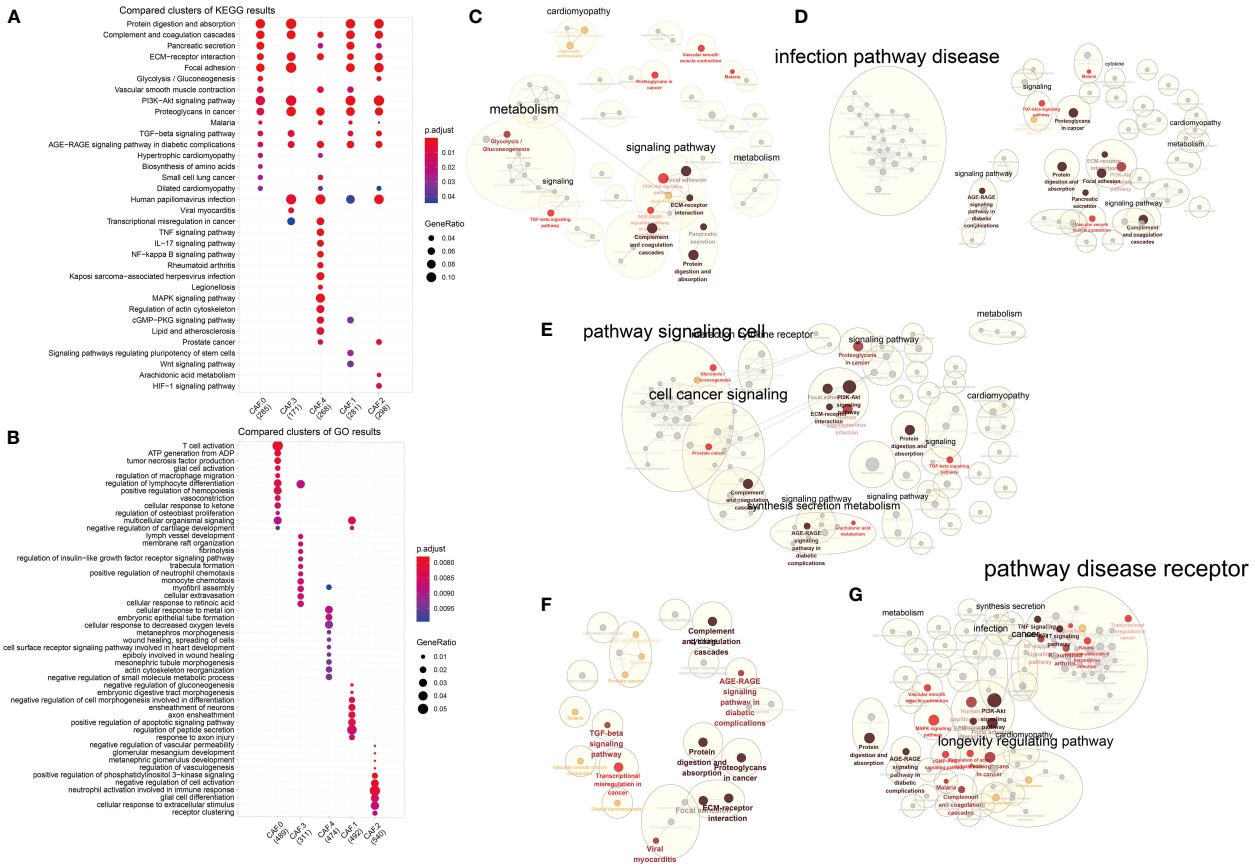


FIGURE 5
Functional enrichment analysis for CAF subsets. **(A)** Compared clusters of KEGG results. The color represents the P-value, and the size of the circle represents the ratio of genes. **(B)** Compared clusters of GO results. The color represents the P-value, and the size of the circle represents the ratio of genes. **(C–G)** Networks of functional enrichment analysis elaborated by Cytoscape.

C3, there were highly expressed genes involved in monocyte chemotaxis and regulation of insulin-like growth factor receptor signaling pathway. For CAF-C4, there were highly expressed genes involved in cellular response to decreased oxygen levels, cellular response to metal ion, negative regulation of small molecule metabolic process, mitogen-activated protein kinase (MAPK) signaling pathway, tumor necrosis factor (TNF) signaling pathway, and IL-17 signaling pathway. We established networks to elaborate how related genes were functionally enriched (Figures 5C–G).

Survival analysis of different CAF subclusters

Volcanic maps of differential genes for comparison of CAF between two groups were shown in Figure 6A. We can see the distribution of hazard ratios (HRs) based on sub-CAFs for tumors in TCGA database from Figure 6B. For ACC (adrenocortical carcinoma), GBM (glioblastoma multiforme), LGG (brain lower-grade glioma), LUSC (lung squamous cell carcinoma), and UVM (uveal melanoma), HRs predicted by sub-CAFs were all statistically significant. We collected PAAD data from seven databases and analyzed the correlation between CAF subgroup marker genes and patient prognosis. We found that both C2 and C4 subgroups were negatively correlated with patient survival (Figures 6C–E).

CAF subclusters exhibited distinct immunotherapy

To get the immune features of the subset of single-cells, 10 cohorts with different tumors before or after immunotherapy (ICB) were collected in TIDE database to further analysis. The results showed that each sub-CAF had different levels of immunotherapy response (Figure 7). The expression of some ICP gene HAVCR2 was positively correlated with the GSVA z-

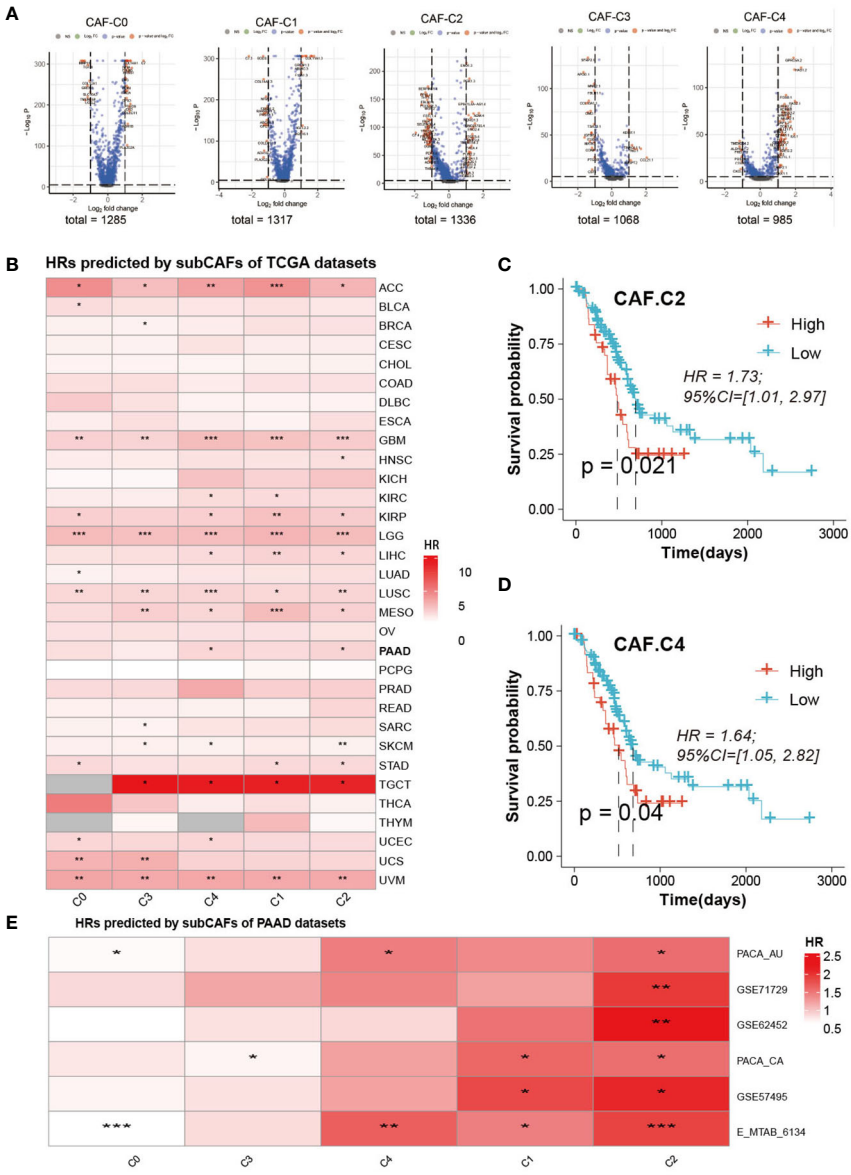


FIGURE 6
Prognosis analysis and Prediction analysis of subsets. **(A)** The volcano figure of differential expression analysis of five CAF-clusters. **(B)** HRs predicted by subCAFs of TCCA datasets. The color represents the HR value. *P < 0.05, **P < 0.01, ***P < 0.001. **(C)** Survival curve predicted for CAF-C2. **(D)** Survival curve predicted for CAF-C4. **(E)** HRs predicted by subCAFs of PAAD datasets. The color represents the HR value. *P < 0.05; **P < 0.01; ***P < 0.001.

score of these CAF subclusters (Figure 7A). We calculated the cell subset score of each sample in the immune therapy dataset for five CAF clusters and analyzed the correlation with prognosis by univariate cox analysis (Figure 7B). From Figure 7B, Nathanson2017_CTLA4 was found to have prognostic value in the four CAF clusters (CAF-C0, CAF-C1, CAF-C2, and CAF-C3). Therefore, we selected the Nathanson2017_CTLA4 immunotherapy dataset for CAF-C2 scoring, divided into high and low groups, and drew the K-M curve, from which we observe the poor prognosis of the low group (Figure 7C). We also developed a diagnostic model for immunotherapy response, as shown in Figure 7D.

Quantitative real-time PCR

We selected the marker genes (ADM, ERO1A, ENO2, BNIP3, and UPP1) of CAF-C2 to detect their expression in

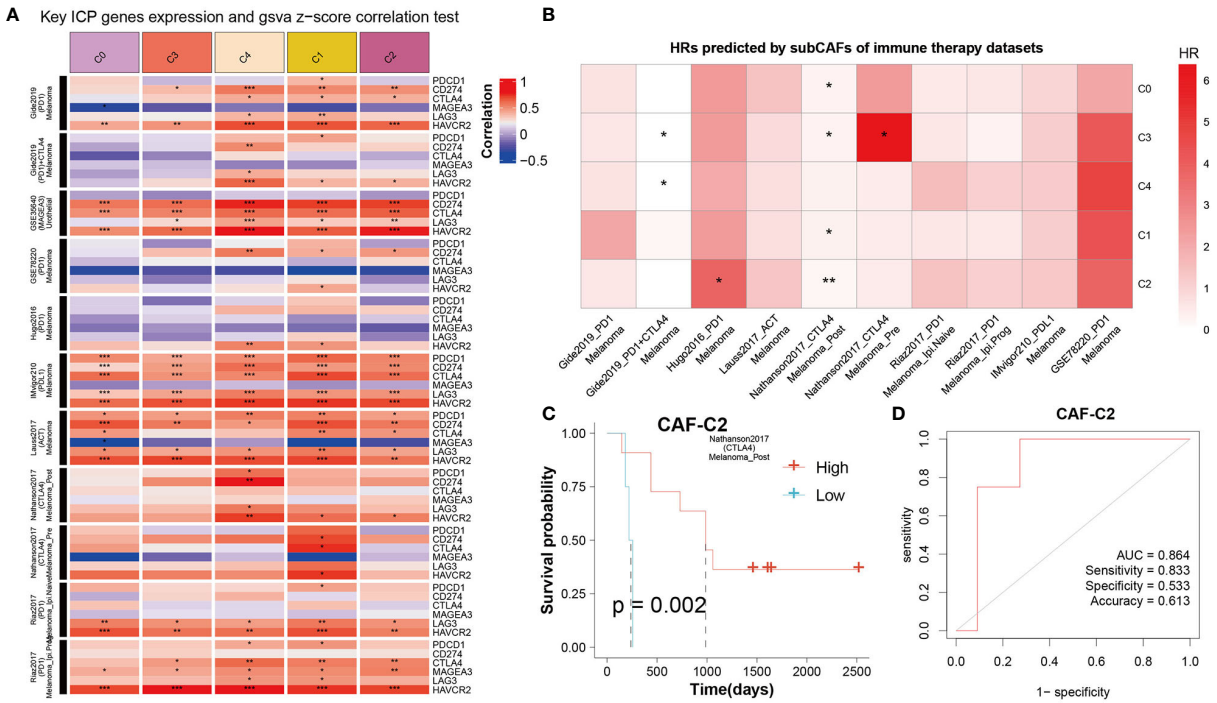


FIGURE 7
The correlation between CAF clusters and immune therapy. **(A)** Key ICP genes expression and GSVA z-score correlation test. Red means positive correlation, and blue means negative correlation. *P < 0.05, **P < 0.01, and ***P < 0.001. **(B)** HRs predicted by sub-CAFs of immune therapy datasets. The color represents the HR value. *P < 0.05 and **P < 0.01. **(C)** K-M curve for Nathanson2017_CTLA4 immunotherapy dataset based on CAF-C2 scoring. **(D)** The diagnostic model for immunotherapy response based on CAF-C2.

human pancreatic CAF-stellate cell (CAF118), human pancreatic cell (HPC-Y5), and human pancreatic cancer cell line (SW1990). Compared with HPC-Y5, ADM (Figure 8A), ERO1A (Figure 8B), ENO2 (Figure 8C), BNIP3 (Figure 8D), and UPP1 (Figure 8E) were significantly higher expressed in SW1990 and CAF118.

Discussion

PAAD is an aggressive malignancy, of which 95% are PDAC. In recent years, its morbidity and mortality rates have increased by an average of 0.3% per year due to changes in lifestyle and factors such as aging population and increased life expectancy

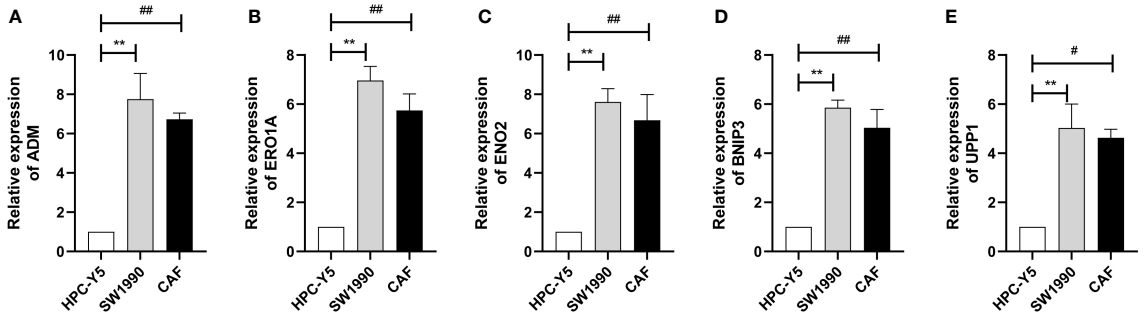


FIGURE 8
Quantitative real-time PCR. **(A–E)** Quantitative real-time PCR assays using cell lines for ADM **(A)**, ERO1A **(B)**, ENO2 **(C)**, BNIP3 **(D)**, and UPP1 **(E)**. **P < 0.01; #P < 0.05; ##P < 0.01.

(40). Because of the lack of specific symptoms and biological markers, early diagnosis of PAAD is very difficult. PAAD progresses rapidly and is inoperable by the time most patients are diagnosed (41). At the same time, pancreatic cancer is not sensitive to most treatments (42), so its prognosis is extremely poor (43). Some studies have shown that the unique interaction network and high heterogeneity of pancreatic cancer cells and that their microenvironment may play an important role in the origin, progression, and drug resistance of pancreatic cancer cells, and elucidating the inherent complex mechanisms has completed the common goal of scholars in this field (44, 45). Despite a lot of work, the results have been poor, with PAAD showing the lowest improvement in 5-year survival in recent years compared with other cancers (46). One of the important reasons lies in the limitations of traditional research methods in exploring the heterogeneity of tumors. Single-cell sequencing technology brings hope to break through this dilemma. It can deeply analyze the distribution, status, and interaction of different subgroups of cells, which makes up for some shortcomings of traditional sequencing technology and provides a new research method. At present, some studies based on single-cell sequencing technology have gradually achieved results and gradually gained a new understanding of the occurrence and progress of PAAD, providing possible targets for early diagnosis and effective treatment and promoting the development of precision medicine in the field of PAAD.

Molecular subtyping of PAAD is still in its early stage. In the clinical evaluation and prognostic analysis of PAAD, TNM stage and other clinical features are commonly used. However, because of individual differences in pancreatic cancer, there is no widely used molecular classification of pancreatic cancer that is associated with prognosis or has different sensitivity to treatment (47). Therefore, it is necessary to develop better methods for clinical diagnosis and prognosis assessment of PAAD so that patients can early detect cancer and take reasonable and effective treatment measures.

In recent years, with the continuous optimization and progress of the second-generation sequencing technology, the study of tumor bioinformatics has developed rapidly (48). However, there are a lot of mesenchymal components in PAAD tissue, which often leads to direct sequencing or inaccurate sequencing results (49, 50). Genomics studies have revealed common genomic pathway changes in PAAD, as well as more common or targeted somatic mutations in addition to the four major driver genes (51–53). Waddell et al. found that unstable patients may be more suitable for treatment with drugs involved in genomic damage repair pathways, such as Poly ADP-ribose polymerase (PARP) inhibitors or platinum, compared with other three types (53). More studies on PAAD typing have focused on transcriptomics. Because of the high content and complex composition of mesenchyma in PAAD tissues, there are some differences in sequencing analysis results. Sequencing analysis of samples with high or enriched tumor cells showed that PAAD tended to be divided into two types: classical and basal-like (54, 55). Studies of samples with relatively low levels of tumor cells, however, showed that pancreatic cancer types tended to be more diverse (55–57). This may be mainly due to the complexity of interstitial components, such as the differences in immune cell infiltration and interstitial activation. Canonical and basal-like transcriptome types are of great significance in predicting the prognosis of patients, but they have not yet played a good role in the classification and guidance of specific clinical treatment.

To further illuminate the subtyping of PAAD based on CAFs, we used Seurat 1656 CAFs in GSE154778 and 6228 CAFs in CRA001160. Our data discriminated five CAF subclusters and corresponding marker genes. To explore the mechanism of these CAF subclusters involved in the development of PAAD, we assess the characteristics of signaling patterns for the five CAF subclusters and found that these CAF subclusters were all closely related to EMT and endocrine. RHIM et al. traced that PAAD cells could develop EMT and obtain mesenchymal phenotype through *in vivo* pedigree, some cells after EMT initiated stem cell program, and PAAD cells with $CD24^{+}CD44^{+}$ stem cell phenotype were more likely to enter the blood circulation and survive (58). Breast cancer cells can also exhibit fibroblast characteristics and have the ability to differentiate into myofibroblasts (59). Our results further suggested that CAFs may be derived from EMT. We found that the expression of metabolism-related genes of CAF in each subclusters was different. Metabolic changes are an important feature in the identification of cancer cells. Many studies have found that CAFs are associated with energy metabolism of cancer cells, and tumor cells can better adapt to their rapid growth by modifying the TME. Sun et al. found that hypoxia can improve the glycolysis activity of CAFs, and lactic acid in hypoxia CAFs, as a metabolic coupling between CAFs and breast cancer cells, can improve the mitochondrial activity of cancer cells through relevant signaling pathways, thus promoting the invasion of breast cancer cells (60). In addition, in autophagy-related paracrine mode, CAFs provide substrates (such as lactic acid, pyruvate, and ketone bodies) for adjacent cancer cells derived from their own excess glycolysis activity (61). Research has shown that, in breast cancer, prostate cancer, head and neck carcinoma and lymphoma, and tumor, the catabolism of fibroblasts, the anabolic metabolism coupling between cancer cells, and the metabolic coupling drive fibroblasts of oxidative stress, glycolysis, autophagy, and aging; the decomposition in the metabolic production of fibroblasts for tumor growth provides a rich nutrition of microenvironment. The formation of mitochondrial fuel (lactic acid, ketone bodies, fatty acids, glutamine, and other amino acids) through a local matrix promotes tumor growth (62). CAFs can play an important role in the progression of cancer cells through a variety of metabolic pathways, which may provide new strategies for the treatment of PAAD.

In summary, we evaluated the heterogeneity of subclusters based on CAF for PAAD. The signaling patterns, immune status, metabolism features, and enrichment pathways of these subclusters were estimated and determined. Nonetheless, some limitations of

the current study should not be ignored. The number of cells from the databases obtained in this study is limited, which varies from patient to patient. Therefore, more sample size is needed to support the conclusion. In addition, further high-throughput single-cell sequencing analysis and *in vivo* studies should be used to confirm the conclusions of this study.

Conclusions

We explored the heterogeneity of five subclusters based on CAF in signaling patterns, immune status, metabolism features, enrichment pathways, and prognosis for PAAD.

Data availability statement

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

Author contributions

HY conceived, designed, and supervised the study. GZ performed data analysis. CW, JJ and WZ arranged the figures and drafted the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to the data from The Cancer Genome Atlas.

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/fonc.2022.1045477/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654

2. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: Good or bad news from the 2018 global cancer statistics? *Cancer Commun (Lond)* (2019) 39:22. doi: 10.1186/s40880-019-0368-6

3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338

4. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer* (2006) 6:392–401. doi: 10.1038/nrc1877

5. Farrow B, Albo D, Berger DH. The role of the tumor microenvironment in the progression of pancreatic cancer. *J Surg Res* (2008) 149:319–28. doi: 10.1016/j.jss.2007.12.757

6. Di Maggio F, El-Shakankery KH. Desmoplasia and biophysics in pancreatic ductal adenocarcinoma: Can we learn from breast cancer? *Pancreas* (2020) 49:313–25. doi: 10.1097/mpa.0000000000001504

7. Zhan HX, Zhou B, Cheng YG, Xu JW, Wang L, Zhang GY, et al. Crosstalk between stromal cells and cancer cells in pancreatic cancer: New insights into stromal biology. *Cancer Lett* (2017) 392:83–93. doi: 10.1016/j.canlet.2017.01.041

8. Ligorio M, Sil S, Malagon-Lopez J, Nieman LT, Misale S, Di Pilato M, et al. Stromal microenvironment shapes the intratumoral architecture of pancreatic cancer. *Cell* (2019) 178:160–175.e127. doi: 10.1016/j.cell.2019.05.012

9. Erkan M, Hausmann S, Michalski CW, Fingerle AA, Dobritz M, Kleeff J, et al. The role of stroma in pancreatic cancer: Diagnostic and therapeutic implications. *Nat Rev Gastroenterol Hepatol* (2012) 9:454–67. doi: 10.1038/nrgastro.2012.115

10. Neesse A, Algül H, Tuveson DA, Gress TM. Stromal biology and therapy in pancreatic cancer: A changing paradigm. *Gut* (2015) 64:1476–84. doi: 10.1136/gutjnl-2015-309304

11. Rucki AA, Zheng L. Pancreatic cancer stroma: Understanding biology leads to new therapeutic strategies. *World J Gastroenterol* (2014) 20:2237–46. doi: 10.3748/wjg.v20.i9.2237

12. Neesse A, Michl P, Frese KK, Feig C, Cook N, Jacobetz MA, et al. Stromal biology and therapy in pancreatic cancer. *Gut* (2011) 60:861–8. doi: 10.1136/gut.2010.226092

13. Thomas D, Radhakrishnan P. Tumor-stromal crosstalk in pancreatic cancer and tissue fibrosis. *Mol Cancer* (2019) 18:14. doi: 10.1186/s12943-018-0927-5

14. Feig C, Gopinathan A, Neesse A, Chan DS, Cook N, Tuveson DA. The pancreas cancer microenvironment. *Clin Cancer Res* (2012) 18:4266–76. doi: 10.1158/1078-0432.Ccr-11-3114

15. Kane S, Engelhart A, Guadagno J, Jones A, Usoro I, Brutcher E. Pancreatic ductal adenocarcinoma: Characteristics of tumor microenvironment and barriers to treatment. *J Adv Pract Oncol* (2020) 11:693–8. doi: 10.6004/jadpro.2020.11.7.4

16. Räsänen K, Vaheri A. Activation of fibroblasts in cancer stroma. *Exp Cell Res* (2010) 316:2713–22. doi: 10.1016/j.yexcr.2010.04.032

17. Özdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu CC, Simpson TR, et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell* (2014) 25:719–34. doi: 10.1016/j.ccr.2014.04.005

18. Rhim AD, Oberstein PE, Thomas DH, Mirek ET, Palermo CF, Sastra SA, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* (2014) 25:735–47. doi: 10.1016/j.ccr.2014.04.021

19. Neuzillet C, Tijeras-Raballand A, Ragulan C, Cros J, Patil Y, Martinet M, et al. Inter- and intra-tumoural heterogeneity in cancer-associated fibroblasts of human pancreatic ductal adenocarcinoma. *J Pathol* (2019) 248:51–65. doi: 10.1002/path.5224

20. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov* (2019) 9:1102–23. doi: 10.1158/2159-8290.Cd-19-0094

21. Dominguez CX, Müller S, Keerthivasan S, Koeppen H, Hung J, Gierke S, et al. Single-cell RNA sequencing reveals stromal evolution into LRRC15(+) myofibroblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discov* (2020) 10:232–53. doi: 10.1158/2159-8290.Cd-19-0644

22. Chen Y, Kim J, Yang S, Wang H, Wu CJ, Sugimoto H, et al. Type I collagen deletion in αSMA(+) myofibroblasts augments immune suppression and accelerates progression of pancreatic cancer. *Cancer Cell* (2021) 39:548–565.e546. doi: 10.1016/j.ccell.2021.02.007

23. Yin L, Lu Y, Cao C, Lu Z, Wei J, Zhu X, et al. CA9-related acidic microenvironment mediates CD8+ T cell related immunosuppression in pancreatic cancer. *Front Oncol* (2021) 11:832315. doi: 10.3389/fonc.2021.832315

24. Ye Z, Liu D, Liu D, Lv Y, Zhang Y, Zhang J, et al. Immune infiltration of CD8+ T cells in patients with diabetic pancreatic cancer reduces the malignancy of cancer tissues: An *in silico* study. *Front Endocrinol (Lausanne)* (2021) 12:826667. doi: 10.3389/fendo.2021.826667

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

26. Gao Y, Chen S, Vafaei S, Zhong X. Tumor-infiltrating immune cell signature predicts the prognosis and chemosensitivity of patients with pancreatic ductal adenocarcinoma. *Front Oncol* (2020) 10:557638. doi: 10.3389/fonc.2020.557638

27. Tang P, Qu W, Wu D, Chen S, Liu M, Chen W, et al. Identifying and validating an acidosis-related signature associated with prognosis and tumor immune infiltration characteristics in pancreatic carcinoma. *J Immunol Res* (2021) 2021:3821055. doi: 10.1155/2021/3821055

28. Zhao G, Lu H, Chang Z, Zhao Y, Zhu T, Chang L, et al. Single-cell RNA sequencing reveals the cellular heterogeneity of aneurysmal infrarenal abdominal aorta. *Cardiovasc Res* (2021) 117:1402–16. doi: 10.1093/cvr/cvaa214

29. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* (2009) 10:515–34. doi: 10.1093/biostatistics/kxp008

30. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* (2021) 12:1088. doi: 10.1038/s41467-021-21246-9

31. Potts HG, Lemieux ME, Rice ES, Warren W, Choudhury RP, Mommersteeg MTM. Discordant genome assemblies drastically alter the interpretation of single-cell RNA sequencing data which can be mitigated by a novel integration method. *Cells* (2022) 11:608. doi: 10.3390/cells11040608

32. Corridoni D, Antanaviciute A, Gupta T, Fawkner-Corbett D, Aulicino A, Jagielowicz M, et al. Single-cell atlas of colonic CD8(+) T cells in ulcerative colitis. *Nat Med* (2020) 26:1480–90. doi: 10.1038/s41591-020-1003-4

33. Tu SL, Upton C. Bioinformatics for analysis of poxvirus genomes. *Methods Mol Biol* (2019) 2023:29–62. doi: 10.1007/978-1-4939-9593-6_2

34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An r package for comparing biological themes among gene clusters. *Omics* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

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

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

37. Zhang LP, Ren H, Du YX, Wang CF. Prognostic value of the preoperative fibrinogen-to-albumin ratio in pancreatic ductal adenocarcinoma patients undergoing R0 resection. *World J Gastroenterol* (2020) 26:7382–404. doi: 10.3748/wjg.v26.i46.7382

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

39. Li GM, Zhang CL, Rui RP, Sun B, Guo W. Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy. *Eur Rev Med Pharmacol Sci* (2018) 22:3553–69. doi: 10.26355/eurrev_201806_15182

40. Grossberg AJ, Chu LC, Deig CR, Fishman EK, Hwang WL, Maitra A, et al. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. *CA Cancer J Clin* (2020) 70:375–403. doi: 10.3322/caac.21626

41. Singhi AD, Koay EJ, Chari ST, Maitra A. Early detection of pancreatic cancer: Opportunities and challenges. *Gastroenterology* (2019) 156:2024–40. doi: 10.1053/j.gastro.2019.01.259

42. Bear AS, Vonderheide RH, O'Hara MH. Challenges and opportunities for pancreatic cancer immunotherapy. *Cancer Cell* (2020) 38:788–802. doi: 10.1016/j.ccell.2020.08.004

43. Stark AP, Sacks GD, Rochefort MM, Donahue TR, Reber HA, Tomlinson JS, et al. Long-term survival in patients with pancreatic ductal adenocarcinoma. *Surgery* (2016) 159:1520–7. doi: 10.1016/j.surg.2015.12.024

44. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* (2016) 16:582–98. doi: 10.1038/nrc.2016.73

45. Chang JH, Jiang Y, Pillarisetty VG. Role of immune cells in pancreatic cancer from bench to clinical application: An updated review. *Med (Baltimore)* (2016) 95:e5541. doi: 10.1097/md.0000000000005541

46. Leinwand J, Miller G. Regulation and modulation of antitumor immunity in pancreatic cancer. *Nat Immunol* (2020) 21:1152–9. doi: 10.1038/s41590-020-0761-y

47. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. *Nat Rev Gastroenterol Hepatol* (2019) 16:207–20. doi: 10.1038/s41575-019-0109-y

48. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet* (2014) 30:418–26. doi: 10.1016/j.tig.2014.07.001

49. Bevilacqua C, Ducos B. Laser microdissection: A powerful tool for genomics at cell level. *Mol Aspects Med* (2018) 59:5–27. doi: 10.1016/j.mam.2017.09.003

50. Burghel GJ, Quinn AM, Smith P, Harris J, Banks C, Wallace A. The importance of neoplastic cell content assessment and enrichment by macrodissection in cancer pharmacogenetic testing. *J Clin Pathol* (2019) 72:721–2. doi: 10.1136/jclinpath-2019-206094

51. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* (2008) 321:1801–6. doi: 10.1126/science.1164368

52. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* (2012) 491:399–405. doi: 10.1038/nature11547

53. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* (2015) 518:495–501. doi: 10.1038/nature14169

54. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* (2015) 47:1168–78. doi: 10.1038/ng.3398

55. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, et al. Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. *Gastroenterology* (2018) 155:1999–2013.e1993. doi: 10.1053/j.gastro.2018.08.033

56. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* (2011) 17:500–3. doi: 10.1038/nm.2344

57. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* (2016) 531:47–52. doi: 10.1038/nature16965

58. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. *Cell* (2012) 148:349–61. doi: 10.1016/j.cell.2011.11.025

59. Petersen OW, Nielsen HL, Gudjonsson T, Villadsen R, Rank F, Niebuhr E, et al. Epithelial to mesenchymal transition in human breast cancer can provide a nonmalignant stroma. *Am J Pathol* (2003) 162:391–402. doi: 10.1016/s0002-9440(10)63834-5

60. Sun K, Tang S, Hou Y, Xi L, Chen Y, Yin J, et al. Oxidized ATM-mediated glycolysis enhancement in breast cancer-associated fibroblasts contributes to tumor invasion through lactate as metabolic coupling. *EBioMedicine* (2019) 41:370–83. doi: 10.1016/j.ebiom.2019.02.025

61. Penkert J, Ripperger T, Schieck M, Schlegelberger B, Steinemann D, Illig T, et al. On metabolic reprogramming and tumor biology: A comprehensive survey of metabolism in breast cancer. *Oncotarget* (2016) 7:67626–49. doi: 10.18632/oncotarget.11759

62. Martinez-Outschoorn UE, Lisanti MP, Sotgia F. Catabolic cancer-associated fibroblasts transfer energy and biomass to anabolic cancer cells, fueling tumor growth. *Semin Cancer Biol* (2014) 25:47–60. doi: 10.1016/j.semcancer.2014.01.005

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 16 December 2022
DOI 10.3389/fimmu.2022.1040661

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Daheng Yang,
Nanjing Medical University, China
Yirui Chen,
Hangzhou Medical College, China

*CORRESPONDENCE
Tianling Ding
dtl_953105@163.com
Yanping Shao
shaoyp2022@163.com

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

RECEIVED 09 September 2022
ACCEPTED 31 October 2022
PUBLISHED 16 December 2022

CITATION
Hua J, Ding T and Shao Y (2022) A transient receptor potential channel-related model based on machine learning for evaluating tumor microenvironment and immunotherapeutic strategies in acute myeloid leukemia.
Front. Immunol. 13:1040661.
doi: 10.3389/fimmu.2022.1040661

A transient receptor potential channel-related model based on machine learning for evaluating tumor microenvironment and immunotherapeutic strategies in acute myeloid leukemia

Jingsheng Hua[1], Tianling Ding[2*] and Yanping Shao[3*]

[1]Department of Hematology, Taizhou Municipal Hospital, Taizhou, China, [2]Department of Hematology, Huashan Hospital, Fudan University, Shanghai, China, [3]Department of Hematology, Taizhou Hospital of Zhejiang Province affiliated to Wenzhou Medical University, Taizhou, China

Background: Acute myeloid leukemia (AML) is an aggressive hematopoietic malignancy. Transient receptor potential (TRP) channels in AML still need to be further explored. A TRP channel-related model based on machine learning was established in this study.

Methods: The data were downloaded from TCGA-LAML and Genome-Tissue Expression (GTEx). TRP-related genes (TRGs) were extracted from previous literature. With the use of Single-Sample Gene Set Enrichment Analysis (ssGSEA), TRP enrichment scores (TESs) were calculated. The limma package was used to identify differentially expressed genes (DEGs), and univariate Cox regression analysis was performed to identify prognostic DEGs. The above prognostic DEGs were analyzed by Random Survival Forest and least absolute shrinkage and selection operator (Lasso) analysis to create the TRP signature. The Kaplan–Meier and receiver operating characteristic (ROC) curves were plotted to investigate the efficiency and accuracy of prognostic prediction. Moreover, genomic mutation analysis was based on GISTIC analysis. Based on ESTIMATE, TIMER, MCPcounter, and ssGSEA, the tumor microenvironment and immunological characteristics were expressly evaluated to explore immunotherapeutic strategies. Enrichment analysis for TRP signature was based on the Kyoto Encyclopedia of Genes Genomes (KEGG), Gene Ontology (GO), over-representation analysis (ORA), and Gene Set Enrichment Analysis (GSEA). Genomics of Drug Sensitivity in Cancer (GDSC) and pRRophetic were used to carry out drug sensitivity analysis. Conclusively, SCHIP1 was randomly selected to perform *in vitro* cyto-functional experiments.

10.3389/fimmu.2022.1040661

Results: The worse clinical outcomes of patients with higher TESs were observed. There were 107 differentially expressed TRGs identified. Our data revealed 57 prognostic TRGs. Eight TRGs were obtained to establish the prognostic TRP signature, and the worse clinical outcomes of patients with higher TRP scores were found. The efficiency and accuracy of TRP signature in predicting prognosis were confirmed by ROC curves and five external validation datasets. Our data revealed that the mutation rates of DNMT3A, IDH2, MUC16, and TTN were relatively high. The level of infiltrating immune cell populations, stromal, immune, and ESTIMATE scores increased as the TRP scores increased. Nevertheless, AML patients with lower TRP scores exhibited more tumor purity. The TRP scores were found to be correlated with immunomodulators and immune checkpoints, thus revealing immune characteristics and immunotherapeutic strategies. The IC50 values of six chemotherapeutics were lower in the high TRP score (HTS) group. Finally, it was found that SCHIP1 may be the oncogenic gene.

Conclusion: The results of this study will help in understanding the role of TRP and SCHIP1 in the prognosis and development of AML.

KEYWORDS

acute myeloid leukemia, TRP, signature, schip1, machine learning

Introduction

Acute myeloid leukemia (AML) is an aggressive hematopoietic malignancy caused by the malignant transformation of hematopoietic stem cells or progenitor cells, which is highly heterogeneous (1). It is the most common acute leukemia in adults, with an annual incidence of approximately four per 100,000 cases (1). Patients with AML generally have a poor prognosis (1). Therefore, it is of great significance for clinical treatment to find molecular markers that can judge the prognosis and effectively distinguish whether patients can benefit from treatment. Advances in genomics have greatly improved our understanding of the pathogenesis of AML, which is one of the targets in the search for diagnosis and treatment of AML.

The prognosis of the same AML type may be very heterogeneous. Therefore, it is of great importance to evaluate the characteristics of each AML patient. In the past 40 years, many new achievements have been made in pathogenesis, but there is no innovative progress in the treatment of AML. The traditional treatment of AML mainly includes three parts: induction regimen therapy, monitoring after induction therapy, and treatment after complete remission (CR) (2). In the post-CR treatment, patients under the age of 60 should choose the appropriate treatment according to the risk stratification, indicating the important role of genetic risk stratification in guiding the treatment of AML. Traditional cytogenetic classification includes better karyotypes, intermediate karyotypes, and poor karyotypes. With the deepening of research on leukemia, people can have more profound knowledge of the pathogenesis of leukemia, and the risk classification of AML combined with genetic changes is more recommended by most guidelines (1). However, the risk classification of AML needs to be further explored.

Bioinformatics is an interdisciplinary subject involving mathematics, statistics, computer science, biology, and other disciplines. After decades of development, bioinformatics is still a subject with great development prospects (3). In recent years, a large number of databases containing biological information of various species have been established worldwide. As the largest database for cancer research, The Cancer Genome Atlas (TCGA) (4) database stores rich sequencing data and clinical information, which can be downloaded by researchers all over the world for free for research so as to improve the cognitive ability of doctors and researchers on the disease. AML project, as one of the earliest and most well-developed projects in TCGA, stores a large amount of sequencing and clinical data, which can be used to prospectively explore some disease-related information and provide directions for clinical and experimental research.

Transient receptor potential (TRP) was first discovered in *Drosophila* (5). When *Drosophila* bearing the mutant gene is exposed for an extended time to light, its photoreceptor will show a transient increase of voltage, so it is named transient receptor potential channel (5). Mammalian TRP channels are

composed of 28 cation permeation channels (6), each with six transmembrane peptides, which assemble into tetramers to form ion channels (6). TRP channels have various types, such as TRPC, TRPA, TRPM, TRPN, TRPP, and TRPV (7). TRP channels are involved in various physiological and pathological processes of the body and respond to physical or chemical stimuli in the cellular environment by sensing them (8). TRP channel is closely related to circulatory (9), urinary (10), digestive (11, 12), nervous (13, 14), and other systems. It is a cation channel widely existing in the body, mainly permeating Ca^{2+}, Mg^{2+}, and other cations (15). By affecting the change of cation concentration, the TRP channel changes the strength of the corresponding pathway signal in the cell, leading to the change of cell function (15). As for AML, TRP was seldom reported in AML. TRP ion channel TRPM2 could enhance the proliferation of AML cell lines through multiple pathways (16). TRP Melastatin Subfamily Member 4 may be an alternative therapeutic approach for AML (17).

Therefore, TRP may be used as a prognostic and therapeutic target. However, the TRP channel in AML still needs to be further explored.

In this study, bioinformatics analysis was used to screen out genes' expression of AML, including TRP-related genes (TRGs) from public databases, and to analyze the characteristics of TRP enrichment in TCGA-LAML. By combining survival information and gene expression, 57 prognostic TRGs were preliminarily identified as possible AML target genes. Through the Random Survival Forest model and least absolute shrinkage and selection operator (Lasso) analysis, a predictive model consisting of eight genes was established and validated in five external datasets, thus proving good predictive ability. The risk prognosis model score was used to group the high- and low-risk groups, and it was found that the risk groups differed in immune profiles and treatment.

Material and methods

Data collection and preprocessing for acute myeloid leukemia

The transcriptome expression profile and corresponding clinical information of patients diagnosed with AML were downloaded from TCGA-LAML dataset in the UCSC Xena platform (https://xenabrowser.net/) (18). There were a total of 149 AML patients with corresponding data included in our study (N = 149). In addition, the transcriptome expression profile of corresponding normal control samples was downloaded from Genome-Tissue Expression (GTEx) project (https://www.gtexportal.org) (19). Meanwhile, five AML cohorts were collected, including GSE12417 (N = 79), GSE12417 (N = 163), GSE37642 (N = 136), GSE37642 (N417), and TARGET (N = 187), from Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) or Therapeutically Applicable Research to Generate Effective Treatments (TARGET; https://ocg.cancer.gov/programs/target/data-matrix) (20). The GEO data were generated from the Affymetrix (21) or Agilent (22) platform. Background correction and normalization for GEO data were carried out using Robust Multichip Average (RMA) algorithm (23). The data forms of TCGA and TARGET were transformed from fragments per kilobase of transcript per million fragments mapped (FPKM) to transcripts per kilobase million (TPM), of which the signal strength was similar to the value processed by RMA (24).

Establishment of transient receptor potential enrichment score

The list of TRGs was extracted from the previous literature (25), which was used for enrichment score calculation. There were eight TRGs included in our study: TRPM1, TRPM2, TRPM3, TRPM4, TRPM5, TRPM6, TRPM7, and TRPM8. To identify TRP-related patterns, TRP enrichment scores (TESs) were calculated for each AML patient using Single-Sample Gene Set Enrichment Analysis (ssGSEA) algorithm (26). According to the optimal cutoff value of TESs calculated by R code (27), patients with AML were divided into the high-TES group (≥cutoff value) or low-TES group (<cutoff value).

Establishment of transient receptor potential signature

The limma package was used to identify differentially expressed genes (DEGs) between the high-TES and low-TES groups ($logFC > 1$, $p < 0.05$) (28). Thereafter, univariate Cox regression analysis was performed to identify prognostic DEGs ($p < 0.05$) (29), and the Random Survival Forest model was utilized to screen out prognostic DEGs with higher importance (variable importance >0.3) based on the randomForestSRC package (30). To establish the TRP signature, the weight of regression coefficients of the prognostic genes identified by the Random Forest Algorithm was calculated using Lasso analysis (31), thus establishing the signature and computing the TRP score.

Efficacy of transient receptor potential signature

The TRP score for 149 patients in TCGA-LAML cohort was estimated according to the method described above. The optimal cutoff was considered based on R code (27) as the threshold value to distinguish subgroups with high TRP scores (HTS) or low TRP scores (LTS). We compared survival differences

between the two subgroups to assess the relationship between TRP score and overall survival (OS) by plotting Kaplan–Meier survival curves (32). Through the timeROC package, the 1-, 3-, and 5-year survival receiver operating characteristic (ROC) curves were plotted to investigate the efficiency and accuracy of prognostic prediction for the TRP score. To further verify the independence of the TRP score predicting prognosis for AML, univariate or multivariate Cox regression analyses of the TRP score and clinicopathological characteristics [age, gender, and white blood cell (WBC)] were performed.

Genomic mutation analysis for acute myeloid leukemia with transient receptor potential score

Somatic mutation profiles of AML were obtained from cBioPortal (http://www.cbioportal.org/datasets) (33). Meanwhile, copy number variation (CNV) analysis was carried out after extracting data from FireBrowse (http://firebrowse.org/) (34). The genomic characteristics were assessed using Genomic Identification of Significant Targets in Cancer (GISTIC) analysis (35).

Evaluation of immunological characteristics

We used the ESTIMATE (The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression) algorithm to assess the abundance of immune cells, stromal cell infiltration level, and tumor purity and expressed them as immune score, stromal score, and ESTIMATE score, respectively (36). In addition, in order to comprehensively analyze the infiltration of immune cells in AML, we further analyzed the levels of six kinds of cells by using the TIMER 2.0 (Tumor Immune Estimation Resource 2.0) network server (http://timer.cistrome.org/) (37). We also used MCPcounter (38) and ssGSEA (26, 39) to assess the relative proportions of 10 immune cells and the infiltration levels of 28 immune cells, respectively. We extracted several immunomodulators from literature reported previously to explore the association between TRP score and immune processes (40).

Enrichment analysis for transient receptor potential signature

Downloading from the MSigDB database, we acquired gene sets using for Kyoto Encyclopedia of Genes Genomes (KEGG) or Gene Ontology (GO) analyses (41). We implemented over-representation analysis (ORA) (42) and Gene Set Enrichment Analysis (GSEA) by using the clusterProfiler package (43).

Drug sensitivity analysis

The Genomics of Drug Sensitivity in Cancer (GDSC) website was utilized to screen a wide range of drugs (44). The prediction model was constructed based on Ridge's regression between drug sensitivity and expression profile of cell lines using the pRRophetic algorithm (45, 46). Subsequently, the IC50 value of corresponding chemotherapeutics for patients with AML was calculated.

Cell culture

We randomly selected one gene, SCHIP1, from the TRP signature to perform *in vitro* cyto-functional experiments. We used one AML cell line, called K562, for *in vitro* assays. We incubated the AML cell line K562 in the incubator at an atmosphere of 37°C and 5% CO_2 and cultured it in 90% Roswell Park Memorial Institute 1640 (RPMI 1640) medium with 10% fetal bovine serum (FBS).

Cell transfections

To perform cell transfections, Hieff Trans™ *in vitro* siRNA Transfection Reagent supplied by Yeasen Biotechnology (Shanghai, China) was used, and the sequences of siRNA were as follows: si-NC (control group) sense (5′-UUCUUCGAACGUGUCACGUTT-3′), si-NC antisense (5′-ACGUGACACGUUCGGAGAATT-3′), si-SCHIP1 sense (5′-GGAGUCUGAAUCCUU GGAUTT-3′), and si-SCHIP1 antisense (5′-AUCCAAGGAUUCAGACUCCTT-3′). According to the kit instructions, the transfection steps were as follows: cells were collected, plates were spread on a six-well plate, and the number of cells on the transfection day was 5×10 (5) to 2×10 (6). OPTI-MEM medium, siRNA, and transfection reagent were used to prepare siRNA-PEI cationic nucleic acid transfection reagent complex and added to the cell suspension. After 4–6 h in a 5% CO_2 incubator at 37°C, 2 ml of complete medium was added and incubated in an incubator for 72 h. The efficiency of SCHIP1 knockdown in K562 cells was confirmed by Western blotting assays.

Western blotting assays

Protein was extracted through protein extraction reagents containing inhibitors. Ten microliters of protease inhibitor mixture, 10 µl of phenylmethylsulfonyl fluoride (PMSF), and 10 µl phosphatase mixture were added to 1 ml of the extraction reagent. The bicinchoninic acid (BCA) method was performed for protein detection: 25 µl of standard and sample to be tested was added to the microwells, 200 µl of BCA working solution

was added to each well, the samples were incubated at 37°C for 30 min, and then the absorbance was detected at 562 nm on a microplate reader. TEMED containing 10% separation glue and 5% concentrate glue for gluing was successively used. After loading, it was electrophoresed with glycine buffer. After electrophoresis, a polyvinylidene difluoride (PVDF) membrane was used for the transmembrane of the gel. After the membrane transfer, the membrane was blocked with 5% non-fat milk and then washed three times with TBST. After blocking, the cells were incubated with primary antibodies at 4°C overnight. Before and after incubation with a secondary antibody for 1 h at room temperature, the membrane was washed with TBST three times. Finally, the color was defined according to the chemiluminescence kit, photos were taken, and statistics and analysis were performed on the gel imaging system.

Cell Counting Kit-8 assays

Cells were collected at a concentration of 1×10 (4) cells/ml. Each well of the 96-well plate was inoculated with 100 µl of cell suspension, and each group had three wells. Ten microliters of si-SCHIP1 or si-NC was added to the corresponding wells and then placed into the incubator for routine culture. The next day, 10 µl of Cell Counting Kit-8 (CCK8) solution was added at a fixed time and incubated in the incubator for 0.5–4 h. Finally, the absorbance at 450 nm was measured by a microplate reader, and the cell viability was calculated.

Statistical analysis

Normally distributed variables and non-normally distributed data between two groups were compared by t-test and Wilcoxon test, respectively. OS status estimated by Kaplan–Meier survival curves and Cox regression used for survival analysis were compared by the survminer package. ROC curves were plotted by the timeROC package, and heatmaps were plotted by the pheatmap package. R package ggplot2 (v4.1.2) was used to visualize the data. *In vitro* assays were performed for more than three independent experiments or replicates. $p < 0.05$ was considered statistically significant.

Results

Characteristics of transient receptor potential enrichment in TCGA-LAML

We calculated TESs for each AML patient using the ssGSEA algorithm. The correlations among the TRGs, clinicopathological characteristics, and TESs are exhibited in Figure 1A. Compared with AML patients with lower TESs, the expressions of TRPM1, TRPM2, TRPM5, and TRPM5 were relatively high; on the contrary, the expressions of TRPM6 and TRPM7 were relatively low (Figure 1A). We distinguished AML patients into the HTS group and LTS group. From Figure 1B, we can observe the significantly worse clinical outcomes of patients with higher TESs, while the prognosis of patients with lower TESs was better. Therefore, TES may be a driving factor for the malignant progression of AML. There were 107 differentially expressed TRGs identified by differential analysis ($logFC > 1$, $p < 0.05$), which could be reflected in the volcano map (Figure 1C). We carried out an enrichment analysis to explore the biological function of these differentially expressed TRGs. The GO analysis (Figure 1D) showed that these TRGs were significantly enriched in several immune-related pathways (neutrophil activation, neutrophil degranulation, neutrophil activation involved in immune response, neutrophil-mediated immunity, defense response to bacterium, defense response to fungus, negative regulation of immune system process, leukocyte migration, macrophage activation, and macrophage differentiation). KEGG analysis (Figure 1E) revealed that these TRGs were significantly enriched in some classical tumor-related pathways (Transcriptional misregulation in cancer, IL-17 signaling pathway, Arachidonic acid metabolism, Influenza A, C-type lectin receptor signaling pathway, Bladder cancer, Serotonergic synapse, Malaria, Shigellosis, and Melanoma).

Establishment of transient receptor potential signature

The univariate Cox regression analysis was performed on the differentially expressed TRGs obtained above. The results revealed 57 prognostic TRGs (Figure 2A), including 24 potential tumor-protective factors (hazard ratio (HR) < 1) and 33 potential tumor-promoting factors (HR > 1). Thereafter, the distribution of error rates generated by the Random Survival Forest model is shown in Figure 2B, thus identifying the variable importance (variable importance >0.3, Figure 2B) of 12 TRGs (ZNF608, NAPSB, CPNE8, ANXA8, LPO, PDCD6IPP1, SLC2A5, SCHIP1, HOXA4, TRH, LST1, and METTL7B). Lasso analysis was used to construct the TRP signature, and the TRP score for 149 patients with AML was calculated. Ultimately, eight TRGs (ANXA8, CPNE8, HOXA4, LPO, LST1, METTL7B, NAPSB, PDCD6IPP1, SCHIP1, SLC2A5, TRH, and ZNF608) were obtained to establish the prognostic signature, and Figure 2C displays the lambda selection diagram. The heatmap displays the distribution of the eight TRGs of the signature, clinicopathological characteristics, and TRP score. It can be clearly observed that high expression of LPO and TRH may be associated with lower TRP scores, in contrast to high expression of NAPSB, METTL7B, SLC2A5, SCHIP1, PDCD6IPP1, and HOXA4, which was associated with higher TRP scores (Figure 2D).

10.3389/fimmu.2022.1040661

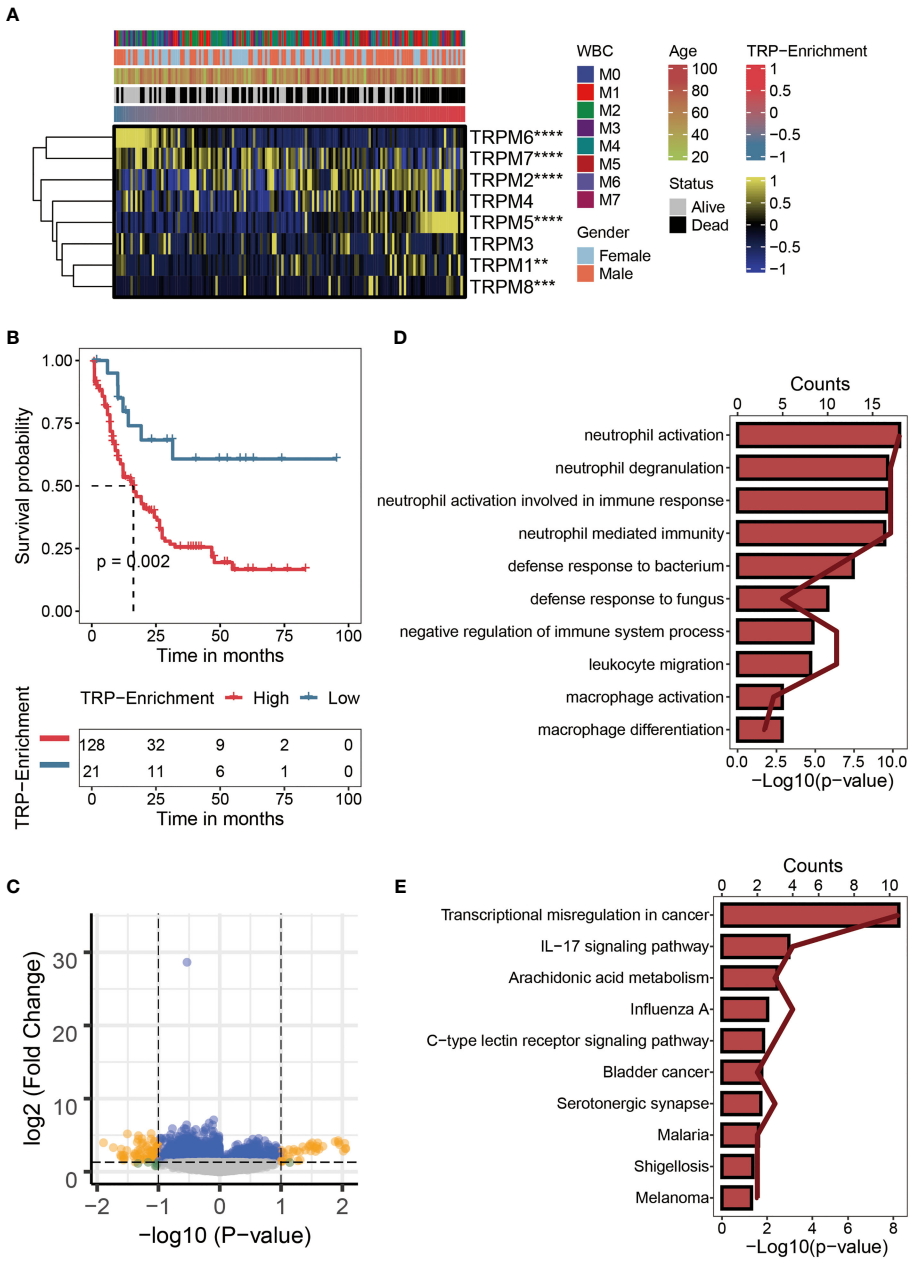


FIGURE 1
Characteristics of TRP enrichment scores in TCGA-LAML cohort. **(A)** Correlation between TRP enrichment scores and the expression values of eight TRP genes in TCGA-LAML cohort. Yellow represents high gene expression; blue represents low gene expression. **(B)** Kaplan–Meier curve showing the correlation between TRP enrichment scores and survival status of AML patients. The blue curve represents the group with lower TRP enrichment scores, and the red curve represents group with higher TRP enrichment scores. **(C)** Volcano map of differential analysis between high and low TRP enrichment groups. There were 107 differentially expressed genes between the two groups. Yellow dots indicate genes whose expression values differ between the two groups, while blue dots indicate genes whose expression values do not differ between the two groups. **(D)** GO enrichment map of 107 differentially expressed genes. **(E)** KEGG enrichment map of 107 differentially expressed genes. TRP, transient receptor potential; AML, acute myeloid leukemia; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes Genomes. **,<0.01; *** ,<0.001; and ****,<0.0001.

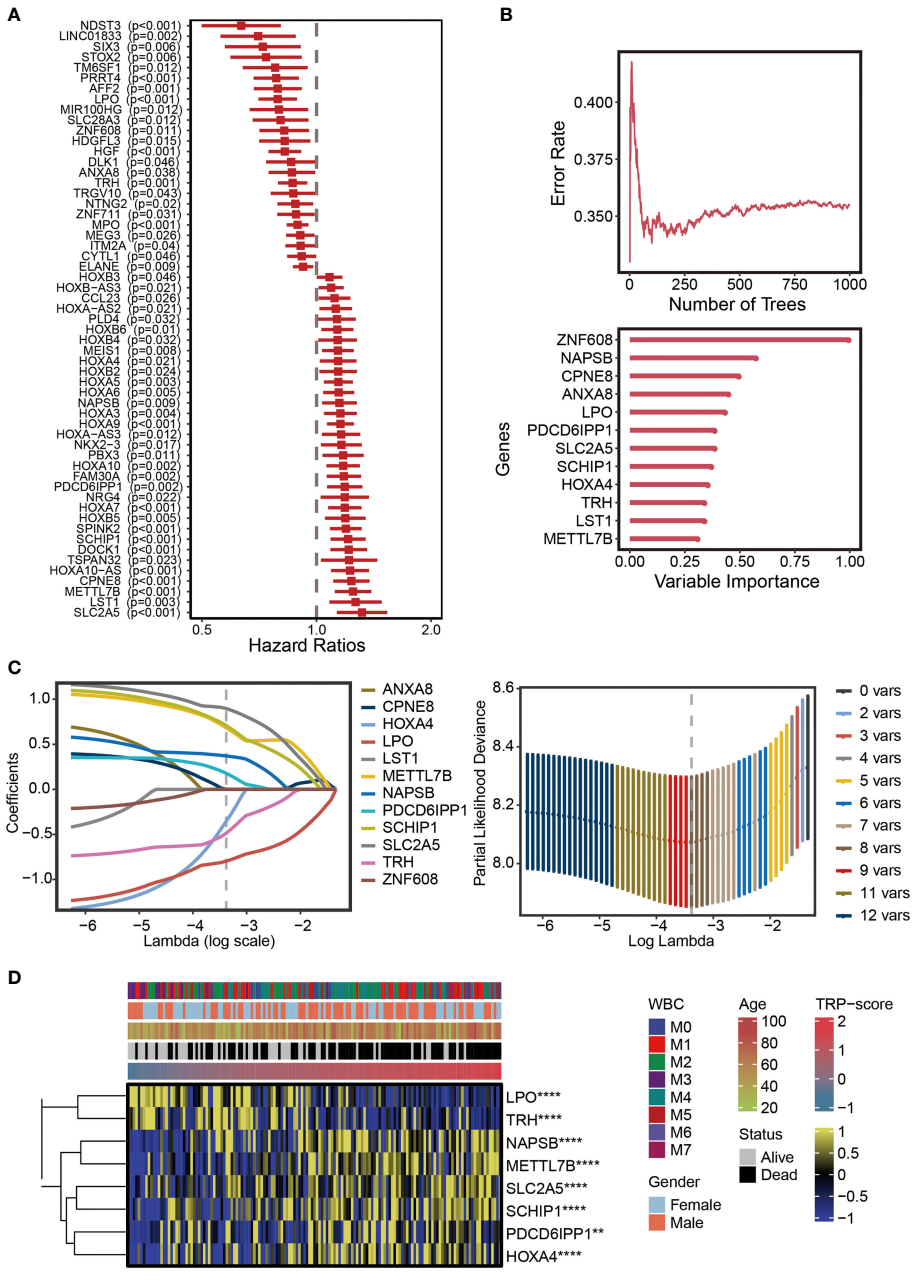


FIGURE 2
Establishment of TRP signature. **(A)** Forest plot for univariate Cox regression analysis of 57 prognostic TRP-related genes. **(B)** The distribution of error rates in Random Survival Forest model and the variable relative importance of 12 TRP-related genes (variable importance >0.3). **(C)** Lambda selection diagram for least absolute shrinkage and selection operator (Lasso) analysis. **(D)** The heatmap displaying the distribution of the eight TRP-related genes of the signature, clinicopathological characteristics, and TRP enrichment scores. Yellow represents high gene expression; blue represents low gene expression. TRP, transient receptor potential. **,<0.01; and ****,<0.0001.

Efficacy of transient receptor potential signature

The optimal cutoff of TRP scores was set as a threshold value to distinguish AML patients into the HTS or LTS groups. The details of clinic information are listed in Supplementary Table S1. The survival curves showed significantly worse clinical outcomes of patients with higher TRP scores, while the prognosis of patients with lower TRP scores was better (Figure 3A). The area under the curve (AUCs) values of 1-year (AUC = 0.738), 3-year (AUC = 0.796), and 5-year (AUC = 0.858) survival ROC curves predicted by the TRP signature were

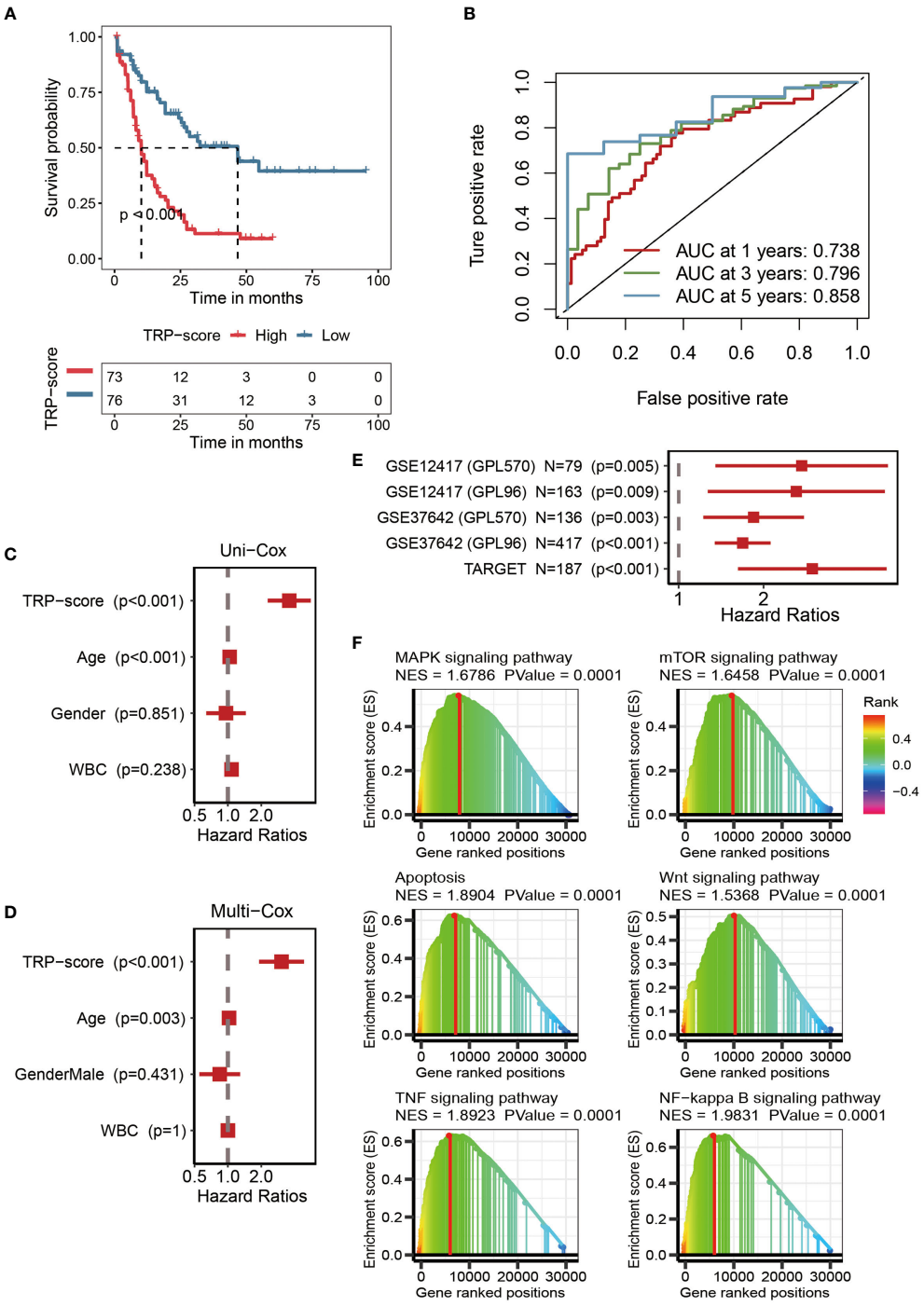


FIGURE 3
Efficacy of TRP signature. **(A)** Kaplan–Meier survival curve showing survival probability of high TRP score or low TRP score subgroups. The blue curve represents the group with lower TRP scores, and the red curve represents group with higher TRP scores. **(B)** The 1-year (0.738), 3-year (0.796), and 5-year (0.858) survival ROC curves predicted by the TRP signature. **(C)** The forest figure for univariate Cox regression analysis of TRP score and clinicopathological features. **(D)** The forest figure for multivariate Cox regression analysis of TRP score and clinicopathological features. **(E)** Univariate Cox regression analysis of the TRP signature in five external validation datasets (GSE12417-GPL570, GSE12417-GPL96, GSE37642-GPL570, GSE37642-GPL96, and TARGET). **(F)** GSEA showing cancer-related pathways positively regulated by TRP signature. TRP, transient receptor potential; ROC, receiver operating characteristic; GSEA, Gene Set Enrichment Analysis.

all higher than 0.7, suggesting the efficiency of TRP signature in predicting prognosis for AML (Figure 3B). Furthermore, TRP signature was an independent prognostic factor for AML patients as demonstrated by univariate (Figure 3C) and multivariate (Figure 3D) Cox regression analyses. Finally, univariate Cox regression analysis was conducted on five external validation datasets (GSE12417-GPL570, GSE12417-GPL96, GSE37642-GPL570, GSE37642-GPL96, and TARGET), and the HRs of the five sets were all greater than 1, demonstrating the accuracy of the TRP signature that we constructed in prognostic prediction (Figure 3E, Supplementary Figure S1). Based on GSEA, six cancer-related pathways (MAPK signaling pathway, TOR signaling pathway, Apoptosis, Wnt signaling pathway, TNF signaling pathway, and NF-kappa B signaling pathway) were identified, which may be positively regulated by this signature, which provided insights for exploring the mechanism of AML (Figure 3F).

Genomic mutation analysis for transient receptor potential signature

We assessed the genomic characterization landscape of the HTS group or LTS group by the GISTIC algorithm, as shown in Figure 4A. Further, we plotted the detailed amplificated or deleted CNV onco-plots of the HTS and LTS groups (Figure 4B). From Figure 4B, we can observe that the results of the two subgroups were similar. DNMT3A, FLT3, RUNX1, NPM1, TP53, NRAS, CACNA1B, IDH2, MUC16, TTN, ALOX12B, ASXL1, ATP10B, BBS12, and BRINP3 were the top 15 genes with the highest mutation rate in AML patients with high TRP scores (Figure 4C). MUC16, IDH2, KIT, TTN, DNMT3A, PRUNE2, UBR4, WT1, AHNAK, AHNAK2, CC2D2A, MACF1, NF1, PCLO, and VPS13D were the top 15 genes with the highest mutation rate in AML patients with low TRP scores (Figure 4C). Thus, the mutation rates of DNMT3A, IDH2, MUC16, and TTN in the two subgroups were relatively high.

Evaluation of immunological characteristics for transient receptor potential signature

After analysis based on MCPcounter, ssGSEA, and TIMER algorithms, the abundance of infiltrating immune cell populations with different TRP scores was displayed in the heatmap (Figure 5A). From a general view, the level of infiltrating immune cell populations (Figure 5A), stromal score (Figure 5B), immune score (Figure 5B), and ESTIMATE score (Figure 5B) increased as the TRP scores increased. Nevertheless, AML patients with lower TRP scores exhibited more tumor purity (Figure 5B). As for gene set variation analysis (GSVA), we focused on immune-related pathways positively regulated by TRP signature. The results showed that the TRP signature may be associated with adaptive immune response, immune response, innate immune response, T-cell receptor signaling pathway, interleukin-1-mediated signaling pathway, interferon-gamma-mediated signaling pathway, positive regulation of T cell proliferation, and T-cell activation (Figure 5C).

Immunotherapy and chemotherapy of transient receptor potential signature

Considering that immunomodulators (IMs) play a critical role in tumor immunotherapy, we compared the correlation between immunomodulator levels (Co-stm, Co-ihb, Ligand, Receptor, Cell adhesion, Antigen presentation, and Other) and the prognostic TRP signature (Figure 6A). To further evaluate the relationship between TRP score and immunotherapy, we calculated the correlation between the TRP scores and the expression level of four classical immune checkpoints, and we found that the score was correlated with PDCD1 ($R = 0.37$, $p = 2.5 \times e^{-6}$), CTLA4 ($R = 0.44$, $p = 1.5 \times e^{-8}$), CD274 ($R = 0.46$, $p = 2.5 \times e^{-9}$), and PDCD1LG2 ($R = 0.48$, $p = 7.1 \times e^{-10}$), which can provide an important reference for the immunotherapy of AML (Figure 6B). The IC50 values of six chemotherapeutics (PLX-4720, 5-Fluorouracil-1073, Dabrafenib-1373, Temozolomide-1375, LGK974-1598, and Foretinib-2040) were contrasted using violin figures, and our data revealed that the IC50 values of the chemotherapeutics mentioned above were lower in the HTS group than in the LTS group, suggesting that patients with higher TRP scores were more likely to benefit from these six chemotherapeutics (Figure 6C).

In vitro assays

To verify the effect of TRP score *in vitro*, we selected SCHIP1, as it represents genes of TRP score in further work. First, our results showed that there was a significant difference in the expression of SCHIP1 between the tumor and normal, and the SCHIP1 also had a poor prognosis in TCGA-AML cohort (Figures 7A, B). Then, after cell transfection, we observed the cell morphology under the microscope (Figure 7C). As shown in Figure 7D, 24 h after transfection, the cells in the NC group had regular shape and uniform size, and there was no significant difference between the si-NC and NC groups, while the cells in the si-SCHIP1 group had heterogeneous size and irregular shape, and some cells showed apoptosis. Forty-eight hours after transfection, the cell morphology of the NC group and si-NC group was regular, and there was no significant difference between the two groups, while the si-SCHIP1 group showed significant apoptosis. We used Western blotting assays to detect the knockout efficiency of SCHIP1 gene, and the results showed

10.3389/fimmu.2022.1040661

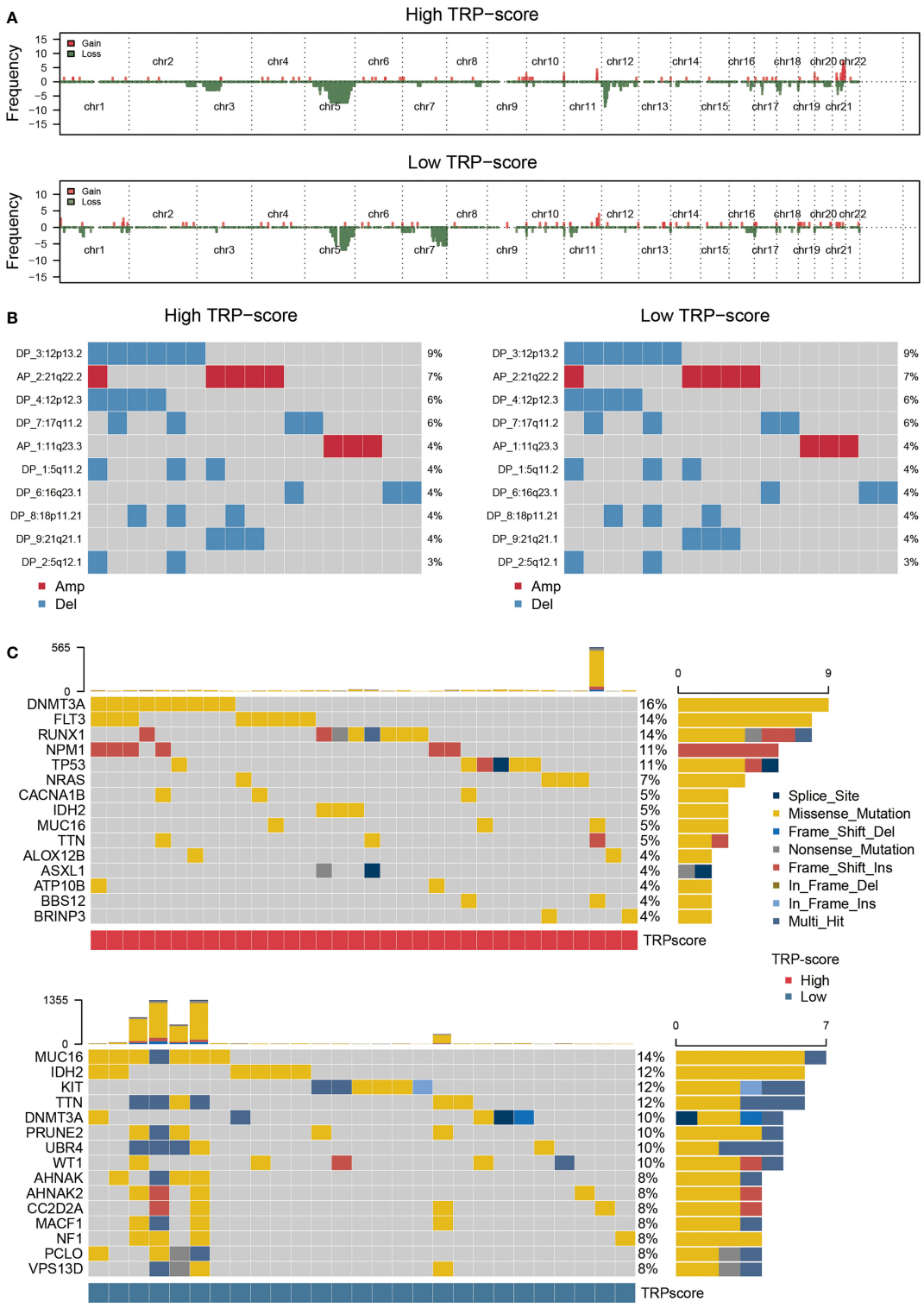


FIGURE 4
Genomic mutation analysis for TRP signature. **(A)** Genomic characterization landscape of groups with high TRP scores or low TRP scores. **(B)** The detailed amplificated or deleted CNV onco-plots of groups with high TRP scores or low TRP scores. **(C)** Waterfall plot of somatic mutations in AML between high and low TRP score groups. TRP, transient receptor potential; AML, acute myeloid leukemia; CNV, copy number variation.

10.3389/fimmu.2022.1040661

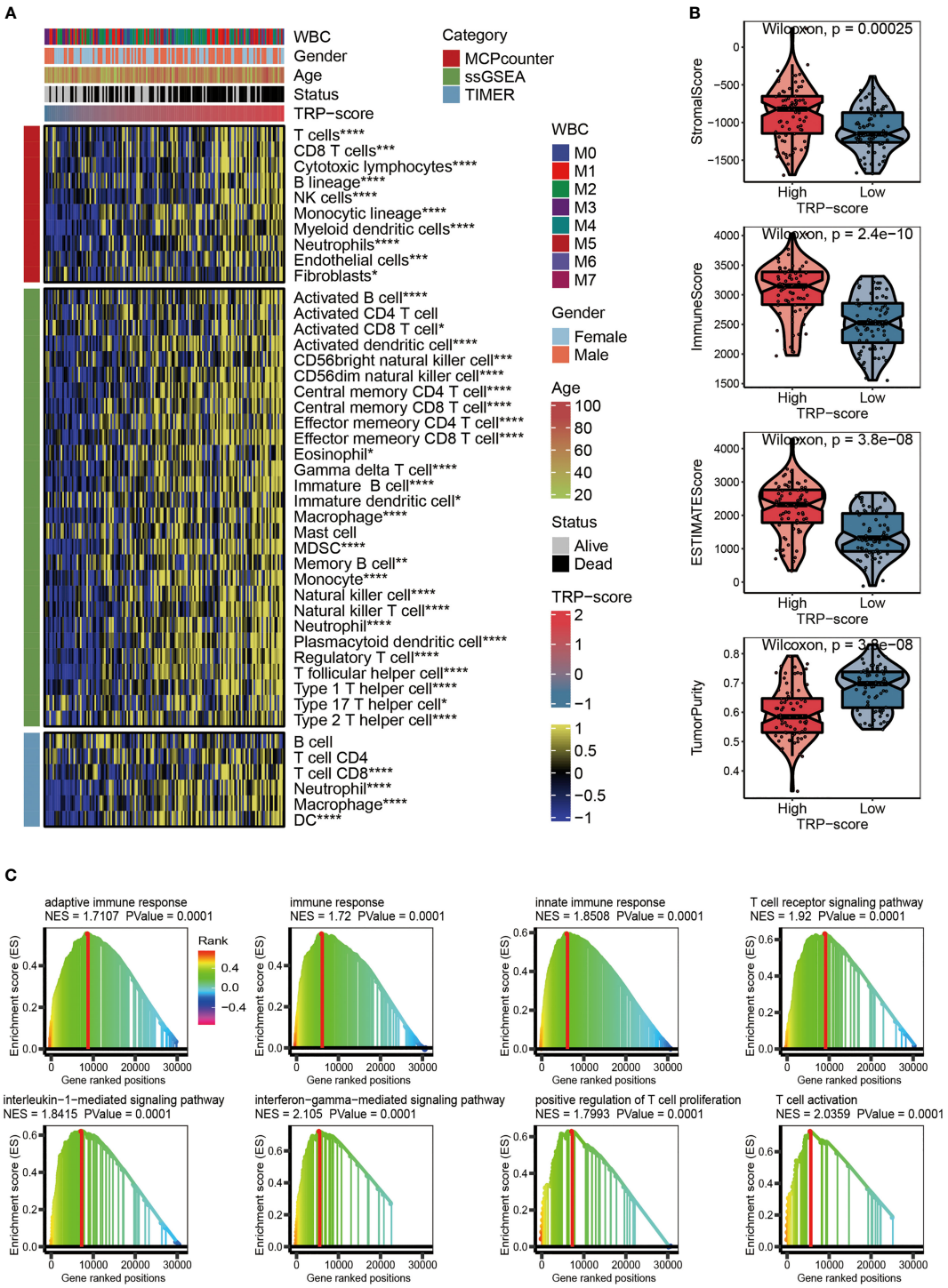


FIGURE 5
Evaluation of immunological characteristics for TRP signature. **(A)** Heatmap displaying the abundance of infiltrating immune cell populations with different TRP scores. **(B)** The violin chart comparing the differences between high and low TRP scores on stromal score, immune score, ESTIMATE score, and tumor purity. **(C)** GSVA for immune-related pathways positively regulated by TRP signature. TRP, transient receptor potential; GSVA, gene set variation analysis. *,<0.05; **,<0.01; ***,<0.001; and ****,<0.0001.

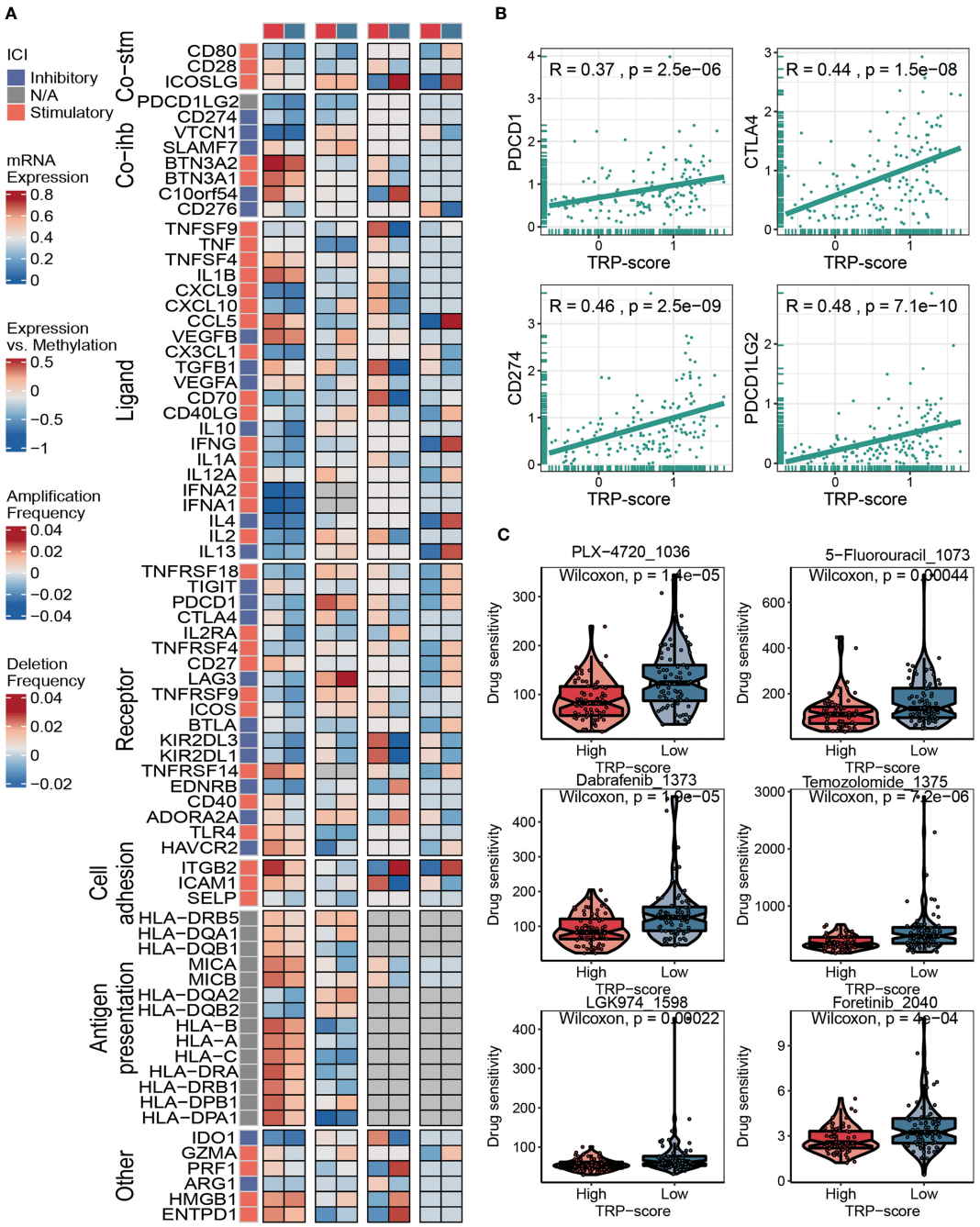


FIGURE 6
Immunotherapy and chemotherapy of TRP signature for AML. **(A)** Correlation of TRP score with seven immunomodulators in AML. **(B)** Correlation between expression of four immune checkpoints and TRP scores. **(C)** Box plots of estimated IC50 for six chemotherapeutic agents in the high or low TRP score groups. TRP, transient receptor potential; AML, acute myeloid leukemia.

that compared with the control group, SCHIP1 gene in the si-SCHIP1 group was significantly knocked down after cell transfection (Figures 7E, F). We used CCK8 assay to test the cell viability of each group (Figure 7G). We found that after 24 h, the cell viability of the si-SCHIP1 group was significantly lower than that of the control group ($p < 0.01$). After 48 h, the cell viability in the si-SCHIP1 group was also significantly decreased compared with the control group ($p < 0.01$), while the cell viability in the si-NC group was significantly increased compared with the si-SCHIP1 group ($p < 0.01$).

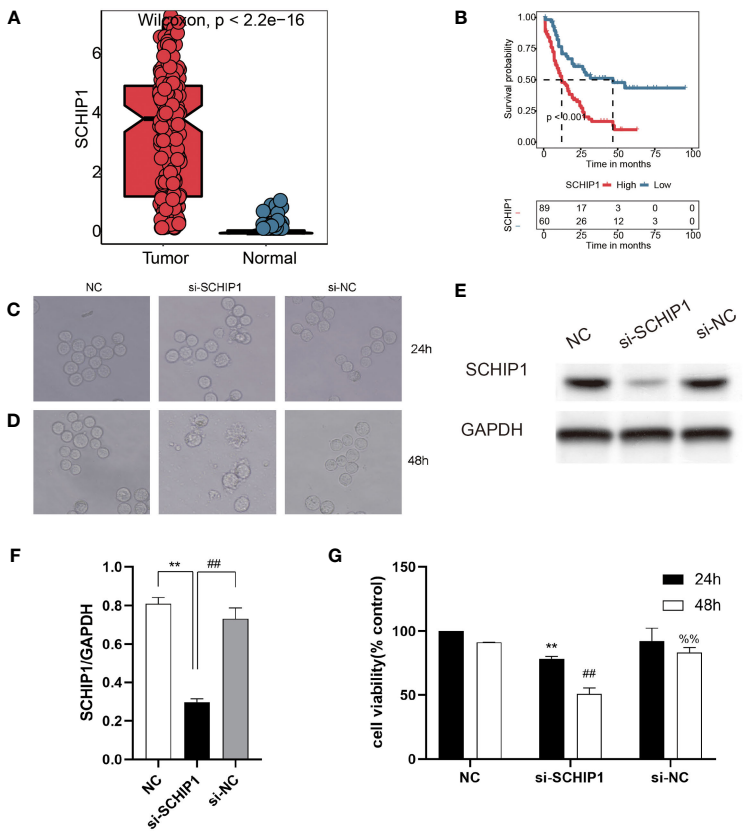


FIGURE 7
In vitro cyto-functional experiments for SCHIP1 of AML. **(A, B)** The expression of SCHIP1 in TCGA between normal and tumor. **(B)** The prognosis of SCHIP1 in TCGA. Cell morphology after transfection at 24 h **(C** or **D)**. NC denotes blank control group, si-SCHIP1 denotes knockdown SCHIP1 group, and si-NC denotes control group. **(E, F)** Western blotting assays verifying the transfection efficiency. **(G)** CCK8 assays comparing the survival rate of different groups of cells. AML, acute myeloid leukemia; TCGA, The Cancer Genome Atlas; CCK8, Cell Counting Kit-8. **,<0.01; si-SCHIP1 vs NC for SCHIP1/GAPDH and 24h cell viability; ##, <0.01; si-NC vs si-SCHIP1 for SCHIP1/GAPDH and 48h cell viability; %%, <0.01; si-NC vs si-SCHIP1 for 48h cell viability.

Discussion

Since the discovery of AML, a great deal of research has been carried out on its etiology, development, and treatment. With the development of technology and in-depth research, many important prognostic factors have been found, such as age, chromosome typing, genotyping, and initial and white blood cell count, and patients are grouped according to these prognostic factors to guide diagnosis and treatment (47). However, due to the limitation of traditional clinicopathological features, the clinical prognosis of patients with AML is still highly heterogeneous. According to the European LeukemiaNet (ELN) risk classification system, about half of patients are classified into the intermediate risk group (48). AML is one of the most common malignant diseases of the circulatory system. Different types of AM L may have different clinical manifestations and prognoses. In conjunction with this change, there is a growing acceptance of early risk stratification for AML to guide further treatment. Clear risk stratification of AML is the prerequisite for subsequent correct diagnosis and treatment. In this study, we established risk stratification based on TRP scores. The survival curves showed significantly worse clinical outcomes for patients with higher TRP scores, while the prognosis of patients with lower TRP scores was better. The AUC values of 1-year (AUC = 0.738), 3-year (AUC = 0.796), and 5-year (AUC = 0.858) survival ROC curves predicted by the TRP scores were all higher than 0.7, suggesting the efficiency of TRP signature in predicting prognosis for AML. Furthermore, the TRP score was an independent prognostic factor for AML patients demonstrated by univariate and multivariate Cox regression analyses. The accuracy of the TRP score we constructed in prognostic prediction was recognized by five external validation datasets. In conclusion, the TRP score system may be a novel and reliable stratification system for AML.

As a rapidly developing interdisciplinary, bioinformatics uses computer science and mathematics to drive the development of

biology. Traditional biological studies and clinical studies are often limited, and the selection of research objectives is often through theoretical speculation or literature support. However, bioinformatics research is more macroscopic. Based on the massive data obtained at the gene level or protein level, high-throughput sequencing technology and other technologies are used to screen out some more potential research targets, which provides possible directions for research. TCGA database is a joint project of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) so as to help researchers better understand cancer and promote related cancer prevention, diagnosis, and treatment progress (4). The GTEx database stores a large number of human normal tissue sequencing samples, which can be used to analyze the genetic differences between tumor samples and normal samples (49). The GEO database is a project of the National Center for Biotechnology Information (NCBI), which stores data mainly from microarray or sequencing data uploaded by various research institutions and individuals. GPL refers to the type of sequencing platform used for sequencing data or gene chip data, and GSE refers to the sequencing data dataset of a series of samples (50). In this study, we carried out a comprehensive bioinformatics analysis based on data from TCGA-LAML dataset in the UCSC Xena platform, GTEx, GEO, and TARGET datasets. We also used a number of algorithms (KEGG, GO, ORA, GSEA, GISTIC, ESTIMATE, TIMER, MCPcounter, ssGSEA, and pRRophetic) to assess functional enrichment pathways, somatic mutations, immune characteristics, and drug sensitivity in AML. Bioinformatics analysis contributed to our results.

In addition to using a large number of bioinformatics tools for analysis, this study also selected a gene, SCHIP1, for the wet experiment, which is also a highlight of this study. SCHIP1 is located at chromosome 3q25 and is a relatively unusual protein initially discovered through interactions with the tumor inhibitor Merlin/NF2 in the mouse brain, and it is a new member of the Hippo pathway (51, 52). SCHIP1 plays different roles in many diseases. SCHIP1 has a variety of functions and plays an important role in the organization of Langhock during early brain development and adulthood, and SCHIP1 is also a cytoplasmic chaperone for cortical cytoskeletal tonic proteins (53). Studies have shown that SCHIP1 plays an important role in proteinuria (54). SCHIP1 also promotes the development and progression of several tumors, including adrenal tumors, acute lymphoblastic leukemia, renal cell carcinoma, and colorectal cancer (55–58). Zhang et al. proposed that IQCJ-Schip1-AS1 could affect the proliferation of colorectal cancer cells through the pathways of cell cycling, DNA replication, and p53 (58). In addition, SCHIP1 is an NF2/Merlin interacting protein in *Drosophila*, and its coiled-coil domain interacts with NF2/Merlin to influence the Hippo pathway (52). After the knockdown of SCHIP1, we found that the apoptosis of AML cells increased and the cell growth rate slowed down, which indicated that SCHIP1 may be a malignant promoter of AML.

There are limitations to the study. First, we constructed and validated the risk prognostic model by retrospectively studying the public database, while more prospective studies are needed for clinical practicability. Second, due to the older AML project data in TCGA database, the lack of clinical information is serious. At the same time, there are few AML data with rich clinical information, and the lack of clinically relevant data is inevitable in this study. Finally, different from solid tumors, which usually detect differential genes by comparing tumor tissues with adjacent tissues, hematological tumors are inevitably affected by other external factors due to the lack of normal bone marrow cells in the samples themselves.

Conclusions

This study determined a risk stratification system based on TRP score through detailed bioinformatics analysis and initially confirmed that SCHIP1 is the oncogene of AML.

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

TD and YS conceived, designed, and supervised the study. JH performed the data analysis and drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants from Research on the Application of Public Welfare Technology of Zhejiang Province (GF22H085719) and the Health Science and Technology Plan of Zhejiang Provincial (2021KY400).

Acknowledgments

We greatly thank the Research on the Application of Public Welfare Technology of Zhejiang Province (GF22H085719) and the Health Science and Technology Plan of Zhejiang Provincial (2021KY400) for supporting the research and work.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1040661/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Kaplan–Meier curves for five external validation data sets.

References

1. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* (2017) 4:424–47. doi: 10.1182/blood-2016-08-733196

2. Freireich EJ, Wiernik PH, Steensma DP. The leukemias: a half-century of discovery. *J Clin Oncol* (2014) 32:3463–9. doi: 10.1200/jco.2014.57.1034

3. Ouzounis CA, Valencia A. Early bioinformatics: the birth of a discipline–a personal view. *Bioinformatics* (2003) 19:2176–90. doi: 10.1093/bioinformatics/btg309

4. Wang Z, Jensen MA, Zenklusen JC. A practical guide to the cancer genome atlas (TCGA). *Methods Mol Biol* (2016) 1418:111–41. doi: 10.1007/978-1-4939-3578-9_6

5. Himmel NJ, Cox DN. Transient receptor potential channels: current perspectives on evolution, structure, function and nomenclature. *Proc Biol Sci* (2020) 287:20201309. doi: 10.1098/rspb.2020.1309

6. Cao E. Structural mechanisms of transient receptor potential ion channels. *J Gen Physiol* (2020) 152:e201811998. doi: 10.1085/jgp.201811998

7. Li H. TRP channel classification. *Adv Exp Med Biol* (2017) 976:1–8. doi: 10.1007/978-94-024-1088-4_1

8. Billeter AT, Hellmann JL, Bhatnagar A, Polk HCJr. Transient receptor potential ion channels: powerful regulators of cell function. *Ann Surg* (2014) 259:229–35. doi: 10.1097/SLA.0b013e3182a6359c

9. Tjong YW, Yao X. Methods for evaluation of vascular endothelial cell function with transient receptor potential (TRP) channel drugs. *Methods Mol Biol* (2018) 1722:195–210. doi: 10.1007/978-1-4939-7553-2_13

10. Everaerts W, Vriens J, Owsianik G, Appendino G, Voets T, De Ridder D, et al. Functional characterization of transient receptor potential channels in mouse urothelial cells. *Am J Physiol Renal Physiol* (2010) 3:F692–701. doi: 10.1152/ajprenal.00599.2009

11. Holzer P. Transient receptor potential (TRP) channels as drug targets for diseases of the digestive system. *Pharmacol Ther* (2011) 131:142–70. doi: 10.1016/j.pharmthera.2011.03.006

12. Stokłosa P, Borgström A, Kappel S, Peinelt C. TRP channels in digestive tract cancers. *Int J Mol Sci* (2020) 21:1877. doi: 10.3390/ijms21051877

13. Julius D. TRP channels and pain. *Annu Rev Cell Dev Biol* (2013) 29:355–84. doi: 10.1146/annurev-cellbio-101011-155833

14. Basso L, Altier C. Transient receptor potential channels in neuropathic pain. *Curr Opin Pharmacol* (2017) 32:9–15. doi: 10.1016/j.coph.2016.10.002

15. Vangeel L, Voets T. Transient receptor potential channels and calcium signaling. *Cold Spring Harb Perspect Biol* (2019) 11:a035048. doi: 10.1101/cshperspect.a035048

16. Chen SJ, Bao L, Keefer K, Shanmughapriya S, Chen L, Lee J, et al. Transient receptor potential ion channel TRPM2 promotes AML proliferation and survival through modulation of mitochondrial function, ROS, and autophagy. *Cell Death Dis* (2020) 4:247. doi: 10.1038/s41419-020-2454-8

17. Wang F, Wu P, Gong S, Chen Y, Gao J, Wang S, et al. Aberrant TRPM4 expression in MLL-rearranged acute myeloid leukemia and its blockade induces cell cycle arrest *via* AKT/GLI1/Cyclin D1 pathway. *Cell Signal* (2020) 72:109643. doi: 10.1016/j.cellsig.2020.109643

18. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data *via* the xena platform. *Nat Biotechnol* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8

19. Human genomics. the genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* (2015) 348:648–60. doi: 10.1126/science.1262110

20. Bolouri H, Farrar JE, Triche T Jr, Ries RE, Lim EL, Alonzo TA, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* (2018) 1:103–12. doi: 10.1038/nm.4439

21. Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG. The affymetrix GeneChip platform: an overview. *Methods Enzymol* (2006) 410:3–28. doi: 10.1016/s0076-6879(06)10001-4

22. Wolber PK, Collins PJ, Lucas AB, De Witte A, Shannon KW. The agilent in situ-synthesized microarray platform. *Methods Enzymol* (2006) 410:28–57. doi: 10.1016/s0076-6879(06)10002-6

23. Katz S, Irizarry RA, Lin X, Tripputi M, Porter MW. A summarization approach for affymetrix GeneChip data using a reference training set from a large, biologically diverse database. *BMC Bioinf* (2006) 7:464. doi: 10.1186/1471-2105-7-464

24. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *Rna* (2020) 26:903–9. doi: 10.1261/rna.074922.120

25. Samanta A, Hughes TET, Moiseenkova-Bell VY. Transient receptor potential (TRP) channels. *Subcell Biochem* (2018) 87:141–65. doi: 10.1007/978-981-10-7757-9_6

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

27. Zhang LP, Ren H, Du YX, Wang CF. Prognostic value of the preoperative fibrinogen-to-albumin ratio in pancreatic ductal adenocarcinoma patients undergoing R0 resection. *World J Gastroenterol* (2020) 26:7382–404. doi: 10.3748/wjg.v26.i46.7382

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

29. van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: basic concepts and methods of cox regression. *Kidney Int* (2008) 74:705–9. doi: 10.1038/ki.2008.294

30. Taylor JM. Random survival forests. *J Thorac Oncol* (2011) 6:1974–5. doi: 10.1097/JTO.0b013e318233d835

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

32. Ranstam J, Cook JA. Kaplan-Meier Curve. *Br J Surg* (2017) 104:442. doi: 10.1002/bjs.10238

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

34. Chabanais J, Labrousse F, Chaunavel A, Germot A, Maftah A. POFUT1 as a promising novel biomarker of colorectal cancer. *Cancers (Basel)* (2018) 10:411. doi: 10.3390/cancers10110411

35. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 4:R41. doi: 10.1186/gb-2011-12-4-r41

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

37. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* (2017) 21:e108–10. doi: 10.1158/0008-5472.Can-17-0307

38. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 1:218. doi: 10.1186/s13059-016-1070-5

39. Xiao B, Liu L, Li A, Xiang C, Wang P, Li H, et al. Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma. *Front Oncol* (2020) 10:607622. doi: 10.3389/fonc.2020.607622

40. Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, Dippel D, et al. Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. *Int J Stroke* (2018) 6:612–32. doi: 10.1177/1747493018778713

41. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 6:417–25. doi: 10.1016/j.cels.2015.12.004

42. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *Omics* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

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

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

45. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol* (2014) 15:R47. doi: 10.1186/gb-2014-15-3-r47

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

47. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* (2016) 23:2209–21. doi: 10.1056/NEJMoa1516192

48. Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* (2010) 3:453–74. doi: 10.1182/blood-2009-07-235358

49. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* (2013) 6:580–5. doi: 10.1038/ng.2653

50. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* (2013) Database issue:D991–5. doi: 10.1093/nar/gks1193

51. Goutebroze L, Brault E, Muchardt C, Camonis J, Thomas G. Cloning and characterization of SCHIP-1, a novel protein interacting specifically with spliced isoforms and naturally occurring mutant NF2 proteins. *Mol Cell Biol* (2000) 20:1699–712. doi: 10.1128/mcb.20.5.1699-1712.2000

52. Chung HL, Choi KW. Schip1, a new upstream regulator of hippo signaling. *Cell Cycle* (2016) 15:2097–8. doi: 10.1080/15384101.2016.1191252

53. Klingler E, Martin PM, Garcia M, Moreau-Fauvarque C, Falk J, Chareyre F, et al. The cytoskeleton-associated protein SCHIP1 is involved in axon guidance, and is required for piriform cortex and anterior commissure development. *Development* (2015) 11:2026–36. doi: 10.1242/dev.119248

54. Perisic L, Rodriguez PQ, Hultenby K, Sun Y, Lal M, Betsholtz C, et al. Schip1 is a novel podocyte foot process protein that mediates actin cytoskeleton rearrangements and forms a complex with Nherf2 and ezrin. *PLoS One* (2015) 3: e0122067. doi: 10.1371/journal.pone.0122067

55. Suarez-Merino B, Hubank M, Revesz T, Harkness W, Hayward R, Thompson D, et al. Microarray analysis of pediatric ependymoma identifies a cluster of 112 candidate genes including four transcripts at 22q12.1-q13.3. *Neuro Oncol* (2005) 1:20–31. doi: 10.1215/s1152851704000596

56. Leung W, Neale G, Behm F, Iyengar R, Finkelstein D, Kastan MB, et al. Deficient innate immunity, thymopoiesis, and gene expression response to radiation in survivors of childhood acute lymphoblastic leukemia. *Cancer Epidemiol* (2010) 3:303–8. doi: 10.1016/j.canep.2010.03.008

57. Fisher KE, Yin-Goen Q, Alexis D, Sirintrapun JS, Harrison W, Benjamin Isett R, et al. Gene expression profiling of clear cell papillary renal cell carcinoma: Comparison with clear cell renal cell carcinoma and papillary renal cell carcinoma. *Mod Pathol* (2014) 2:222–30. doi: 10.1038/modpathol.2013.140

58. Zhang J, Bian Z, Jin G, Liu Y, Li M, Yao S, et al. Long non-coding RNA IQCJ-SCHIP1 antisense RNA 1 is downregulated in colorectal cancer and inhibits cell proliferation. *Ann Transl Med* (2019) 9:198. doi: 10.21037/atm.2019.04.21

10.3389/fimmu.2022.1040661

Glossary

AML	acute myeloid leukemia
GTEx	Genome-Tissue Expression
TRG	TRP-related gene
ssGSEA	Single-Sample Gene Set Enrichment Analysis
TES	TRP enrichment score
DEG	differentially expressed gene
Lasso	least absolute shrinkage and selection operator
ROC	receiver operating characteristic
GISTIC	Genomic Identification of Significant Targets in Cancer
KEGG	Kyoto Encyclopedia of Genes Genomes
GO	Gene Ontology
ORA	over-representation analysis
GSEA	Gene Set Enrichment Analysis
GDSC	Genomics of Drug Sensitivity in Cancer
CR	complete remission
TCGA	The Cancer Genome Atlas
TRP	transient receptor potential
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
RMA	Robust Multichip Average
FPKM	fragments per kilobase of transcript per million fragments mapped
TPM	transcripts per kilobase million
HTS	high TRP score
LTS	low TRP score
OS	overall survival
CNV	copy number variation
ESTIMATE	Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression
TIMER	Tumor Immune Estimation Resource
RPMI	Roswell Park Memorial Institute
FBS	fetal bovine serum
PMSF	phenylmethylsulfonyl fluoride
PVDF	olyvinylidene difluoride
SCHIP1,	Schwannomin-Interacting Protein 1

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fimmu.2022.1061814

OPEN ACCESS

EDITED BY
Ping Zheng,
The University of Melbourne, Australia

REVIEWED BY
Valeria I Segatori,
National University of Quilmes, Argentina
Shiyu Mao,
Tongji University, China

*CORRESPONDENCE
Qiang Zhao
zhaoqiang@tjmuch.com

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

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

RECEIVED 05 October 2022
ACCEPTED 05 December 2022
PUBLISHED 20 December 2022

CITATION
Yang J, Han L, Sha Y, Jin Y, Li Z, Gong B, Li J, Liu Y, Wang Y and Zhao Q (2022) A novel ganglioside-related risk signature can reveal the distinct immune landscape of neuroblastoma and predict the immunotherapeutic response. *Front. Immunol.* 13:1061814. doi: 10.3389/fimmu.2022.1061814

A novel ganglioside-related risk signature can reveal the distinct immune landscape of neuroblastoma and predict the immunotherapeutic response

Jiaxing Yang[1,2,3,4†], Lei Han[2,3,4,5†], Yongliang Sha[1,2,3,4], Yan Jin[1,2,3,4], Zhongyuan Li[1,2,3,4], Baocheng Gong[1,2,3,4], Jie Li[1,2,3,4], Yun Liu[1,2,3,4], Yangyang Wang[1,2,3,4] and Qiang Zhao[1,2,3,4*]

[1]Department of Pediatric Oncology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [2]National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [3]Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [4]Tianjin's Clinical Research Center for Cancer, Tianjin, China, [5]Department of Cancer Molecular Diagnostics Core, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

Introduction: Gangliosides play an essential role in cancer development and progression. However, the involvement of gangliosides in the prognosis and tumor microenvironment (TME) of neuroblastoma is not entirely understood.

Methods: Consensus clustering analysis was performed to identify ganglioside-mediated molecular subtypes. LASSO-Cox analysis was conducted to identify independent prognostic genes, and a novel risk signature was constructed. The risk signature was validated internally and externally. We further explored the independent prognosis value, immune landscape, drug susceptibility, and tumor dedifferentiation of the risk signature. The role of the signature gene B3GALT4 in neuroblastoma was explored *in vitro*.

Results: Seventeen ganglioside-related genes were differentially expressed between INSS stage 4 and other stages, and two ganglioside-related clusters with distinct prognoses were identified. A novel risk signature integrating ten ganglioside-related prognostic genes was established. Across the train set and external validation sets, the risk signature presented high predictive accuracy and discrimination. The risk signature was an independent prognostic factor and constructed a nomogram combining multiple clinical characteristics. In the high-score group, the deficiency in antigen processing and presenting machinery, lack of immune cell infiltration, and escaping NK cells contributed substantially to immune escape. The low-score group was more responsive to immune checkpoint blockade therapy, while the high-score group showed substantial sensitivity to multiple chemotherapeutic drugs. Besides, the risk

score was significantly positively correlated with the stemness index and reduced considerably in all-trans retinoic acid-treated neuroblastoma cell lines, indicating high dedifferentiation in the high-score group. Additionally, neuroblastoma cells with downregulation of B3GALT4 present with increased proliferation, invasion, and metastasis abilities *in vitro*.

Conclusion: The novel ganglioside-related risk signature highlights the role of ganglioside in neuroblastoma prognosis and immune landscape and helps optimize chemotherapy and immunotherapy for neuroblastoma.

KEYWORDS

neuroblastoma, gangliosides, prognosis, immune landscape, dedifferentiation

1 Introduction

Neuroblastoma is the most common extracranial solid tumor in children, accounting for only 6-10% of all pediatric malignancies but 12-15% of pediatric cancer-related deaths (1–3). The prognostic heterogeneity of neuroblastoma has been wildly characterized. The 5-year event-free survival (EFS) rate in the low-intermediate risk neuroblastoma exceeds 80%, while the high-risk group, which accounts for half of the total cases, has a 5-year EFS of only 50% (4). Further survival improvement needs more precise prognostic information on neuroblastoma, and novel genetic and molecular predictive biomarkers are urgently required besides known clinical risk factors.

Gangliosides are glycosphingolipids prevalent on the surface of cells, characterized by one or more sialic acid residues on carbohydrate moieties. It is particularly in specialized membrane domains known as lipid rafts and has a role in cell adhesion and signal transduction (5–7). Gangliosides are implicated in cancer development and progression, including tumor proliferation, invasion, angiogenesis, and metastasis (8). Ganglioside GM3 could decrease the phosphorylation of epidermal growth factor receptors and inhibit the proliferation of bladder cancer (9). Ganglioside GM2 is highly expressed in pancreatic ductal adenocarcinoma and correlated with the activation of TGF-β1 signaling and the promotion of tumor invasion (10). Besides, ganglioside GM3 and GD3 are involved in angiogenesis regulation and metastasis in solid tumors (11, 12). However, the role of gangliosides in neuroblastoma is not entirely understood, and the double-edged sword function of gangliosides in regulating malignant characteristics in neuroectodermal-derived malignancies is a critical trait (8, 13, 14). The monosialogangliosides GM3 and GM1 suppress neuroblastoma, glioma, and astrocytoma proliferation by interacting with different growth factor receptors (13). By contrast, GD3 and GD2 of the b-series gangliosides contribute predominantly to tumor-promoting activities in malignancies arising from neuroectodermal cells (13). GD3 is involved in maintaining and enhancing neural stem cell and glioblastoma self-renewal abilities through EGFR activation (15, 16). Besides, GD3 and GD2 promote proliferation, motility, and invasion in various malignancies, including breast cancer, small cell lung cancer, melanoma, and osteosarcoma (13, 17–19). Importantly, GD2 has become one of the most critical tumor markers and immunotherapeutic targets for neuroblastoma. The anti-GD2 monoclonal antibody immunotherapy has been wildly conducted in neuroblastoma clinical management and presents a considerable improvement in high-risk neuroblastoma prognosis. Given the critical and complicated effects of gangliosides on tumors, it is vital to conduct in-depth studies on the role of gangliosides in neuroblastoma to predict prognosis and inform clinical management.

The remarkable efficacy of anti-GD2 monoclonal antibody suggests the tremendous potential of immunotherapy in neuroblastoma. Nonetheless, immunotherapeutic approaches in neuroblastoma continue to face several obstacles. Neuroblastoma has low immunogenicity due to low mutational load and MHC-I expression, resulting in a lack of lymphocyte infiltration and immunological activity in the tumor microenvironment (TME) (20). Additionally, various immune evasion strategies in TME can obstruct lymphocyte infiltration and activation (20). Thus, a comprehensive understanding of the TME is critical for precisely targeting neuroblastoma with immunotherapy. Interestingly, previous studies have shown that gangliosides are involved in regulating TME. Gangliosides on the surface of tumor cells or shed from cells can suppress cytotoxic T cells or dendritic cells, contributing to tumor immune evasion (7, 20). Besides, gangliosides and IFN-γ could synergistically inhibit dendritic cell activity, promoting immune suppression in the TME (21). Therefore, elucidating the role of gangliosides in TME in neuroblastoma could facilitate the understanding of tumor progression and the optimization of immune therapies.

While the involvement of gangliosides in neuroblastoma prognosis and TME remains to be explored, the bioinformatic

analysis provides us with a new direction. In the present study, samples in the GSE49710 dataset were clustered based on ganglioside-related gene expression, and a risk signature was constructed to predict neuroblastoma prognosis. Additionally, we investigated the immune landscapes and escape strategies in ganglioside-related risk groups. Immunotherapy response and chemotherapeutic drug sensitivity were further explored in the high-score and low-score groups. Our findings constructed an accurate and effective prognostic signature for neuroblastoma and may help inform the treatment strategy for neuroblastoma.

2 Materials and methods

2.1 Data acquisition and preprocessing

The workflow of this study was presented in Figure S1. Expression data and corresponding clinical information were obtained from the Gene Expression Omnibus (GEO) GSE49710 (n = 498) (22) and ArrayExpress E-MTAB-8248 (n = 223) (23). The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) neuroblastoma gene-expression profile and clinical data (n = 150) were acquired from the UCSC Xena database (http://xena.ucsc.edu/). The GSE49710 cohort was used to construct the risk signature, with the E-MTAB-8248 and TARGET datasets serving as external validation. The clinical baseline characteristics of three data sets were summarized in Table 1. Expression data were normalized, and log2 transformed. The expression profile and corresponding immunotherapy information for the GSE78220 cohort were retrieved from the GEO database. The expression data of neuroblastoma cell lines treated with Dimethyl sulfoxide (DMSO) or all-trans retinoic acid (ATRA) was obtained from GSE155000 in the GEO database. Thirty-four genes associated with gangliosides were identified by the Molecular Signatures Database (MSigDB; https://www.gsea-msigdb.org/gsea/msigdb) and previously published literature (Table S1) (14).

2.2 Consensus clustering analysis based ganglioside-related genes

A total of thirty-three ganglioside-related genes were extracted from the GSE49710 dataset. Given the significant

TABLE 1 Clinical baseline characteristics of three datasets in present study.

Characteristic	GSE49710 (N = 498)	TARGET (N = 150)	E-MTAB-8248 (N = 223)
Gender			
Male	287 (57.6%)	88 (58.7%)	N/A
Female	211(42.4%)	62 (41.3%)	N/A
Stage			
1	121 (24.3%)	0 (0.0%)	29 (13.0%)
2	78 (15.7%)	1 (0.7%)	39 (17.5%)
3	63 (12.7%)	9 (6.0%)	36 (16.1%)
4	183 (36.7%)	120 (80.0%)	89 (39.9%)
4S	53 (10.6%)	20 (13.3%)	30 (13.5%)
Age			
<18 months	305 (61.2%)	32 (21.3%)	104 (46.6%)
≥18 months	193 (38.8%)	118 (78.7%)	119 (53.4%)
MYCN Status			
Normal	401 (80.5%)	119 (79.3%)	176 (78.9%)
Amplified	92 (18.5%)	30 (20.0%)	46 (20.6%)
Unknow	5 (1.0%)	1 (0.7%)	1 (0.5%)
Risk group			
Non–high-risk	322 (64.7%)	31 (20.7%)	133 (59.6%)
High-risk	176 (35.3%)	119 (79.3%)	90 (40.4%)

N/A, not available.

difference between International Neuroblastoma Staging System (INSS) stage 4 and other stages (24), seventeen ganglioside-related genes were identified as differentially expressed genes (DEGs) between these two groups. Unsupervised consensus clustering analysis based on ganglioside-related DEGs was performed on GSE49710 and TARGET datasets using the "Consensus Cluster Plus" R package and the "k-means" method to discern diverse expression patterns, with the repetition number set to 1000 to ensure stability (25). The t-distributed stochastic neighbor embedding (t-SNE) analysis was conducted using the "Rtsne" package to investigate the distribution between distinct clusters.

2.3 Identification of differentially expressed genes between clusters and functional enrichment analysis

The analysis of differentially expressed genes between clusters was conducted using the "limma" R package according to the specified criteria ($|log2FC| \geq 1$ and adjusted p-value < 0.05). The clusterProfiler R package was used to perform functional enrichment analysis on these DEGs, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses (26).

2.4 Construction and validation of the ganglioside-related risk signature

Univariate Cox regression analysis using the "survival" R package was performed to screen for potential prognostic genes within ganglioside-related DEGs in the GSE49710 dataset. The least absolute shrinkage and selection operator (LASSO)-penalized Cox regression analysis was then conducted to identify potential signature genes. Finally, the regression coefficients of ten signature genes were determined using multivariate Cox regression. The risk score was calculated by multiplying the expression value of each signature gene by its corresponding regression coefficient.

According to the median risk score, samples were classified as low-score or high-score groups. The scatter dot plots were generated to visualize the association between risk score and survival status. Survival analysis of overall survival (OS) probability was performed to evaluate the risk signature prognostic value using "survival" and "survminer" R packages. Receiver operating characteristic (ROC) curve analysis was performed by the "timeROC" R package to assess the specificity and sensitivity of the risk signature. The unique capability of risk signature was evaluated using principal component analysis (PCA) and the R package "ggplot2." The ganglioside-related risk signature validation was carried out in GSE49710, TARGET, E-MTAB-8248, and Tianjin cohorts.

2.5 Independent prognostic analysis and construction of a nomogram

The "survival" R package was used to conduct univariate and multivariate Cox regression analysis on datasets to determine the risk signature's predictive significance in the context of recognized prognostic indicators. Furthermore, a nomogram was constructed comprising the risk signature and several established prognostic factors by the "rms" R package to predicate 3-year and 5-year OS. Calibration plots were drawn to assess the accuracy of the nomogram.

2.6 Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was conducted to investigate the biological function differences in low-score and high-score groups using the R packages "clusterProfiler" and "enrichplot." The reference gene set ("c2.cp.kegg.v7.5.1.symbols.gmt") was acquired from MSigDB (http://www.gsea-msigdb.org/gsea/downloads.jsp).

2.7 Immune landscape of the risk signature

The low immunogenicity of neuroblastoma could be attributed to a low mutational burden and impairment in the antigen processing and presenting machinery (APM) (20). APM scores (APS) were derived in the previous study by gene set variation analysis (GSVA) based on 18 APM-related genes to represent antigen processing and presentation efficiency (Table S2) (27). As an integral element of APM, MHC-I mediates the recognition and lysis of neuroblastoma cells by cytotoxic T lymphocytes (CTL). To estimate the quantity of MHC-I activity, we obtained the gene set associated with the MHC-I protein complex pathway from the MSigDB (Table S3). Subsequently, the enrichment score (ES) was calculated in the single-sample gene set enrichment analysis (ssGSEA) using the "gsva" R package to reflect MHC-I activity, which was defined as the MHC score in this study.

The immune infiltration landscape was investigated by calculating infiltrating scores of 30 different types of TME cells by ssGSEA (Table S4) (28, 29). Besides, immune, stromal, and ESTIMATE scores were calculated by the "ESTIMATE" algorithm to reveal the distinct immune microenvironments between risk groups (30).

Immune function-related gene sets, including interferon receptor and natural killer cell cytotoxicity, were gathered from the Immunology Database and Analysis Portal (ImmPort, http://www.immport.org) database (Table S5) (31). Immune function activities were measured by ssGSEA and compared between risk groups.

2.8 Development of the stemness index

The messenger ribonucleic acid stemness index (mRNAsi) was developed using the one-class logistic regression machine learning algorithm (OCLR) based on pluripotent stem cell samples from the Progenitor Cell Biology Consortium dataset (https://www.synapse.org/, accessed on 16 January 2022). The mRNAsi had been widely used for tumor dedifferentiation and stemness prediction (32–34). The workflow was available on https://bioinformaticsfmrp.github.io/PanCanStem_Web/. The mRNAsi value was normalized to 0-1, with increased mRNAsi indicating a greater degree of dedifferentiation. The stemness index model was constructed in this work, and the mRNAsi was estimated in the GSE49710 dataset. Considering the effect of tumor purity on mRNAsi, we corrected for mRNAsi using tumor purity generated by the ESTIMATE algorithm, and the corrected mRNAsi (c_mRNAsi) was calculated as mRNAsi/tumor purity.

2.9 Drug sensitivity analysis

The half-maximal inhibitory concentration (IC50) of commonly used chemotherapeutic agents was predicted by the "pRRophetic" R package to characterize chemosensitivity in high-score and low-score groups (35).

2.10 Immunohistochemistry

A total of forty-six paraffin-embedded neuroblastoma specimens were collected at Tianjin Medical University Cancer Institute and Hospital. This study complied with the Declaration of Helsinki and was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (E20210027). Sections were deparaffinized with xylene for 30 mins and gradient concentrations of alcohol followed by rehydration. The heat-induced epitope retrieval was conducted by the Tris/EDTA buffer (Solarbio, Beijing, China), pH of 9.0, at 120°C for 3 mins, and the sections were immersed in 3% hydrogen peroxide for 30 min and incubated with the primary ST8SIA2 antibody (dilution 1:100; Rabbit polyclonal, 19736-1-AP; Proteintech), B3GALT4 antibody (dilution 1:100; Rabbit monoclonal, ab169759; abcam), and CD8 antibody (dilution 1:4000; Mouse monoclonal, 66868-1-Ig; Proteintech) at 4°C overnight. After washing with PBS and incubation with the secondary antibody (PV-6001; ZSGB-BIO; Beijing, China) at 37°C for 1 hour, the antigens were detected using DAB chromogen and counterstained with hematoxylin for 1 min. The immunoreactivity score (IRS) was generated for semi-quantitative expression and scored by two independent, experienced pathologists blinded to the clinical information. The inconsistencies were discussed to reach a unified result. The IRS considered staining intensity and the percentage of positive tumor cells. The staining intensity was assessed in four grades, including negative staining (0 points), weak staining (1 point), moderate staining (2 points), and strong staining (3 points). The percentage of positive tumor cells in the section was divided into five grades, including 0-5% (0 point), 6-25% (1 point), 26-50% (2 points), 51-75% (3 points) and 76-100% (4 points). The sample IRS was calculated by multiplying scores of the staining intensity and percentage of positive tumor cells. The precents of $CD8^+$ T cells were quantified as the proportion of CD8A-positive cells in all cells on 200× photographs.

2.11 Cell lines and cell culture

Neuroblastoma cell lines 9464D and 975A2 were gifted from Dr. Rimas Orentas at Seattle Children's Research Institute. Neuroblastoma cells were maintained in the high-glucose DMEM medium (Gibco) containing 10% FBS (BI) and 1% penicillin/streptomycin (Gibco). The cells were cultured in a humidified incubator at 37°C in a 5% CO2 atmosphere.

2.12 Cell transfection

The small interfering RNA (siRNA) targeting B3GALT4, and negative control siRNA (si-NC) were purchased from General Biol (Anhui, China). The manufacturer's instruction was followed for cell transfection. Neuroblastoma cells were transfected with siRNA using the transfection reagent Lipofectamine®2000 (Invitrogen). The transfected cells were collected for further experiments after 24h. Quantitative realtime PCR assay and western blot analysis were performed to verify the knockdown efficiency.

2.13 Western blot analysis

Cells were lysed in RIPA lysis buffer (Solarbio) for protein extraction, and the protein concentrations were evaluated by the BCA method. The proteins were separated by 10% SDS-polyacrylamide gel electrophoresis (PAGE) and transferred to PVDF membranes. After incubation in 5% skimmed milk for 1 h at room temperature, the membranes were incubated overnight at 4°C with primary antibodies against B3GALT4 (dilution 1:1000; Rabbit monoclonal, ab169759; abcam). After incubation with the secondary antibody and wash with TBS-T three times, the band images were visualized by the enhanced chemiluminescence kit.

2.14 Real-time quantitative PCR

The total RNA was extracted from cells by Trizol reagent (Invitrogen) and converted to cDNA using the PrimScript RT

Master Mix (Takara). cDNA amplification was carried out by SYBR Green PCR Kit (Takara) according to the program: 5 seconds at 95°C for the denaturation, 34 seconds at 60°C for annealing, followed by 30 seconds at 72°C for extension, and forty cycles were completed in total. The primer sequences were designed as follows: B3GALT4: F: 5'-AACGCCATTCGG GCATCTT-3', R: 5'-GTTGCGGTAGGAATCCTGGAA-3'; GA PDH: F: 5'-ACCCTTAAGAGGGATGCTGC-3', R: 5'-CCCAA TACGGCCAAATCCGT-3'. The 2−ΔΔCt value was employed to quantify the relative gene expression levels with GAPDH as the endogenous control.

2.15 Cell proliferation and colony forming assay

Cell viability was measured by the Cell Counting Kit-8 (CCK8) assay. Neuroblastoma cells were plated into 96-well plates for 24h, 48h, and 72h, followed by adding 100μ CCK8 solution (Solarbio) and incubating for 2 hours at 37°C. The absorbance of each well was measured at a wavelength of 450 nm (OD450) with a microplate reader. Cells were planted and cultured for 2 weeks in each well of a 6-well plate for cell colony formation assay. The colonies were fixed with 4% paraformaldehyde for 15 minutes, stained with 0.1% crystal violet for 20 minutes at room temperature, and quantified by ImageJ software after being photographed.

2.16 Cell invasion and migration assay

The transwell assay was performed to evaluate cell invasion and migration ability. Eight-micrometer pore-size transwell filters (Corning) were put in a 24-well plate for the migration assay, while the upper chambers plated with matrigel (BD Biosciences) for the invasion assay. Cells in 200 μl FBS-free medium were seeded onto each upper chamber, and the lower chamber was added with 600 μl medium with 10% FBS. After being cultured for 24 hours at 37°C, these invasive and metastatic cells in the lower side of the filter were fixed by 4% paraformaldehyde, stained with 0.1% crystal violet solution, and photographed.

2.17 Sample collection for RNA sequencing in the Tianjin cohort

Twenty-six neuroblastoma biopsies and corresponding clinical information were collected at Tianjin Medical University Cancer Institute and Hospital. This study complied with the Declaration of Helsinki and was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (E20210027).

2.18 RNA quantification and qualification

RNA quantification and qualification were performed according to the following steps. RNA purity and concentration were generated by NanoDrop 2000, and RNA integrity and quantity were quantified by the Agilent 2100/ 4200 system.

2.19 Library construction

The messenger RNA was extracted from total RNA and fragmented into 300-350 bp fragments. The reverse transcription was conducted using fragmented RNA and dNTPs (dATP, dTTP, dCTP, and dGTP) to synthesize the first strand cDNA, followed by the synthesis of the second strand cDNA. After the double-strand cDNA remaining overhangs were converted into blunt ends by exonuclease/polymerase, 3' ends of DNA fragments were adenylated, and sequencing adaptors were ligated to the cDNA. Subsequently, the library fragments were purified. The PCR was used to amplify the template, and the product was purified to form the final library.

2.20 Sequencing and quality control of the raw data

After library preparation and sample pooling, Illumina sequencing was performed on the samples. Raw data in the formation of FASTQ were processed through in-house perl scripts. Clean data were formed by reads without low-quality or adapter and ploy-N. The clean data's Q20, Q30, and GC content were assessed. The clean reads were mapped to the silva database to eliminate the rRNA.

2.21 Reads mapping and quantification of gene expression level

Paired-end clean reads were aligned to the reference genome (hg19) using Hisat2 (36). Featurecount was used to count the reads numbers mapped to each gene (37).

2.22 Statistical analysis

All statistical analysis was performed through R software (version 4.1.2). Survival curves were generated by the Kaplan-Meier method and log-rank test for statistical tests. Spearman rank correlation was used to analyze the correlations between continuous variables. The Mann-Whitney Wilcoxon or Kruskal-Wallis test was used to compare continuous variables between groups. The Pearson chi-square test was used to compare

categorical variables across groups. Two-sided $p < 0.05$ was considered statistically significant.

3 Results

3.1 Consensus clustering analysis of ganglioside-related genes identified two clusters of neuroblastoma with different outcomes

The International Neuroblastoma Staging System (INSS) had been wildly used in neuroblastoma clinical management, and INSS stage 4 was an independent risk factor for neuroblastoma (4). Besides prognosis, there were also significant differences in biological characteristics between stage 4 and other stages. To screen crucial ganglioside-related genes in neuroblastoma, we identified seventeen ganglioside-related DEGs between the INSS stage 4 (high stage) and other stages (low stage) (Figure 1A). Consensus clustering analysis was performed to classify patients with distinct ganglioside-mediated patterns in GSE49710 based on seventeen ganglioside-related DEGs expression. k = 2 was selected as the ideal option for cluster construction, and 498 samples were allocated to clusters A and B, with 222 and 276 samples, respectively (Figure 1B, Figure S2; Table S6). As shown in Figure 1C, the two clusters could be clearly distinguished in the t-SNE analysis. Survival analysis revealed the significant survival advantage of cluster B over cluster A (P=0.005, Figure 1D). Additionally, the TARGET dataset was used to verify the consensus clustering result. Two distinct clusters with significantly different prognoses were identified in the TARGET dataset (Figure S3A-M; Table S7), suggesting the stability of the clustering result.

The analysis of the differentially expressed genes was performed between two clusters in the GSE49710 dataset to explore the further difference in the biological function in identified clusters. One hundred and eight DEGs were finally identified according to $|logFC| > 1$ and adjusted p-value < 0.05 (Table S8). Interestingly, GO functional enrichment analysis showed these DEGs were significantly enriched in neural crest cell development of biological processes, the postsynaptic membrane of cellular components, and signaling receptor activator activity of molecular function (Figure S4). In the KEGG functional enrichment analysis, DEGs were enriched in the neuroactive ligand receptor interaction and cAMP signaling pathways (Figure S4). These results suggested a potential role for gangliosides in neuroblastoma differentiation. Consistent with the functional enrichment in receptor ligand activity, the role of gangliosides in cancer cell signaling had been widely characterized (18).

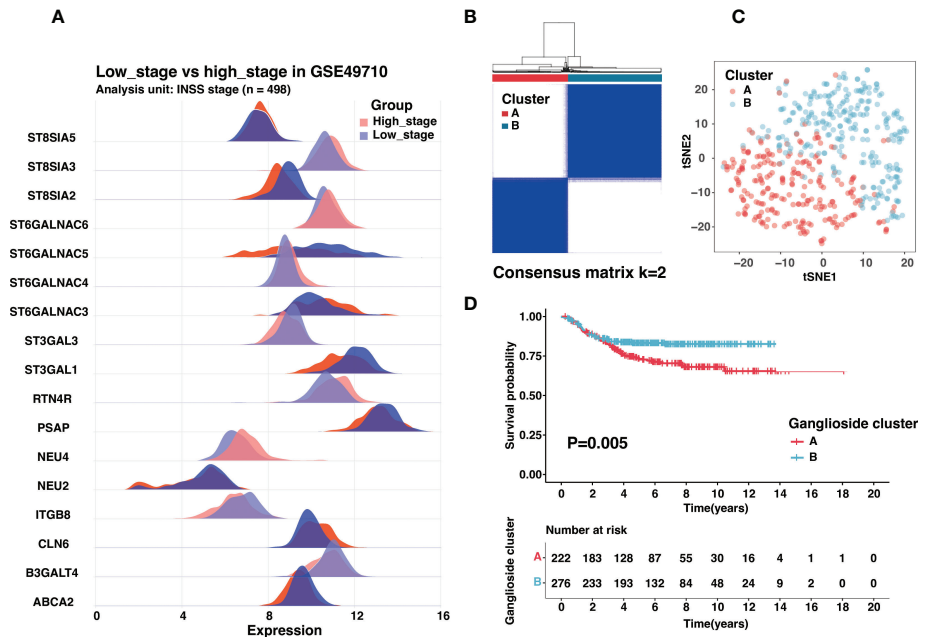


FIGURE 1
Identification of ganglioside-related clusters in the GSE49710 dataset. **(A)** The expression of ganglioside-related genes that were differentially expressed between samples with high stage (INSS stage 4) and low stage (other stages). **(B)** Identification of two ganglioside-related clusters according to the consensus clustering matrix (k = 2) in the GSE49710 cohort. **(C)** The t-SNE analysis revealed a clear distinction between the two clusters. **(D)** Kaplan–Meier curves of overall survival (OS) in the GSE49710 cohort between different clusters.

3.2 Development and validation of the novel ganglioside-related risk signature

?>Given the prognostic significance of different clusters, a ganglioside-related risk signature based on two clusters was developed in the GSE49710 cohort to predict individual prognosis accurately. Firstly, the univariate Cox regression analysis of ganglioside-related DEGs was conducted, and sixteen potential prognostic genes were recognized (Figure 2A; Table S9). Then, Lasso-penalized Cox regression analysis was performed, followed by multivariate Cox regression to identify ten independent prognostic genes and corresponding regression coefficients. These genes included ABCA2, B3GALT4, NEU4, ST3GAL1, ST3GAL3, ST6GALNAC4, ST6GALNAC5, ST6GALNAC6, ST8SIA2, and ST8SIA3 (Figures 2B, C). These ten genes were used to establish the ganglioside-related risk signature, and the risk scores were derived from the expression values of each signature gene and its corresponding regression coefficient. As shown in Figure 2D, cluster A presenting with a poor prognosis, received significantly higher risk scores than cluster B. Samples were then divided into high-score and low-score groups based on the median risk score. The expression of ganglioside-related genes was substantially different between the two groups, with eight elevated genes in the high-score group (Figure 2E).

Neuroblastoma with high expression of the b-series gangliosides, including GD1b and GT1b, typically presented an excellent prognosis (14). Interestingly, B3GALT4 catalyzed the first step in converting GD2 to more complex b-series gangliosides. Considering B3GALT4 serves as the connecting link of ganglioside in neuroblastoma, we further performed immunohistochemistry to validate the expression of B3GALT4 in clinical neuroblastoma specimens. Consistent with the results in the GSE49710 dataset, the immunohistochemistry analysis showed a significantly low expression of B3GALT4 in samples with a high stage (Figures 3A, B). ST8SIA2 was involved in the developmental regulation of polysialic acid and modulated neuroblastoma adhesion and metastasis. Interestingly, mRNA levels of ST8SIA2 were highest in stages 1 and 4s neuroblastoma (38). Consistent with this result, our study also showed that the expression level of ST8SIA2 was significantly upregulated in low-stage samples (Figures 3C, D).

Besides, these signature genes were interconnected (Figure 4A), and the corresponding regression coefficients were presented in Figure 4B and Table S10. The risk signature was validated in the GSE49710, E-MTAB-8248, and TARGET cohorts. Samples in the GSE49710 cohort were divided into two groups based on the median risk score, and an increase in risk score was associated with a decrease in survival time (Figure 4C). As expected, those samples classified as the high-score group had a considerably poorer prognosis than samples classified as the low-score group (Figure 4D). Additionally, ROC analysis demonstrated that the area under the curve (AUC) values of

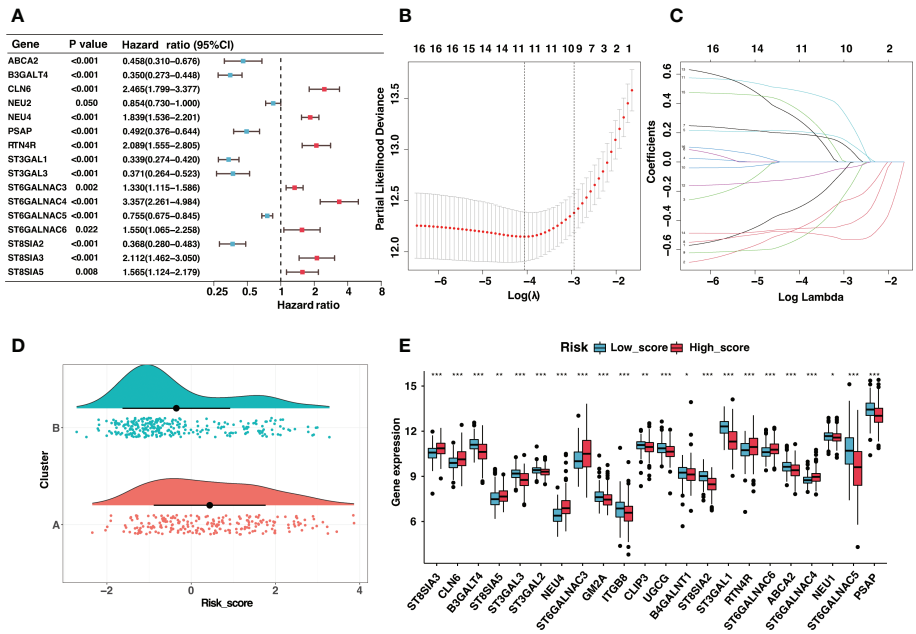


FIGURE 2
Construction of the ganglioside-related risk signature. **(A)** The forest map of ganglioside-related differentially expressed genes (DEGs) in GSE49710 cohort generated by univariate Cox analysis (P < 0.05). **(B,C)** LASSO Cox regression analysis of ganglioside-related prognostic DEGs. **(D)** The distribution of risk scores between two ganglioside-related clusters. **(E)** The expression of ganglioside-related genes between the low-score and high-score group. (*: P < 0.05, **: P < 0.01, ***: P < 0.001).

the risk signature for 3- and 5-year OS prediction were 0.891 and 0.902, respectively (Figure 4E). The risk signature performed better in predicting 3-year OS than established clinical prognostic factors (Figure 4F).

Moreover, validation was conducted on the E-MTAB-8248 and TARGET cohorts. Consistent with the train set, the increase in risk score was accompanied by a decrease in survival time in E-MTAB-8248 (Figure S5A) and TARGET (Figure S5E) cohorts. The risk signature could accurately predict 3-year OS in E-MTAB-8248 (Figure S5B) and TARGET (Figure S5F) cohorts, with AUC values of 0.807 and 0.667, better than clinical characteristics (Figure S5C and S5G). Importantly, the risk signature had a vital prognosis predictive value in both cohorts, exhibiting significantly poor prognosis in the high-score group (Figure S5D and S5H). The PCA analysis suggested that the high-score group could also be separated from the low-score group in all cohorts (Figure S5I-K). In addition, even in children in COG high-risk group or older than 18 months, the prognosis of samples with a high score was significantly worse than that of low-score samples (Figures S6A-D).

Additionally, we validated the reliability of risk signature with tissue samples in our center. A total of twenty-six samples with RNA-seq data were included in this study, and the corresponding clinical characteristics were shown in Tables S11–12. As expected, high-score samples in the Tianjin cohort presented low survival time (Figure 5A). The risk signature performed excellent sensitivity and specificity in predicting 3-year and 5-year OS (Figure 5B). The overall survival and event-free survival were significantly worse for the sample in the high-score group compared with low-score samples ($P < 0.05$, Figures 5C, D). In conclusion, the risk signature could effectively predict neuroblastoma prognosis after thorough evaluation and validation.

3.3 Clinical correlation analysis, independent prognosis analysis, and construction of a nomogram

There had been several established prognostic factors in neuroblastoma, including age, INSS stage, MYCN status, and the clinical risk classification system. We explored the correlation between the risk signature and these prognostic factors. As illustrated in Figure 6A, the risk score was significantly associated with age, MYCN amplification, clinical risk group, stage, and progression. Patients with unfavorable clinical characteristics present high scores (Figures 6B-F). Similar results were observed in the Tianjin cohort, and there were significant high scores in samples with advanced stage and unfavorable histology (Figure S7).

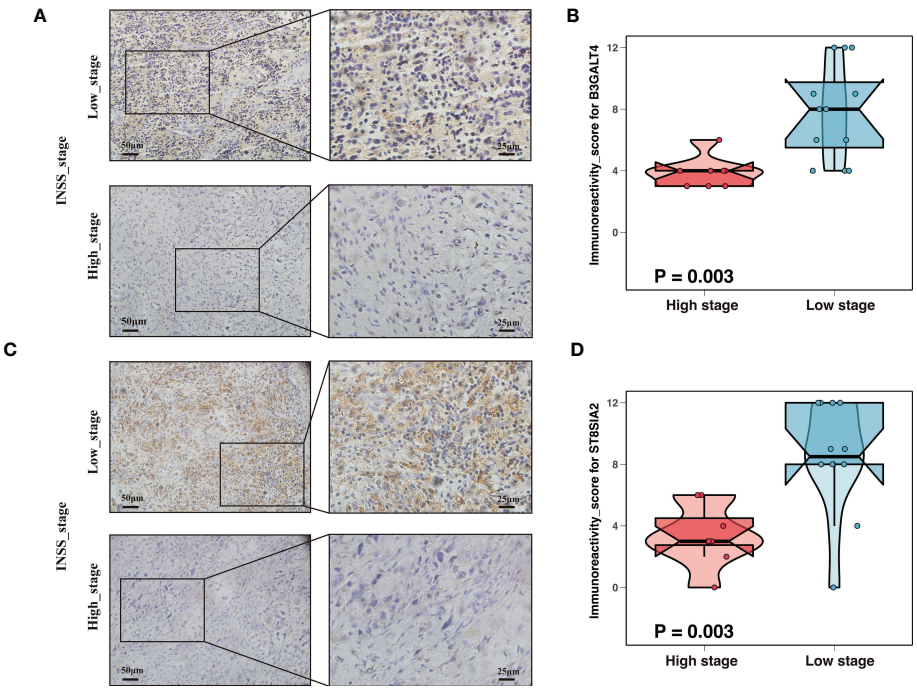


FIGURE 3
Expression validation of ganglioside-related genes by immunohistochemistry (IHC). **(A)** Representative IHC images showing the expression of B3GALT4 in different stages. Magnification, ×200, ×400. **(B)** The comparison of B3GALT4 immunoreactivity score (IRS) between samples with high stage and low stage. **(C)** Representative IHC images showing the expression of ST8SIA2 in different stages. Magnification, ×200, ×400. **(D)** The ST8SIA2 immunoreactivity score (IRS) comparison between samples with high stage and low stage.

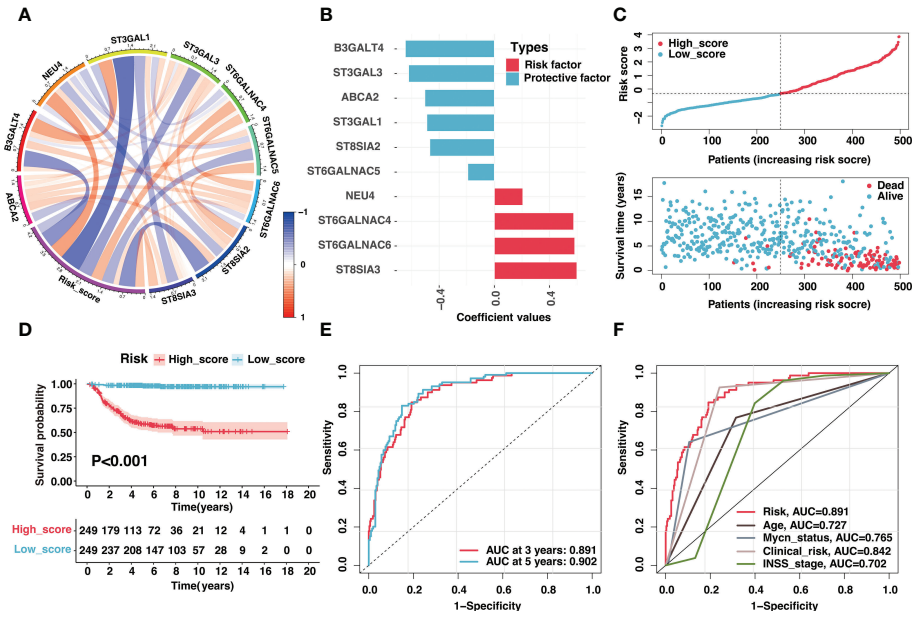


FIGURE 4
Validation of the ganglioside-related risk signature in the GSE49710 dataset. **(A)** The correlation coefficient between risk score and signature genes. **(B)** Coefficient values of ten signature genes. **(C)** The distribution of the risk score (top) and survival time (bottom) in the GSE49710 cohort. **(D)** Kaplan–Meier curves of overall survival (OS) in GSE49710 cohort between different risk groups. **(E)** The receiver operating characteristic (ROC) curves of the risk signature for 3-year and 5-year OS prediction in the GSE49710 cohort. **(F)** The ROC curves of clinical prognostic factors for 3-year OS prediction in the GSE49710 cohort.

In light of the correlation between the risk signature and clinical characteristics, we investigated the independent prognostic significance of the risk signature. After univariate and multivariate cox regression analysis, the risk signature was identified as an independent prognostic factor in the GSE4910 (Figures 7A, B), E-MTAB-8248 (Figure S8A), and TARGET (Figure S8B) cohorts, respectively. To optimize the clinical utilization in individual prognosis prediction, we incorporated the risk signature and several clinical risk factors to construct a nomogram in the GSE49710 cohort (Figure 7C). The nomogram could assign a score to each prognostic factor and predict 3-year and 5-year OS based on the sum of scores in each sample (Figure 7C). The calibration curves were plotted to evaluate the accuracy of the nomogram, and the nomogram prediction curves were quite close to standard curves in GSE49710 (Figure 7D), E-MTAB-8248 (Figure 7E), and TARGET (Figure S8C) cohorts, suggesting an excellent accuracy for prognosis prediction in all datasets. The nomogram incorporated the risk score and multiple established prognostic factors and could precisely predict neuroblastoma prognosis.

3.4 Immune landscape of the ganglioside-related risk signature

The gene set enrichment analysis (GSEA) was conducted to elucidate the biological functions behind the variations in prognosis between low-score and high-score groups. As shown in Figure S9A, the high-score group was significantly enriched in tumorigenic pathways, including cell cycle, DNA replication, homologous recombination, ribosome, and spliceosome. Interestingly, the low-score group was enriched in antigen processing and presentation and cell adhesion pathways, implying underlying immune landscape differences between the two groups (Figure S9B).

Impairment of the antigen-presenting machinery (APM) contributed heavily to the low immunogenicity of neuroblastoma. It was widely accepted that antigen presentation through MHC-I molecules did not function in neuroblastoma due to low expression levels (20, 39). These characteristics made neuroblastoma cells almost undetectable to CD8 T cells. Interestingly, the low-score group was enriched in the antigen processing and presentation pathway, implying a potential role of ganglioside in APM of neuroblastoma.

We introduced the APM scores (APS) developed in previous research (detailed in "Materials and Methods") as a proxy for antigen processing and presentation efficiency. As seen in Figure 8A, the APS was significantly and adversely linked with the risk score, showing that APM in the high-score group was suppressed (Figure 8A). Considering MHC-I molecules as a critical component of APM, the MHC score (described in "Materials and Methods") was developed using ssGSEA to

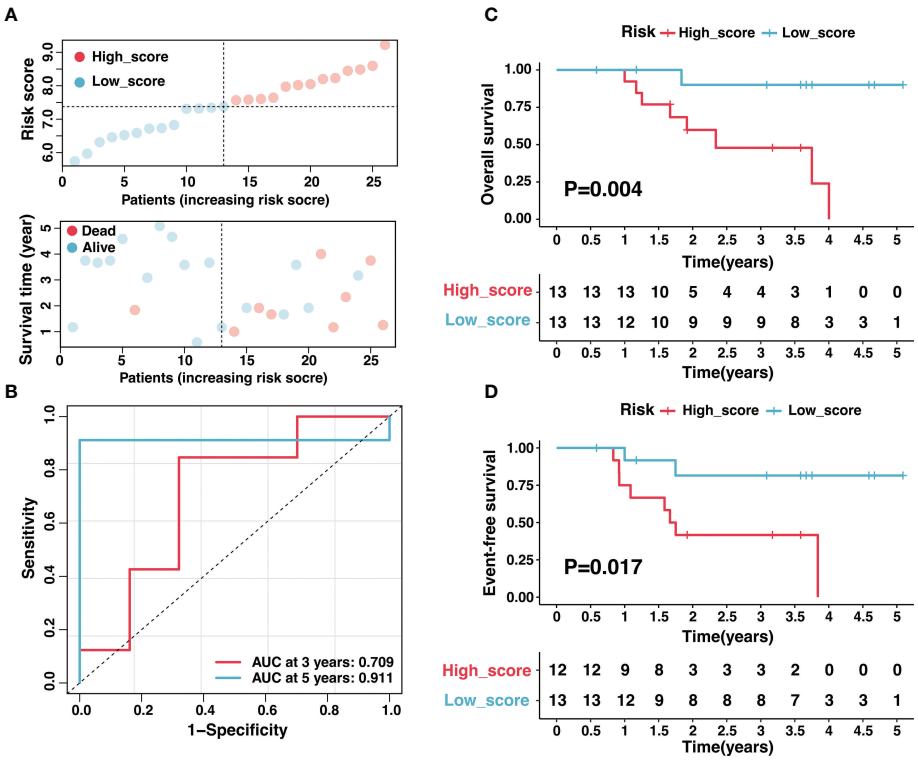


FIGURE 5
Validation of the ganglioside-related risk signature in the Tianjin cohort. **(A)** The distribution of the risk score (top) and survival time (bottom) in the Tianjin cohort. **(B)** The ROC curves of the risk signature for 3-year and 5-year OS prediction in the Tianjin cohort. **(C, D)** Kaplan–Meier curves of overall survival **(C)** and event-free survival **(D)** in Tianjin cohort between different risk groups.

assess MHC-I activity. As expected, there was a strong negative association between risk score and MHC score as well as activated CD8 T cell infiltration (Figures 8B, C), and increased MHC-I activity was associated with high activated CD8 T cell infiltration (Figure 8D). Additionally, consistent with the concept that IFNγ could increase MHC-I expression in neuroblastoma (40), the MHC score was substantially and positively correlated with the activity of interferon receptors (Figure 8E). In general, APM impairment contributed significantly to immune escape in the high-score group, and the low activity of MHC-I may be an important cause.

Moreover, the absence of leukocytes and the presence of immunosuppressive myeloid and stromal cells were also efficient strategies for tumor immune evasion in neuroblastoma. Interestingly, ganglioside soluble GD2 could reduce T cell proliferation, suggesting the role of ganglioside in tumor immune evasion of neuroblastoma. The ESTIMATE method was used to determine the infiltration levels of immune cells and stromal cells. The high-score group presented significantly lower stromal and immune scores than the low-score group (Figure 8F). As illustrated in Figure 8G, most immune infiltrating cells were infiltrated at low levels in the high-score group, demonstrating an immune escape mechanism resulting from the lack of immune cells in the high-score group. In summary, the immune escape strategy in the high-score group was partially attributed to the lack of immune cell infiltration.

Tumor-infiltrating lymphocytes were essential constituents of the tumor immune microenvironment in neuroblastoma. Recent studies revealed reduced $CD8^+$ T lymphocyte infiltration in high-risk and advanced-stage neuroblastoma (20). Children with a rising rate of $CD8^+$ T lymphocytes had a better prognosis, highlighting that strengthening $CD8^+$ T-cell responses would be a promising therapy opportunity (41). Consistent with previous studies, our results showed that samples with high $CD8^+$ T lymphocyte infiltration presented an excellent prognosis in the GSE49710 cohort (Figure 9A). Furthermore, the risk signature was a potent indicator of $CD8^+$ T lymphocyte infiltration. Compared to samples with a high score, samples with a low score displayed a significantly greater infiltration of $CD8^+$ T-cells (Figure 9B). The risk score was adversely and significantly associated with $CD8^+$ T-cell infiltration and CD8A expression (Figure 9C). We further performed immunohistochemistry on tissue samples from the Tianjin cohort to confirm this finding. As indicated in Figures 9D, E, there was a considerable decrease in $CD8^+$ T-cell infiltration in the high-score group compared to the low-score group (Figures 9D, E). In summary, the risk signature

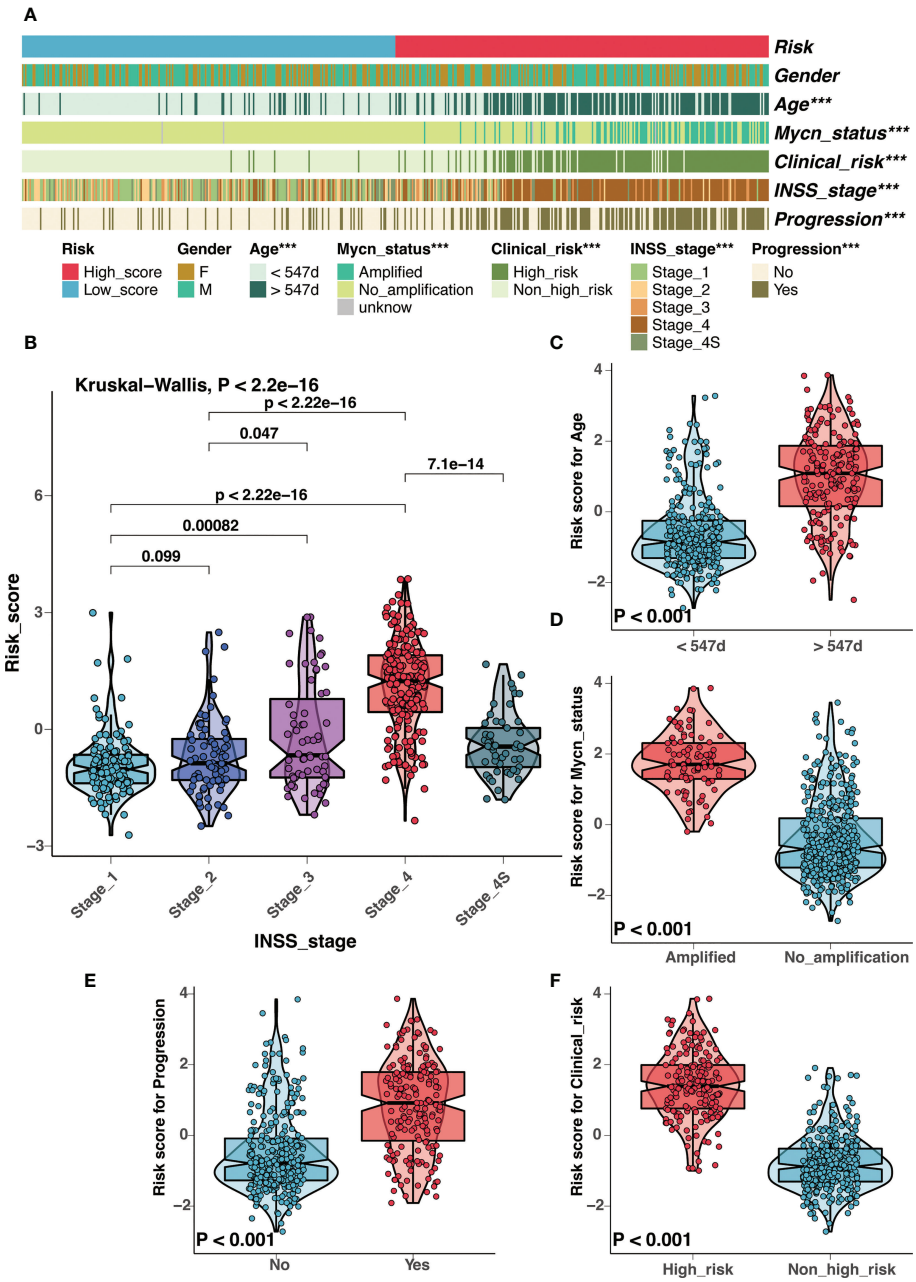


FIGURE 6
Correlation analysis between clinical characteristics and the risk signature in GSE49710 cohort. **(A)** Correlation analysis between the risk signature and clinical characteristics in GSE49710 cohort. **(B-F)** The comparison of risk scores between samples with different clinical characteristics, including INSS stage **(B)**, age **(C)**, MYCN status **(D)**, progression **(E)**, and COG risk groups **(F)**. (***: P < 0.001).

could reliably predict $CD8^+$ T-cell infiltration in the immune microenvironment of neuroblastoma.

It had been shown that gangliosides were involved in regulating NK cell cytotoxicity through multiple mechanisms (42, 43). Importantly, antibody-dependent cell-mediated cytotoxicity (ADCC) mediated by NK cells and neutrophils was a critical mechanism for anti-GD2 antibody efficacy (44). Neuroblastoma cells lacking MHC-I molecules, which served as the ligands for killer inhibitory receptors, should be particularly susceptible to NK cells. Notably, the high-score group with reduced MHC-I activity exhibited limited NK cell infiltration and cytotoxicity (Figures 10A, B). NK-mediated elimination of neuroblastoma seemed to be shielded by other mechanisms that modify the balance of activating and inhibitory signals on NK

10.3389/fimmu.2022.1061814

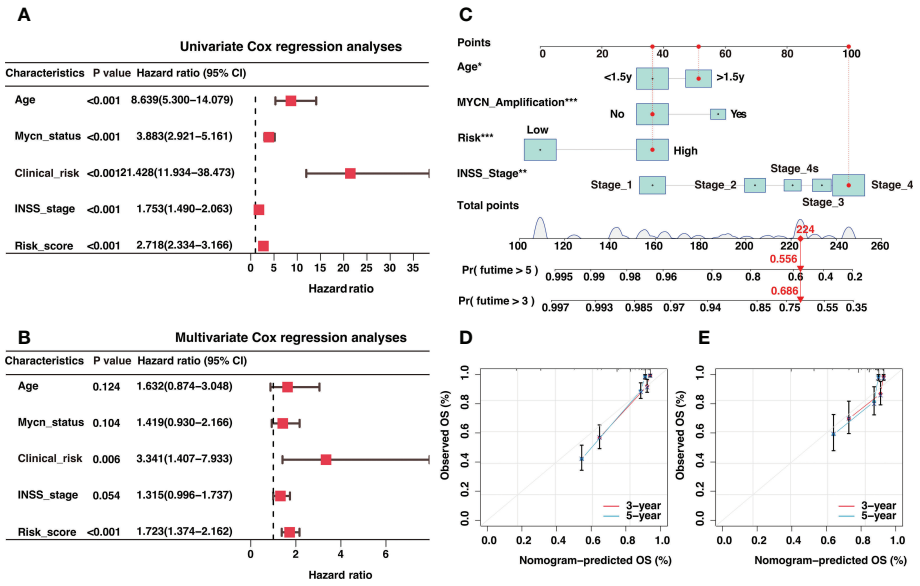


FIGURE 7
Independent prognosis analysis of the risk signature and construction of a nomogram for 3-year and 5-year overall survival (OS) prediction. **(A)** The univariate Cox regression analysis in GSE49710 cohort **(B)**. The multivariate Cox regression analysis in GSE49710 cohort. **(C)** The establishment of a nomogram that predicted 3-year and 5-year OS in the GSE49710 cohort. **(D, E)** Calibration curves of the nomogram in the prediction of 3-year and 5-year OS in GSE49710 **(D)** and E-MTAB-8248 **(E)** cohorts. *: P < 0.05, **: P < 0.01, ***: P < 0.001.

cells (20). Therefore, we compared the expression of ligands for NK-activating receptors between two risk groups. Apart from PVR, ligands for the NK cell-activating receptors DNAM-1 and NKG2D, such as MICA, MICB, and ULBP1, were downregulated in the high-score group (Figure 10C). The downregulation of NK-activating receptors may be the potential reason for the inhibition of NK cells. Taken together, escaping NK cells was an essential driver of immunosuppression in the high-score group.

Immune checkpoints (ICs) in the tumor microenvironment could significantly impact the reactivity of tumor-infiltrating lymphocytes to neuroblastoma (20). Signature genes were significantly associated with several ICs, indicating that ICs may play a role in the ganglioside-related risk signature (Figure 10D). As seen in Figure 10E, the low-score group had increased expression of PD-L1 (CD274), CD200R1, and CD200, while the high-score group had increased expression of B7-H3 (CD276). Briefly, immune evasion strategies were mediated by different ICs in low-score and high-score groups.

3.5 The ganglioside-related risk signature could predict immunotherapeutic response and chemotherapy sensitivity

Immune checkpoint inhibitors (ICIs) had revolutionized cancer treatment. However, only a small number of patients were responding (45). Considering the dramatic differences in the immune landscape between the two groups, we investigated the risk signature's predictive ability for immunotherapeutic benefits. Due to the lack of expression data for immunotherapy of neuroblastoma, we used the immunotherapy dataset of melanoma, which is also a neuroendocrine tumor, to investigate the role of the risk signature in immunotherapeutic response prediction. Figure 11A depicted the distribution of treatment response at different risk scores in GSE78220, and the risk score was significantly lower in the immunotherapy-responsive group than in the non-responsive group (Figure 11B). Additionally, samples with high scores had a poor prognosis (Figure 11C). The risk signature had excellent discrimination in predicting immunotherapy response, presenting an AUC of 0.728 (Figure 11D). In short, our findings showed that the ganglioside-related risk signature could accurately predict responsiveness to immune checkpoint blockade therapy. The low-score group showed a better response to immunotherapy than the high-score group.

Chemotherapy was the cornerstone of neuroblastoma treatment, and we examined the susceptibility of the two groups to commonly used chemotherapeutic drugs in the established treatment regimen. The IC50 values of four chemotherapeutic drugs were compared between two groups: cisplatin, doxorubicin, etoposide, vinblastine. Interestingly, the IC50 values for these chemotherapeutic drugs were significantly lower in the high-score group (Figure 11E). Function

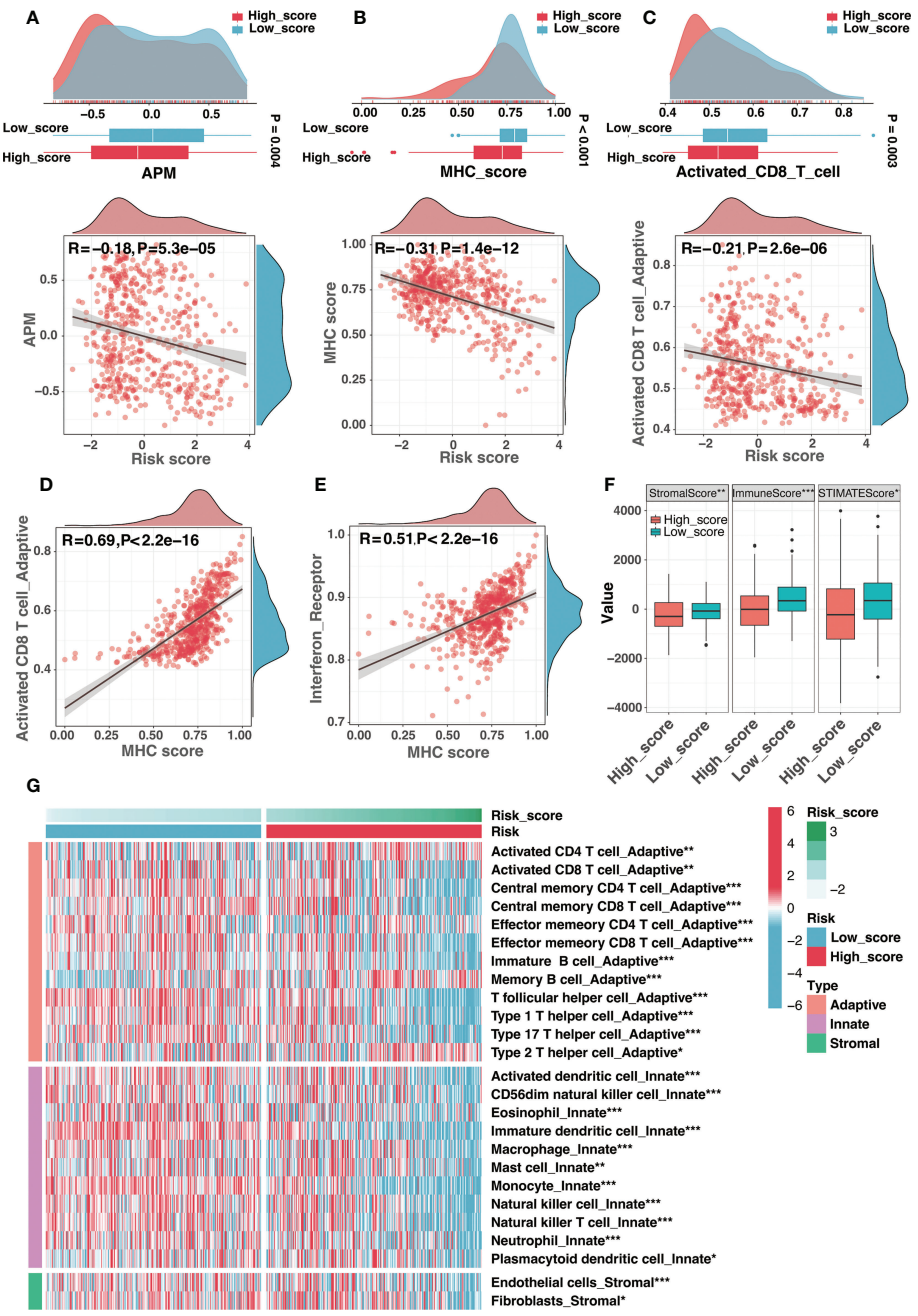


FIGURE 8
The immune landscape of the risk signature. **(A)** The comparison of antigen-presenting machinery (APM) score between different risk groups (top) and correlation analysis between the risk score and the APM score (bottom). **(B)** The comparison of MHC score between different risk groups (top) and correlation analysis between the risk score and the MHC score (bottom). **(C)** The comparison of activated CD8 T cell infiltration between different risk groups (top) and correlation analysis between the risk score and the activated CD8 T cell infiltration (bottom). **(D, E)** The correlation analysis between the MHC score and the activated CD8 T cell infiltration **(D)** as well as interferon receptor activity **(E)**. **(F)** Comparison of the stromal, immune, and ESTIMATE scores between high-score and low-score groups. **(G)** Different infiltration levels of immune infiltrating cells between low-score and high-score groups. (*:P< 0.05, **: P < 0.01, ***: P < 0.001).

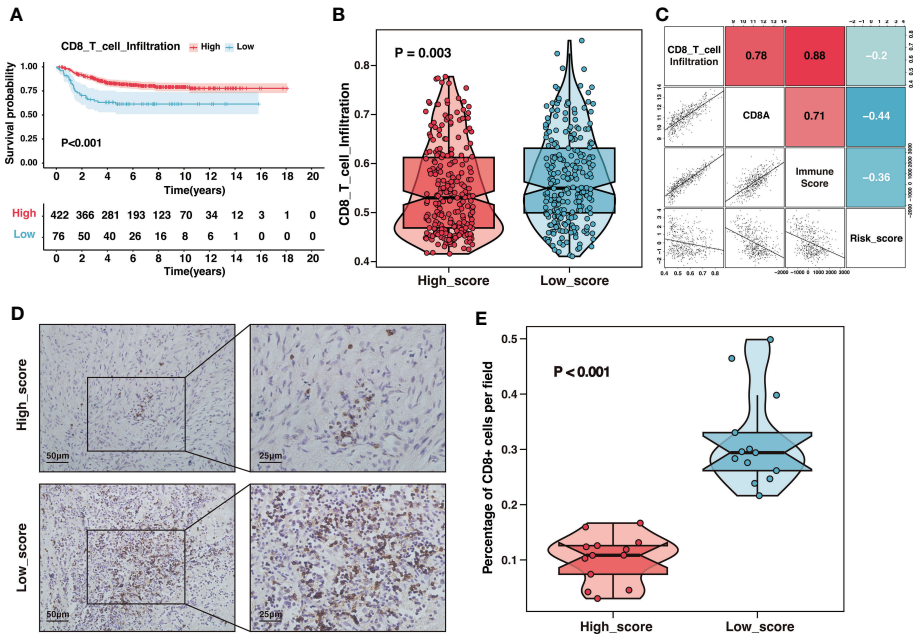


FIGURE 9
The high-score group presented low CD8$^+$ T-cell infiltration. **(A)** Kaplan–Meier curves of overall survival between samples with different infiltration of CD8$^+$ T-cell in the GSE49710 dataset. **(B)** The comparison of CD8$^+$ T-cell infiltration between high-score and low-score groups in the GSE49710 dataset. **(C)** The correlation analysis of the risk score, CD8$^+$ T-cell infiltration, CD8A expression and immune-score in the GSE49710 dataset. **(D)** Representative immunohistochemistry (IHC) images showing the infiltration of CD8$^+$ T-cell between different risk groups in Tianjin cohort. Magnification, ×200, ×400. **(E)** The comparison of CD8$^+$ T-cell infiltration between different risk groups in Tianjin cohort.

enrichment in the cell cycle and DNA replication of the high-score group may be the potential reason (Figure S9). These results suggested that the high-score group could still benefit from the established chemotherapy regime.

3.6 The high-score group presented with a high degree of dedifferentiation

The previous study showed that ganglioside was implicated in maintaining neural stem cell self-renewal capacity (15). Furthermore, the critical implications of gangliosides in tumor stem cells have been frequently highlighted (16, 46–48). Recent research suggested that the low MHC-I expression of neuroblastoma may reflect the undifferentiated state of the neural crest (20). These findings implied that gangliosides might contribute to the undifferentiated state of neuroblastoma. Interestingly, DEGs between ganglioside-related clusters were highly enriched in neural crest cell development (Figure S4). Therefore, we explored the role of ganglioside-related risk signature in reflecting the degree of neuroblastoma dedifferentiation. Firstly, we investigated the connection between the risk score and the corrected stemness index mRNAsi (c_mRNAsi). Figure 12A showed a significant positive correlation between the risk score and c_mRANsi (R = 0.62, P < 0.001). Similarly, all signature genes were significantly correlated with the c_mRNAsi (Figure 12B). We subsequently validated this result in a dataset of neuroblastoma cell lines. It had been known that all-trans retinoic acid (ATRA) could induce differentiation in both primary neuroblastomas and cell lines. As shown in Figure 12C, neuroblastoma cells treated with ATRA exhibited a reduced risk score in both BE2C and NGP cell lines. In addition, the risk score was significantly positively associated with multiple stemness markers of neuroblastoma, including CD133, EZH2, and OCT4 (Figures 12D-F). These results indicated that the high-score group presented with a high degree of dedifferentiation.

3.7 Downregulated B3GALT4 promoted the progress of neuroblastoma cells

Considering B3GALT4 as the connecting link of ganglioside and with the highest absolute value of regression coefficient in signature genes, the siRNA of B3GALT4 was transfected into 9464D and 975A2 cells, and the role of B3GALT4 in neuroblastoma was explored. The western blot and RT-qPCR were performed to verify the downregulation of B3GALT4 expression after transfection for further experiments (Figures 13A, B). As shown in Figure 13C, the CCK-8 assay

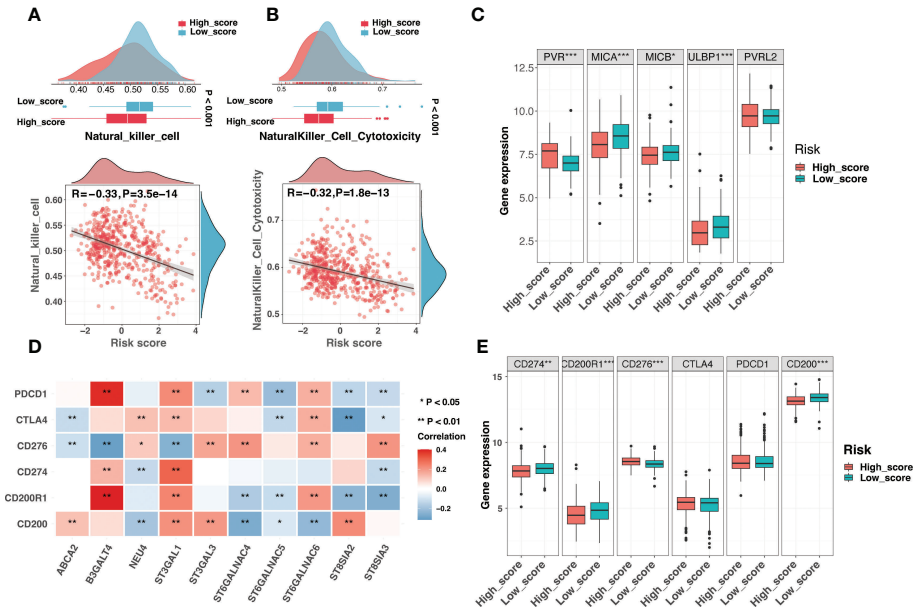


FIGURE 10
Comparison of NK cell activity and immune checkpoints in the risk signature **(A)** The comparison of the NK cell infiltration level between different risk groups (top) and correlation analysis between the risk score and the NK cell infiltration level (bottom). **(B)** The comparison of the NK cell cytotoxicity between different risk groups (top) and correlation analysis between the risk score and the NK cell cytotoxicity (bottom). **(C)** The boxplot exhibited different expression levels of ligands for NK cell-activating receptors between low-score and high-score groups. **(D)** The correlation analysis between the expression of signature genes and established immune checkpoints. **(E)** The boxplot demonstrated different expression levels of immune checkpoints between low-score and high-score groups. (*: P < 0.05, **: P < 0.01, ***: P < 0.001).

demonstrated the knockdown of B3GALT4 significantly promoted cell proliferation in both cell lines (Figure 13C). Consistently, colony formation assay showed that neuroblastoma cells with downregulation of B3GALT4 present more cell clones than the control group in both cell lines (Figures 13D, E). Besides, cells detected in the lower chamber were significantly increased after transfected with si-B3GALT4 compared to the control group in both invasion and migration assays (Figures 13F, G). Our findings suggested that B3GALT4 could inhibit the progression of neuroblastoma, verifying the protective role of B3GALT4 in the risk signature.

4 Discussion

Neuroblastoma is an extraordinarily lethal childhood tumor characterized by high heterogeneity. Precise and efficient prognostic prediction is critical to guide treatment. In the present study, we identified two ganglioside-related clusters with differential expression patterns and outcomes, indicating a substantial prognostic significance of gangliosides in neuroblastoma. To effectively predict the individualized prognosis of neuroblastoma based on gangliosides, we conducted the Lasso-penalized Cox regression analysis on ganglioside-related DEGs to identify independent prognostic genes and developed a ten-gene risk signature. The risk signature showed excellent discrimination and accuracy in GSE49710, TARGET, E-MTAB-8248, and Tianjin cohorts. Additionally, the risk signature was significantly related to several previously identified prognostic markers and was demonstrated to be an independent prognostic factor for neuroblastoma. Besides, a nomogram incorporating multiple established clinical prognostic characteristics was developed and verified. In summary, we developed a novel ganglioside-related risk signature in neuroblastoma that enables reliable and individualized prognosis prediction.

There are ten genes in the ganglioside-related risk signature. However, few of them are identified as prognostic genes in neuroblastoma. ABCA2 encodes a membrane-associated protein belonging to the ATP-binding cassette transporter superfamily and is involved in the metabolism of gangliosides (49). Additionally, ABCA2 is overexpressed in pediatric acute lymphoblastic leukemia and may contribute to multidrug resistance (50). B3GALT4 is involved in synthesizing GM1/GD1 gangliosides and has been identified as a prognostic marker for osteosarcoma and neuroblastoma (51, 52). Our

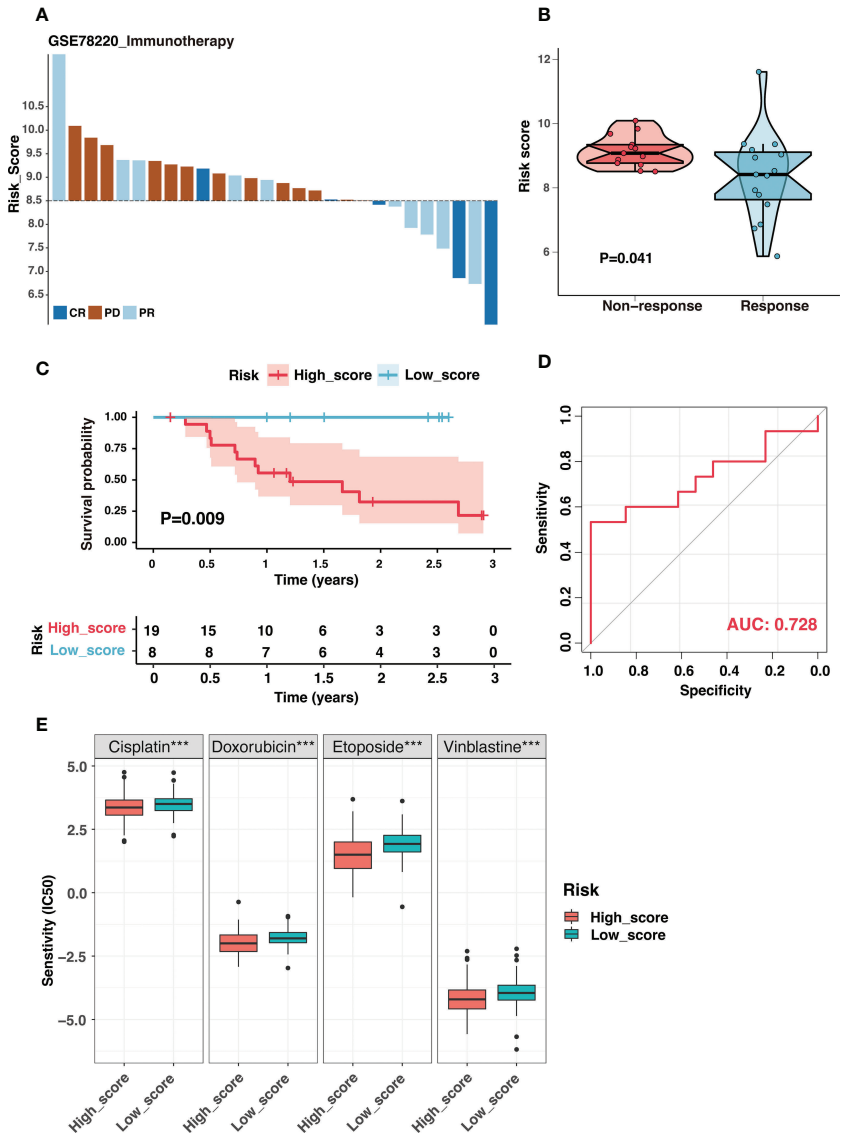


FIGURE 11
The ganglioside-related risk signature could predict immunotherapeutic response and chemotherapy sensitivity. **(A)** The distribution of immunotherapy responses at different risk scores in the GSE78220 cohort. **(B)** The comparison of risk scores between non-response and response groups in the GSE78220 cohort. **(C)** Kaplan–Meier curves of overall survival in GSE78220 cohort between different risk groups. **(D)** The receiver operating characteristic (ROC) curves for immunotherapy response prediction in the GSE78220 cohort. **(E)** The boxplot demonstrated different IC50 (the half maximal inhibitory concentration) values of four chemotherapeutic drugs, including cisplatin, doxorubicin, etoposide and vinblastine, between low-score and high-score groups. (***: P < 0.001).

results also showed the inhibition of B3GALT4 could significantly increase neuroblastoma cell proliferation, migration, and invasion *in vitro*. Notably, NEU4 has been identified as a potential regulator of neuronal development, with overexpression promoting the acquisition of a stem cell-like phenotype in neuroblastoma cells (53). The sialyltransferases are required to synthesize gangliosides, and their aberrant expression is closely related to a poor prognosis in tumors (54, 55). Among them, ST3GAL1 overexpression promotes epithelial-mesenchymal transition, migration, and invasion in ovarian cancer (56), and ST3GAL3 downregulation inhibits pancreatic cancer cell migration and invasion (57). Additionally, the sialyltransferases ST6GalNAc4, ST6GalNAc5, and ST6GalNAc6 contribute to the synthesis and metabolism of

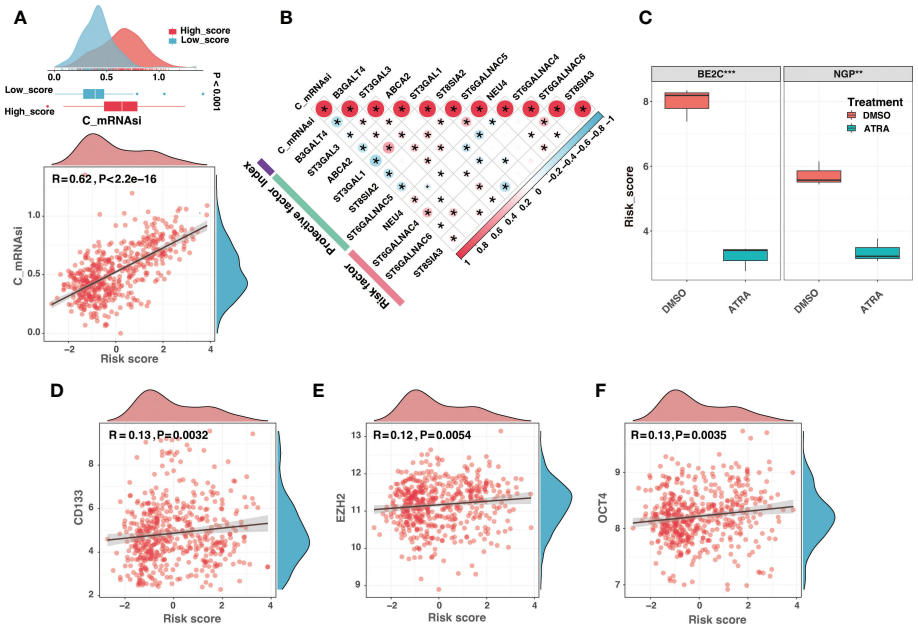


FIGURE 12
The high-score group presented with a high degree of dedifferentiation. **(A)** The comparison of the corrected mRNAsi (c_mRNAsi) between different risk groups (top) and correlation analysis between the risk score and the c_mRNAsi (bottom). **(B)** The correlation analysis between the c_mRNAsi and signature gene expression. **(C)** The comparison of risk scores between neuroblastoma cells treated with dimethyl sulfoxide (DMSO) or all-trans retinoic acid (ATRA) in BE2C and NGP cell lines. **(D-F)** The risk score was significantly positively associated with multiple stemness markers of neuroblastoma, including CD133 **(D)**, EZH2 **(E)**, and OCT4 **(F)**. (*: P < 0.05, **: P < 0.01, ***: P < 0.001).

the gangliosides GD1a and GM1b. Overexpression of ST6GalNAc4 has been crucial for tumor cell glycosylation modification and lung cancer metastasis, although the roles of ST6GalNAc5 and ST6GalNAc6 in malignancies remain unclear (55). Furthermore, the polysialyltransferase ST8SIA2 is also implicated in small cell lung cancer and glioma metastasis and invasion (55). Consistent with ST8SIA3 as a risk factor in the present study, it mediates the sialylation of GM3 and GD3 and promotes survival, proliferation, clonogenicity, and migration of glioblastoma cells (55). To summarize, these signature genes are intimately engaged in the synthesis and metabolism of gangliosides and play a critical role in tumor development and progress, supporting the predictive value of the risk signature in neuroblastoma.

Our result revealed that the low-score group was enriched in the antigen processing and presentation pathway, implying potential differences in the immune microenvironment between the two groups. Gangliosides have been identified as potent inhibitors of the cellular immune response. Soluble GD2 shed from neuroblastoma cells has been shown to suppress T cell proliferation and contribute to tumor immune evasion in neuroblastoma (8, 13, 20). Moreover, anti-GD2 antibodies have been introduced into the standard treatment regime for high-risk neuroblastoma. Given the critical role of gangliosides in the neuroblastoma immune microenvironment, we investigated and compared the immune landscape between the low-score and high-score groups.

Neuroblastoma has low immunogenicity, characterized by a low mutational burden and abnormalities in the antigen processing and presenting machinery. Our findings indicated that the high-score group exhibited an impairment in APM and low MHC-I activity, suggesting that the high-score group was less immunogenic than the low-score group. Interestingly, increasing researches indicate that the fundamental reason for MHC-I suppression may be the embryonic origin of neuroblastoma. Neuroblastoma seems to represent the underdeveloped neural crest state, characterized by low MHC-I expression. Neuroblastoma cell line differentiation is associated with increased MHC-I expression (20, 58). In the present study, low MHC-I expression in the high-score group may represent a substantial degree of dedifferentiation. High c_mRNAsi in the high-score group supports this hypothesis. We also found that ATRA-induced differentiated neuroblastoma cell lines exhibited reduced scores. The risk score was also significantly positively associated with multiple stemness markers of neuroblastoma. Generally, these results suggested that gangliosides may play a

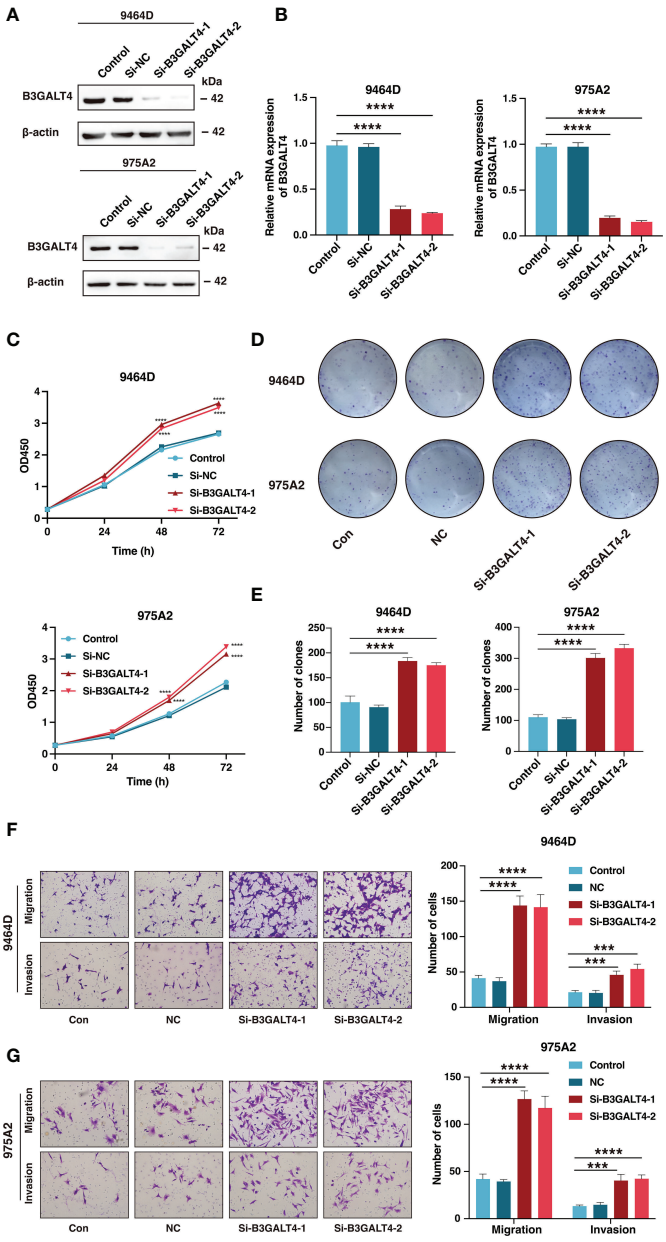


FIGURE 13
The downregulation of B3GALT4 could promote the progression of neuroblastoma. **(A, B)** The western blot analysis **(A)** and quantitative real-time PCR **(B)** was performed to validate the downregulation of B3GALT4 after transfection with siRNA in 9464D and 975A2 cells. **(C)** The CCK-8 assay was performed to measure the proliferation capacity of 9464D and 975A2 cells. **(D, E)** The colony formation assay **(D)** and corresponding statistical analysis **(E)** of 9464D and 975A2 cells. **(F, G)** The transwell assays were conducted to determine the effect of down-regulated B3GALT4 on neuroblastoma migration and invasion capacity in 9464D **(F)** and 975A2 cells **(G)**. (***, P< 0.001; ****, P< 0.0001).

role in the formation and dedifferentiation of neuroblastoma, and the ganglioside-related risk signature could reflect the degree of neuroblastoma dedifferentiation.

Low immunogenicity leads to insufficient infiltration of lymphocytes into the tumor and poor anti-tumor reactivity (20). The high-score group with low immunogenicity showed significantly reduced immune scores and infiltration levels. The $CD8^+$ T-cell, a crucial component of the immune response to tumors, has been identified as one of the most significant immunotherapy targets for tumors (59). Recent research has demonstrated that $CD8^+$ T-cell infiltration is substantially related to neuroblastoma prognosis (20). We discovered that the risk signature could predict $CD8^+$ T-cell infiltration accurately and validated it using tissue samples in

the Tianjin cohort. Interestingly, several immunosuppressive and stromal cells were substantially infiltrated in the low-score group. ST3GAL1 has been implicated in transforming tumor-associated macrophage differentiation to a more suppressive phenotype (55). Consistently, our results also showed that the expression level of ST3GAL1 was significantly upregulated in the low-score group. Immunosuppressive cell infiltration represents a potential immune escape mechanism in the low-score group, and targeting these cells may be an effective immunotherapeutic strategy.

Generally, the cytotoxic activity of NK cells is inhibited by the binding of killer-cell immunoglobulin-like receptors to MHC-I molecules on normal cells. In contrast, tumor cells typically lack MHC-I and are thus vulnerable to NK-mediated killing (20, 39, 60). Surprisingly, the high-score group with low MHC-I activity had a low NK cell infiltration and cytotoxicity level, indicating a potent NK cell suppression in this group. The imbalance of NK cell activation and inhibitory signaling is a critical mechanism of immune escape in neuroblastoma (20, 39). We found ligands for NK cell activating receptors, including MICA, MICB, and ULBP1 were downregulated in the high-score group. The inhibition of NK cell activation may be the potential reason for escaping NK cells in the high-score group. Additionally, previous research showed that overexpression of B7-H3 molecules could inhibit NK cell cytotoxicity, and we found that the high-score group exhibited increased B7-H3 expression (60). The escaping NK cell significantly contributes to immunosuppression in the high-score group. Fortunately, the anti-GD2 antibody Dinutuximab could restore the NK cell balance and promote NK cell cytotoxicity against neuroblastoma potently, highlighting the need for anti-GD2 immunotherapy in the high-score group (20).

The presence of immune checkpoints is a critical mechanism by which cancers escape the immune system, and immune checkpoint inhibitors have made breakthroughs in adult malignancies. However, the efficacy of ICIs in neuroblastoma is not satisfactory, and identifying individuals who respond to immunotherapy is essential. We found that PD-L1 was significantly overexpressed in the low-score group, indicating that the low-score group may benefit from ICIs (20). The ganglioside-related risk signature could effectively predict immunotherapy response in the ICI-treated cohort, with better response to immunotherapy in the low-score group. Therefore, ICIs are potential options for immunotherapy in the low-score group, further verification by large-scale and multi-center investigations are required in neuroblastoma.

There are some limitations to this study. There is a shortage of molecular sequencing data for anti-GD2 therapy in neuroblastoma, and the risk signature for predicting anti-GD2 antibody immunotherapy response requires additional confirmation. It is encouraging to note that the anti-GD2 immunotherapy for neuroblastoma has been introduced in China, and a relevant clinical trial is in progress in our institution (CTR20221154). Furthermore, since this study is based on retrospective data, it requires further validation from a comprehensive perspective research. The treatment strategies that we advocated for the different subgroups in the risk signature, including chemotherapy and immunotherapy, should be assessed in prospective clinical trials. Additionally, thorough laboratory investigations are necessary to elucidate the comprehensive biological functions of signature genes.

5 Conclusion

In this work, we developed a novel ganglioside-related risk signature that enabled precise prognostic prediction of neuroblastoma. Additionally, the risk signature identified distinct immune landscapes and immune evasion strategies between risk groups and could be used to predict immunotherapy response. This work emphasizes the critical role of gangliosides in the prognosis and immune microenvironment of neuroblastoma, which may inform clinical evaluation and therapeutic decision-making.

Data availability statement

Publicly available datasets were analyzed in this study. The names of the repositories and accession numbers can be found within the article/Supplementary Materials.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (E20210027). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

LH and QZ designed the study. JY, LH, YS, YJ, ZL, BG, JL, YL, and YW analyzed and interpreted data. JY drafted the manuscript, and major revised by LH and QZ. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Key Research and Development Program of China (2018YFC1313000,

2018YFC1313001) and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-009A).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1061814/full#supplementary-material

FIGURE S1
The workflow of the present study.

FIGURE S2
Consensus clustering analysis based on differentially expressed ganglioside-related genes in the GSE49710 dataset. (A-H) Consensus score matrix of samples when k = 2-9. (I) CDF of the consensus matrix for each k (indicated by colors). (J) Relative alterations in the area under CDF curves. (K) Tracking plot for each k.

FIGURE S3
Consensus clustering analysis based on differentially expressed ganglioside-related genes in the TARGET dataset. (A-H) Consensus score matrix of samples when k = 2-9. (I) CDF of the consensus matrix for each k (indicated by colors). (J) Relative alterations in the area under CDF curves. (K) Tracking plot for each k. (L) The t-SNE analysis revealed a clear distinction between the two clusters in the TARGET cohort. (M) Kaplan–Meier curves of overall survival in TARGET cohort between different ganglioside-related clusters.

FIGURE S4
Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis of differentially expressed genes between ganglioside-related clusters A and B (indicated by colors).

FIGURE S5
Validation of the ganglioside-related risk signature. (A) The distribution of the risk score (top) and survival time (bottom) in the E-MTAB-8248 cohort. (B) The receiver operating characteristic (ROC) curves of the risk signature for 3-year and 5-year OS prediction in the E-MTAB-8248 cohort. (C) The ROC curves of clinical prognostic factors for 3-year OS prediction in the E-MTAB-8248 cohort. (D) Kaplan–Meier curves of overall survival (OS) in the E-MTAB-8248 cohort between different risk groups. (E) The distribution of the risk score (top) and survival time (bottom) in the TARGET cohort. (F) The ROC curves of the risk signature for 3-year and 5-year OS prediction in the TARGET cohort. (G) The ROC curves of clinical prognostic factors for 3-year OS prediction in the TARGET cohort. (H) Kaplan–Meier curves of overall survival (OS) in the TARGET cohort between different risk groups. (I-K) The principal component analysis (PCA) of the risk signature in GSE49710 (I), E-MTAB-8248 (J), and TARGET (K) cohorts.

FIGURE S6
The prognostic value of the ganglioside-related risk signature in samples in COG high-risk group or older than 18 months. (A-B) Kaplan–Meier curves of samples in COG high-risk group in GSE49710 (A) and TARGET (B) cohorts. (C-D) Kaplan–Meier curves of samples older than 18 months in GSE49710 (C) and TARGET (D) cohorts.

FIGURE S7
Correlation analysis between clinical characteristics and the risk signature in the Tianjin cohort. (A-E) The comparison of risk scores between samples with different clinical characteristics in the Tianjin cohort, including INRGSS stage (A), MYCN status (B), COG risk groups (C), pathology (D), and INSS stage (E).

FIGURE S8
Independent prognosis analysis of the risk signature and external validation of the nomogram. (A-B) The univariate and multivariate Cox regression analysis in E-MTAB-8248 (A) and TARGET (B) cohorts. (C) Calibration curves of the nomogram in the prediction of 3-year and 5-year overall survival (OS) in the TARGET cohort.

FIGURE S9
Gene set enrichment analysis (GSEA) of the risk signature. The GSEA in the high-score (A) and low-score (B) groups.

References

1. Kholodenko IV, Kalinovsky DV, Doronin II, Deyev SM, Kholodenko RV. Neuroblastoma origin and therapeutic targets for immunotherapy. *J Immunol Res* (2018) 2018:7394268. doi: 10.1155/2018/7394268

2. Johnsen JI, Dyberg C, Wickström M. Neuroblastoma-a neural crest derived embryonal malignancy. *Front Mol Neurosci* (2019) 12:9. doi: 10.3389/fnmol.2019.00009

3. Zafar A, Wang W, Liu G, Wang X, Xian W, McKeon F, et al. Molecular targeting therapies for neuroblastoma: Progress and challenges. *Med Res Rev* (2021) 41(2):961–1021. doi: 10.1002/med.21750

4. Irwin MS, Naranjo A, Zhang FF, Cohn SL, London WB, Gastier-Foster JM, et al. Revised neuroblastoma risk classification system: A report from the children's oncology group. *J Clin Oncol* (2021) 39(29):3229–41. doi: 10.1200/jco.21.00278

5. Groux-Degroote S, Guérardel Y, Delannoy P. Gangliosides: Structures, biosynthesis, analysis, and roles in cancer. *Chembiochem* (2017) 18(13):1146–54. doi: 10.1002/cbic.201600705

6. Groux-Degroote S, Delannoy P. Cancer-associated glycosphingolipids as tumor markers and targets for cancer immunotherapy. *Int J Mol Sci* (2021) 22 (11):6145. doi: 10.3390/ijms22116145

7. Liu J, Zheng X, Pang X, Li L, Wang J, Yang C, et al. Ganglioside Gd3 synthase (Gd3s), a novel cancer drug target. *Acta Pharm Sin B* (2018) 8(5):713–20. doi: 10.1016/j.apsb.2018.07.009

8. Piazzesi A, Afsar SY, van Echten-Deckert G. Sphingolipid metabolism in the development and progression of cancer: One cancer's help is another's hindrance. *Mol Oncol* (2021) 15(12):3256–79. doi: 10.1002/1878-0261.13063

9. Wang H, Isaji T, Satoh M, Li D, Arai Y, Gu J. Antitumor effects of exogenous ganglioside Gm3 on bladder cancer in an orthotopic cancer model. *Urology* (2013) 81(1):210.e11–5. doi: 10.1016/j.urology.2012.08.015

10. Sasaki N, Hirabayashi K, Michishita M, Takahashi K, Hasegawa F, Gomi F, et al. Ganglioside Gm2, highly expressed in the mia paca-2 pancreatic ductal adenocarcinoma cell line, is correlated with growth, invasion, and advanced stage. *Sci Rep* (2019) 9(1):19369. doi: 10.1038/s41598-019-55867-4

11. Suzuki M, Nagane M, Kato K, Yamauchi A, Shimizu T, Yamashita H, et al. Endothelial ganglioside Gm3 regulates angiogenesis in solid tumors. *Biochem Biophys Res Commun* (2021) 569:10–6. doi: 10.1016/j.bbrc.2021.06.063

12. Ramos RI, Bustos MA, Wu J, Jones P, Chang SC, Kiyohara E, et al. Upregulation of cell surface Gd3 ganglioside phenotype is associated with human melanoma brain metastasis. *Mol Oncol* (2020) 14(8):1760–78. doi: 10.1002/1878-0261.12702

13. Cavdarli S, Groux-Degroote S, Delannoy P. Gangliosides: The double-edge sword of neuro-ectodermal derived tumors. *Biomolecules* (2019) 9(8):311. doi: 10.3390/biom9080311

14. Schengrund CL. Gangliosides and neuroblastomas. *Int J Mol Sci* (2020) 21 (15):5313. doi: 10.3390/ijms21155313

15. Wang J, Yu RK. Interaction of ganglioside Gd3 with an egf receptor sustains the self-renewal ability of mouse neural stem cells *in vitro*. *Proc Natl Acad Sci U.S.A.* (2013) 110(47):19137–42. doi: 10.1073/pnas.1307224110

16. Yeh SC, Wang PY, Lou YW, Khoo KH, Hsiao M, Hsu TL, et al. Glycolipid Gd3 and Gd3 synthase are key drivers for glioblastoma stem cells and tumorigenicity. *Proc Natl Acad Sci U.S.A.* (2016) 113(20):5592–7. doi: 10.1073/pnas.1604721113

17. Shibuya H, Hamamura K, Hotta H, Matsumoto Y, Nishida Y, Hattori H, et al. Enhancement of malignant properties of human osteosarcoma cells with disialyl gangliosides Gd2/Gd3. *Cancer Sci* (2012) 103(9):1656–64. doi: 10.1111/j.1349-7006.2012.02344.x

18. Sasaki N, Toyoda M, Ishiwata T. Gangliosides as signaling regulators in cancer. *Int J Mol Sci* (2021) 22(10):5076. doi: 10.3390/ijms22105076

19. Yoshida S, Fukumoto S, Kawaguchi H, Sato S, Ueda R, Furukawa K. Ganglioside G(D2) in small cell lung cancer cell lines: Enhancement of cell proliferation and mediation of apoptosis. *Cancer Res* (2001) 61(10):4244–52.

20. Wienke J, Dierselhuis MP, Tytgat GAM, Künkele A, Nierkens S, Molenaar JJ. The immune landscape of neuroblastoma: Challenges and opportunities for novel therapeutic strategies in pediatric oncology. *Eur J Cancer* (2021) 144:123–50. doi: 10.1016/j.ejca.2020.11.014

21. Dillinger B, Ahmadi-Erber S, Lau M, Hoelzl MA, Erhart F, Juergens B, et al. Ifn-Γ and tumor gangliosides: Implications for the tumor microenvironment. *Cell Immunol* (2018) 325:33–40. doi: 10.1016/j.cellimm.2018.01.014

22. Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, et al. Comparison of rna-seq and microarray-based models for clinical endpoint prediction. *Genome Biol* (2015) 16(1):133. doi: 10.1186/s13059-015-0694-1

23. Roderwieser A, Sand F, Walter E, Fischer J, Gecht J, Bartenhagen C, et al. Telomerase is a prognostic marker of poor outcome and a therapeutic target in neuroblastoma. *JCO Precis Oncol* (2019) 3:1–20. doi: 10.1200/po.19.00072

24. Kang W, Hu J, Zhao Q, Song F. Identification of an autophagy-related risk signature correlates with immunophenotype and predicts immune checkpoint blockade efficacy of neuroblastoma. *Front Cell Dev Biol* (2021) 9:731380. doi: 10.3389/fcell.2021.731380

25. Wilkerson MD, Hayes DN. Consensusclusterplus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26 (12):1572–3. doi: 10.1093/bioinformatics/btq170

26. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An r package for comparing biological themes among gene clusters. *Omics* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

27. Wang S, He Z, Wang X, Li H, Liu XS. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *Elife* (2019) 8: e49020. doi: 10.7554/eLife.49020

28. Liu Z, Guo Y, Yang X, Chen C, Fan D, Wu X, et al. Immune landscape refines the classification of colorectal cancer with heterogeneous prognosis, tumor microenvironment and distinct sensitivity to frontline therapies. *Front Cell Dev Biol* (2021) 9:784199. doi: 10.3389/fcell.2021.784199

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

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

31. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, et al. Immport: Disseminating data to the public for the future of immunology. *Immunol Res* (2014) 58(2-3):234–9. doi: 10.1007/s12026-014-8516-1

32. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* (2018) 173(2):338–54.e15. doi: 10.1016/j.cell.2018.03.034

33. Daily K, Ho Sui SJ, Schriml LM, Dexheimer PJ, Salomonis N, Schroll R, et al. Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. *Sci Data* (2017) 4:170030. doi: 10.1038/sdata.2017.30

34. Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, et al. Integrated genomic analysis of diverse induced pluripotent stem cells from the progenitor cell biology consortium. *Stem Cell Rep* (2016) 7(1):110–25. doi: 10.1016/j.stemcr.2016.05.006

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

36. Kim D, Langmead B, Salzberg SL. Hisat: A fast spliced aligner with low memory requirements. *Nat Methods* (2015) 12(4):357–60. doi: 10.1038/nmeth.3317

37. Liao Y, Smyth GK, Shi W. Featurecounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* (2014) 30(7):923–30. doi: 10.1093/bioinformatics/btt656

38. Cheung IY, Vickers A, Cheung NK. Sialyltransferase stx (St8siaii): A novel molecular marker of metastatic neuroblastoma. *Int J Cancer* (2006) 119(1):152–6. doi: 10.1002/ijc.21789

39. Vanichapol T, Chutipongtanate S, Anurathapan U, Hongeng S. Immune escape mechanisms and future prospects for immunotherapy in neuroblastoma. *BioMed Res Int* (2018) 2018:1812535. doi: 10.1155/2018/1812535

40. Reid GS, Shan X, Coughlin CM, Lassoued W, Pawel BR, Wexler LH, et al. Interferon-Gamma-Dependent infiltration of human T cells into neuroblastoma tumors *in vivo*. *Clin Cancer Res* (2009) 15(21):6602–8. doi: 10.1158/1078-0432.Ccr-09-0829

41. Chowdhury F, Dunn S, Mitchell S, Mellows T, Ashton-Key M, Gray JC. Pd-L1 and Cd8+Pd1+ lymphocytes exist as targets in the pediatric tumor microenvironment for immunomodulatory therapy. *OncoImmunology* (2015) 4 (10):e1029701. doi: 10.1080/2162402X.2015.1029701

42. Kawasaki Y, Ito A, Withers DA, Taima T, Kakoi N, Saito S, et al. Ganglioside Dsgb5, preferred ligand for siglec-7, inhibits nk cell cytotoxicity against renal cell carcinoma cells. *Glycobiology* (2010) 20(11):1373–9. doi: 10.1093/glycob/cwq116

43. Nicoll G, Avril T, Lock K, Furukawa K, Bovin N, Crocker PR. Ganglioside Gd3 expression on target cells can modulate nk cell cytotoxicity *Via* siglec-7-Dependent and -independent mechanisms. *Eur J Immunol* (2003) 33(6):1642–8. doi: 10.1002/eji.200323693

44. Morandi F, Sabatini F, Podestà M, Airoldi I. Immunotherapeutic strategies for neuroblastoma: Present, past and future. *Vaccines (Basel)* (2021) 9(1):43. doi: 10.3390/vaccines9010043

45. Morad G, Helmink BA, Sharma P, Wargo JA. Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* (2021) 184(21):5309–37. doi: 10.1016/j.cell.2021.09.020

46. Liang YJ, Wang CY, Wang IA, Chen YW, Li LT, Lin CY, et al. Interaction of glycosphingolipids Gd3 and Gd2 with growth factor receptors maintains breast cancer stem cell phenotype. *Oncotarget* (2017) 8(29):47454–73. doi: 10.18632/oncotarget.17665

47. Jiang L, Hao Y, Shao C, Wu Q, Prager BC, Gimple RC, et al. Adar1-mediated rna editing links ganglioside catabolism to glioblastoma stem cell maintenance. *J Clin Invest* (2022) 132(6):e143397. doi: 10.1172/jci143397

48. Battula VL, Shi Y, Evans KW, Wang RY, Spaeth EL, Jacamo RO, et al. Ganglioside Gd2 identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest* (2012) 122(6):2066–78. doi: 10.1172/jci59735

49. Sakai H, Tanaka Y, Tanaka M, Ban N, Yamada K, Matsumura Y, et al. Abca2 deficiency results in abnormal sphingolipid metabolism in mouse brain. *J Biol Chem* (2007) 282(27):19692–9. doi: 10.1074/jbc.M611056200

50. Aberuyi N, Rahgozar S, Moafi A. The role of atp-binding cassette transporter A2 in childhood acute lymphoblastic leukemia multidrug resistance. *Iran J Ped Hematol Oncol* (2014) 4(3):118–26.

51. Zhang M, Liu Y, Kong D. Identifying biomolecules and constructing a prognostic risk prediction model for recurrence in osteosarcoma. *J Bone Oncol* (2021) 26:100331. doi: 10.1016/j.jbo.2020.100331

52. Sha Y, Han L, Sun B, Zhao Q. Identification of a glycosyltransferase signature for predicting prognosis and immune microenvironment in neuroblastoma. *Front Cell Dev Biol* (2021) 9:769580. doi: 10.3389/fcell.2021.769580

53. Silvestri I, Testa F, Zappasodi R, Cairo CW, Zhang Y, Lupo B, et al. Sialidase Neu4 is involved in glioblastoma stem cell survival. *Cell Death Dis* (2014) 5(8): e1381. doi: 10.1038/cddis.2014.349

54. Zhang X, Yang X, Chen M, Zheng S, Li J, Lin S, et al. St3gal3 confers Paclitaxel–Mediated chemoresistance in ovarian cancer cells by attenuating Caspase –8/3 signaling. *Mol Med Rep* (2019) 20(5):4499–506. doi: 10.3892/mmr.2019.10712

55. Hugonnet M, Singh P, Haas Q, von Gunten S. The distinct roles of sialyltransferases in cancer biology and onco-immunology. *Front Immunol* (2021) 12:799861. doi: 10.3389/fimmu.2021.799861

56. Wu X, Zhao J, Ruan Y, Sun L, Xu C, Jiang H. Sialyltransferase St3gal1 promotes cell migration, invasion, and tgf-β1-Induced emt and confers paclitaxel resistance in ovarian cancer. *Cell Death Dis* (2018) 9(11):1102. doi: 10.1038/s41419-018-1101-0

57. Guerrero PE, Miró L, Wong BS, Massaguer A, Martínez-Bosch N, Llorens R, et al. Knockdown of α2,3-sialyltransferases impairs pancreatic cancer cell migration, invasion and e-Selectin-Dependent adhesion. *Int J Mol Sci* (2020) 21 (17):6239. doi: 10.3390/ijms21176239

58. Spel L, Nieuwenhuis J, Haarsma R, Stickel E, Bleijerveld OB, Altelaar M, et al. Nedd4-binding protein 1 and Tnfaip3-interacting protein 1 control mhc-1 display in neuroblastoma. *Cancer Res* (2018) 78(23):6621–31. doi: 10.1158/0008-5472.Can-18-0545

59. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent advances in targeting Cd8 T-cell immunity for more effective cancer immunotherapy. *Front Immunol* (2018) 9:14. doi: 10.3389/fimmu.2018.00014

60. Pelosi A, Fiore PF, Di Matteo S, Veneziani I, Caruana I, Ebert S, et al. Pediatric tumors-mediated inhibitory effect on nk cells: The case of neuroblastoma and wilms' tumors. *Cancers (Basel)* (2021) 13(10):2374. doi: 10.3390/cancers13102374

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/fimmu.2023.1097075

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Xuefeng An,
Tianjin Medical University General Hospital, China
Shasha He,
Second Xiangya Hospital, Central South University, China
Yi Liu,
Mayo Clinic Florida, United States

*CORRESPONDENCE
Yingying He
729334324@qq.com

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

RECEIVED 13 November 2022
ACCEPTED 12 January 2023
PUBLISHED 25 January 2023

CITATION
Lu D, Liao J, Cheng H, Ma Q, Wu F, Xie F and He Y (2023) Construction and systematic evaluation of a machine learning-based cuproptosis-related lncRNA score signature to predict the response to immunotherapy in hepatocellular carcinoma.
Front. Immunol. 14:1097075.
doi: 10.3389/fimmu.2023.1097075

Construction and systematic evaluation of a machine learning-based cuproptosis-related lncRNA score signature to predict the response to immunotherapy in hepatocellular carcinoma

Dingyu Lu[1], Jian Liao[2], Hao Cheng[1], Qian Ma[1], Fei Wu[1], Fei Xie[1] and Yingying He[1]*

[1]Oncology Department, Deyang People's Hospital, Deyang, China, [2]Intensive care Unit, Deyang People's Hospital, Deyang, China

Introduction: Hepatocellular carcinoma (HCC) is a common malignant cancer with a poor prognosis. Cuproptosis and associated lncRNAs are connected with cancer progression. However, the information on the prognostic value of cuproptosis-related lncRNAs is still limited in HCC.

Methods: We isolated the transcriptome and clinical information of HCC from TCGA and ICGC databases. Ten cuproptosis-related genes were obtained and related lncRNAs were correlated by Pearson's correlation. By performing lasso regression, we created a cuproptosis-related lncRNA prognostic model based on the cuproptosis-related lncRNA score (CLS). Comprehensive analyses were performed, including the fields of function, immunity, mutation and clinical application, by various R packages.

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

Conclusion: We created a prognostic CLS signature by machine learning, and we comprehensively analyzed the signature in the fields of function, immunity, mutation and clinical application.

KEYWORDS

cuproptosis-related lncRNA score, hepatocellular carcinoma, machine learning, prognostic model, immunotherapy

Introduction

Hepatocellular carcinoma (HCC) ranks fifth in most common carcinoma and second in cancer-related death (1). As a major histological type, HCC is identified by a high mortality rate and rapid progression (2). The main treatments for early and advanced HCC include surgical resection, multi-kinase inhibitors and immunotherapy. However, the therapeutic effect was limited due to the treatment resistance or adverse reactions (3–5). Therefore, it is vital to individually predict the overall survival rate and sensitivity of the drugs to guide clinical treatment and improve the therapeutic effect for HCC patients.

Cuproptosis is an innovative cell death pathway in which copper can directly bind to the tricarboxylic acid (TCA) cycle and cause protein stress, which eventually results in cell death (6). Copper, which is essential for life, plays a vital role in regulating homeostasis. Lack of copper may cause dysfunction of copper-binding enzymes. However, increasing the level of copper may lead to cell death (7). A recent study revealed that the level of intracellular copper may regulate the progression of cancer (8). Thus, increasing the accumulation of intracellular cancer is considered to be a novel therapeutic target for cancer cell killing (9). According to the mechanism, it is necessary to determine the regulators of the novel form of cell death in HCC patients.

Long noncoding RNAs (lncRNAs) consist of more than 200 nucleotides and mostly do not encode proteins (10). The functions of some lncRNAs have been widely studied, and they are involved in regulating chromatin dynamics, genes, cell differentiation, growth and development (11). Thanks to next-generation sequencing, thousands of lncRNAs have been revealed to be abnormally expressed in various cancers (12). Most importantly, many lncRNAs were associated with prognosis in many types of cancer as well as potential therapeutic targets (13–15).

In our study, we constructed a novel machine learning-based cuproptosis-related lncRNA prognostic signature for HCC patients with bioinformatic analysis. We performed functional, immune and mutational analyses to comprehensively evaluate the created model. Moreover, our model can guide the clinical treatment with satisfactory results.

Methods

Data extraction

Ten cuproptosis-related genes were obtained from a previous article. The related data, including transcriptome RNA sequencing and clinical data, were extracted from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) and International Cancer Genome Consortium (ICGC) (https://dcc.icgc.org) online databases. Patients in both datasets were collected based on the following criteria: (a) pathological diagnosed with LIHC (Liver hepatocellular carcinoma); (b) available clinical information (including age, gender, stage, and complete follow-up information); (c) available gene expression matrix. Finally, we collected 340 patients in the TCGA-LIHC cohort and 226 patients in the ICGC-LIHC cohort. The cohort of DNA methylation and copy number were obtained from UCSC Xena (https://xena.ucsc.edu/), which belongs to University of California Santa Cruz.

Establishment of the cuproptosis-related prognostic lncRNA signature

We explored the correlation between 10 cuproptosis-related genes and lncRNAs by performing Pearson's correlation with a P-value < 0.05. The network was constructed by R the package "Igraph". To filter the prognostic lncRNAs and establish the cuproptosis-related prognostic lncRNA signature, we performed LASSO regression. The corresponding coefficients (β) of the signature were obtained. The cuproptosis-related lncRNA score (CLS) was calculated by the following formula: CLS = Σ [expression (cuproptosis-related prognostic lncRNA signature)*β]. The cutoff value was the median CLS value in each data set.

Validation of the cuproptosis-related prognostic lncRNA signature

We constructed the lncRNA signature by using the TCGA dataset as the training cohort. Afterward, the ICGC dataset was used for validation as the testing cohort. To evaluate the capacity of prediction, we calculated the concordance index (C-index) by using the R package "Pec". The area under the curve (AUC) analysis was obtained to assess the reliability of our signature with the R package "timeROC". The heatmap was created by the R package "pheatmap". Kaplan-Meier (K-M) analysis was performed in TCGA and ICGC cohorts with the R package "survival".

RNA isolation and RT-qPCR

We isolated RNA using an RNeasy Mini Kit (QIAGEN, Hilden, Germany). The RNA was reversed to cRNA by utilizing a High-Capacity RNA-to-cDNA^{TM} Kit (Thermo Fisher Scientific, Hilden, Germany). Afterward, we performed RT-qPCR with PowerUp^{TM} SYBR^{TM} Green Master Mix (Thermo Fisher Scientific, Hilden, Germany) based on the manufacturer's instructions. The sequences of the lncRNA primers are shown (Table S2). The relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method.

Nomogram establishment based on CLS

We performed the univariate Cox regression and multivariate Cox regression with the R package "survival". To individually predict the overall survival rate, we established a CLS-based nomogram according to the Cox regression analysis by the R package "RMS". Then, we obtained the calibration curves and AUCs by utilizing the R packages "rms" and "survivalROC" respectively. Moreover, the decision curve analysis (DCA) was analyzed with the R package "rmda" to further evaluate the superiority of the nomogram.

Functional and immune analyses

The correlation heatmap was analyzed by the R package "ggcor". After obtaining the differentially expressed genes, we introduced an online resource called Metascape (https://metascape.org) to determine the enrichment items. Gene set enrichment analysis (GSEA) was used to analyze the enriched pathways. The immune-correlated pathways were isolated from a previous article (16). Other pathways of interest were obtained from a published article (17). We obtained the homologous recombination deficiency (HRD) score, cancer-testis antigen (CTA) score and intratumor heterogeneity from an article (18). The R package "cibersortR" was utilized to obtain the relative abundance of each tumor-infiltrating immune cell (TIC) in each sample. Moreover, the tumor microenvironment was analyzed by ESTIMATE algorithm.

Mutational analyses

The mutational data were extracted from the TCGA using the R package "TCGAbiolinks". We created the mutational waterfall plot and the lollipop chart with the R package "maftools". The tumor mutational burden (TMB) of each sample was calculated. Furthermore, the mutational spectrum of mutational signatures was determined based on the R package "MutationalPattern".

Clinical decision based on CLS

The genomics of drug sensitivity in cancer (GDSC) database (www.cancerRxgene.org) was introduced. The half-maximal inhibitory concentration (IC50) was calculated with the R package "pRRophetic". The immunophenoscore (IPS) was calculated with a reported algorithm (19). We performed subclass mapping analysis (20) to assess the response to PD-1 and CTLA4 in an existing dataset containing comprehensive immunotherapy information in melanoma patients (21).

The response to immunotherapy was detected by tumor immune dysfunction and exclusion (TIDE) mode (http://tide.dfci.harvard.edu) (22). Five biomarkers, including IPS, interferon gamma (IFNG), CD274, CD8 and myeloid-derived suppressor cell (MDSC), were compared with CLS to evaluate the accuracy of prediction according to the AUC analyses. In addition, the database ConnectivityMap (https://clue.io/) was utilized to figure out the potential small molecule drugs and the corresponding mechanism of action.

Statistical analyses

R software (version 4.0.4) was used for all statistical analyses. Adobe Illustrator was used for managing all figures. We performed the correlation analyses by Pearson's correlation. The Wilcoxon test was used to analyze the difference between two groups. The proportion of the data was evaluated *via* the chi-squared test. A P-value less than 0.05 was considered to be significant. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Results

Ten cuproptosis-associated genes and related lncRNAs were identified

According to a recent high-quality article (6), we collected 10 cuproptosis-associated genes for further research (Table S1). First, we analyzed the fold change, mutational frequency, methylation and hazard ratio of ten cuproptosis-associated genes (Figure 1A). *DLAT*, *DLD*, *GLS*, *LIPT1*, *MTF1*, *PDHB* and *FDX1* were highly expressed in HCC, while *PDHA1* and *LIAS* were downregulated in HCC. *CDKN2A* was considered to be the most frequently mutated gene. The lowest methylation level was found in the *GLS* gene. *DLAT* was found to be a risk factor in HCC. Afterward, we performed Pearson's correlation to identify 242 correlated lncRNAs with a P-value < 0.05, and the result

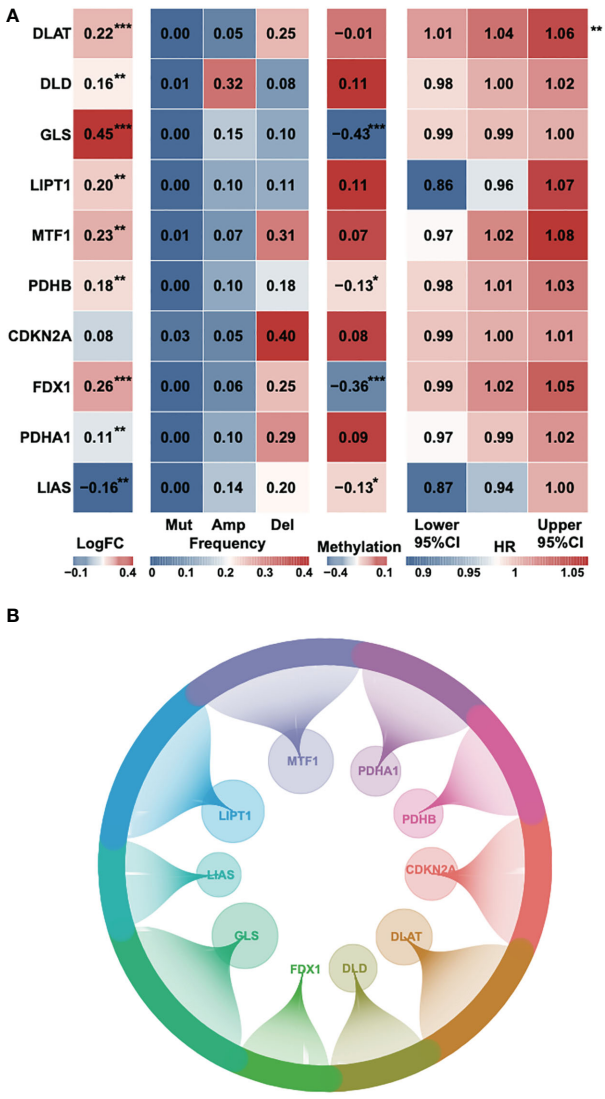


FIGURE 1
Identification of cuproptosis-related genes and corresponding lncRNAs. **(A)** The fold change, mutational frequency, methylation level and Hazard ratio of the ten cuproptosis-related genes. **(B)** The correlated lncRNAs of the ten cuproptosis-related genes. *P-value < 0.05, **P-value < 0.01, ***P-value < 0.001.

was exhibited using a circle plot (Figure 1B). Two hundred and twenty four lncRNAs were selected.

Construction of a prognostic signature based on 13 cuproptosis-related lncRNAs

To identify the most stable prognostic model, we performed Lasso regression and revealed that the 13-lncRNA and 14-lncRNA models were suitable for prognostic signature construction. Since only one lncRNA was not included in the 13-lncRNA model, we eventually selected the 13-lncRNA model as the principle of simplicity (Figure 2A). The lasso regression model of the 13 lncRNAs (lambda=0.04139117) is shown (Figure 2B). Then, we performed ridge regression and obtained the same result (Figure 2C). In addition, we introduced a new scoring system, the cuproptosis-related lncRNA score (CLS), to evaluate the risk level in HCC. By detecting the C-index, which is used for the assessment of prediction capacity and reliability (23), we uncovered that the C-index was the highest in CLS compared to stage, age and sex in both TCGA and ICGC databases (Figure 2D). The results illustrated that CLS may act as a suitable signature with a high prediction capacity in HCC. Furthermore, we also performed AUC analysis to evaluate our model in TCGA and ICGC datasets (Figures 2E, S1A), and the results indicated that CLS was better than some traditional prediction markers. Then, we calculated the CLS in each sample and ranked the order from low to high CLS. The survival status and the expression of 13 lncRNAs in each sample are illustrated in both datasets (Figures 2F, S1B). The results revealed that high CLS patients obtained a worse survival status, and that most lncRNAs in our model were highly expressed in high CLS patients except PLGLA. Afterward, we performed the RT-qPCR to detect the mRNA expression of 13 lncRNAs in the LX2 hepatic stellate cell line and Hep3B HCC cell line (Figure 2G). In addition, we pointed out that the overall survival (OS) rate was lower in high CLS patients by performing Kaplan-Meier analysis in the TCGA and ICGC databases ($P < 0.001$) (Figures 2H, S1C). We subsequently performed AUC analysis to assess the accuracy of our CLS system, the AUCs at 1-, 3-, and 5-year were 0.774, 0.685 and 0.71, respectively, in the TCGA database (Figure 2I) and 0.692, 0.729 and 0.903, respectively, in the ICGC database (Figure S1D), which showed that our CLS system was satisfactory for prognostic prediction.

Establishment of a CLS-based nomogram for HCC

We analyzed the univariate Cox regression and multivariate Cox regression in both TCGA and ICGC cohorts (Figures 3A, B) to figure out the possible independent prognostic factors. We announced that stage and CLS were the independent prognostic factors in HCC patients, and that the CLS was even better than stage. Thus, we created a CLS-based nomogram for HCC patients to predict the prognosis individually (Figure 3C). With the CLS-based nomogram,

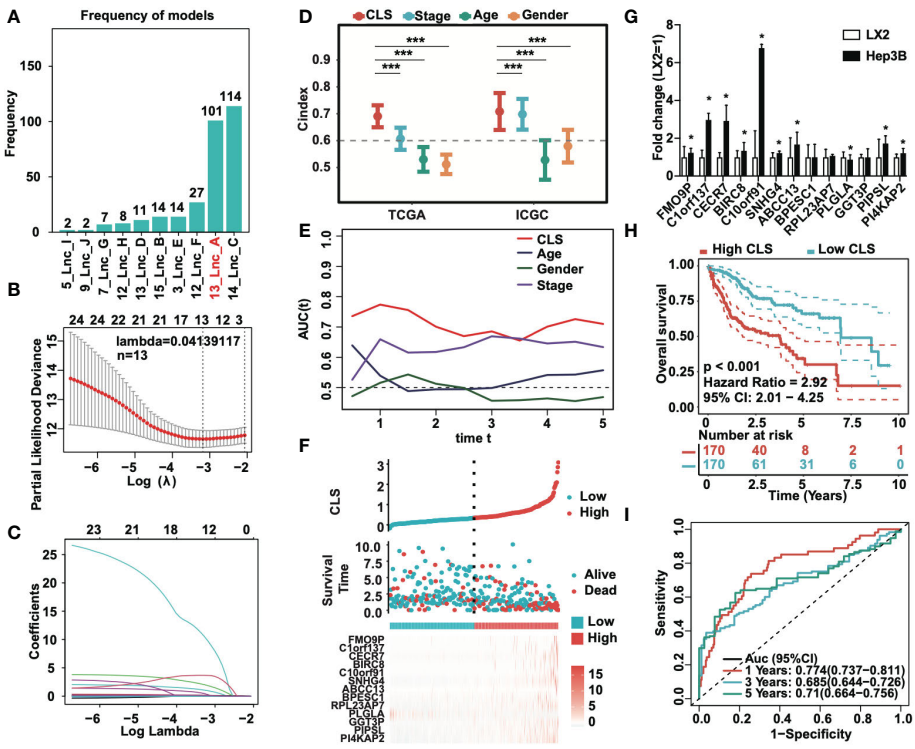


FIGURE 2
Prognostic signature based on CLS was created. **(A)** Lasso regression of the cuproptosis-related lncRNAs. **(B)** Identification of the tuning parameter in Lasso model. **(C)** The coefficients in Lasso model. **(D)** The C-index of CLS, stage, age and gender in TCGA and ICGC databases. **(E)** The AUC of CLS, age, gender and stage in TCGA. **(F)** The survival status and the expression of the 13 cuproptosis-related lncRNAs of each sample ranked from high to low CLS. **(G)** The mRNA expression of 13 cuproptosis-related lncRNAs in HCC cell line Hep3B compared to the hepatic stellate cell line LX2. **(H)** Kaplan-Meier analysis of the high and low CLS patients. **(I)** The 1-, 3- and 5-year AUC of the prognostic signature. *P-value < 0.05, ***P-value < 0.001

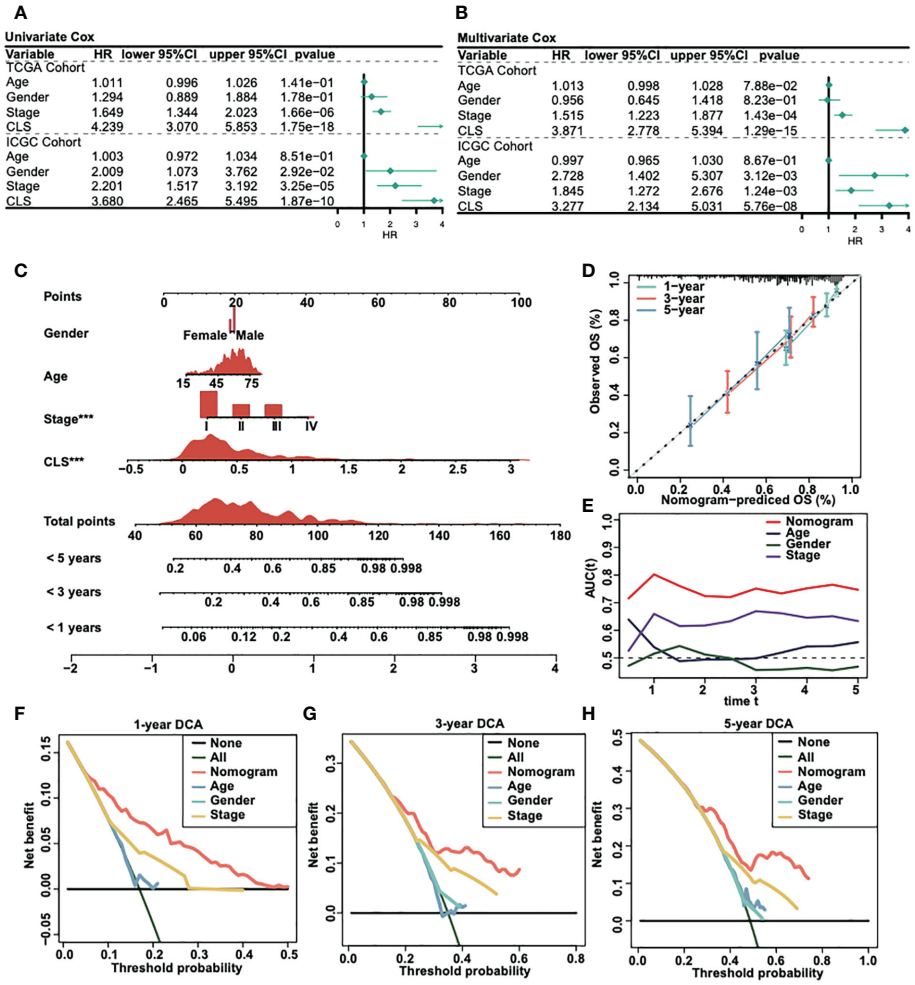


FIGURE 3
Construction of a CLS-based nomogram. **(A)** Univariate Cox regression in TCGA and ICGC cohorts. **(B)** Multivariate Cox regression in TCGA and ICGC cohorts. **(C)** Construction of a nomogram by various parameters. **(D)** Calibration curve of the CLS-based nomogram. **(E)** AUC analysis for the constructed nomogram. **(F)** One-year DCA for the nomogram. **(G)** Three-year DCA for the nomogram. **(H)** Five-year DCA for the nomogram.

we could calculate the survival rate of less than 1-, 3- and 5-year for each HCC patient. Subsequently, we created a calibration curve to assess the accuracy of our constructed nomogram (Figure 3D). The calibration curves illustrated a satisfactory capacity. And the AUC of the nomogram was the largest compared to age, sex and stage (Figure 3E), which demonstrated that the CLS-based nomogram was stable and had a high capacity for prognostic prediction. Furthermore, we performed the 1-, 3- and 5-year DCA (Figures 3F-H), DCA was used to assess the usefulness of the models we interested. We evaluated the usefulness of each model by net benefit (24). In this analysis, the CLS-based nomogram showed a larger net benefit compared to other models, the result revealed that the CLS-based nomogram was worthy of application in the clinic.

Functional analyses of the CLS model

We built a heatmap to exhibit the correlation and the significance between CLS and hallmark gene sets (Figure 4A). For example, CLS had a positive correlation with MTORCI signaling with a p-value less than 0.001. In total, the majority of cancer-related pathways were significantly related to CLS, with a positive/negative correlation. Then, after obtaining the differentially expressed genes, we performed the enrichment analysis using Metascape. The top five enriched items in high CLS samples were mitotic cell cycle, microtubule cytoskeleton organization, cell cycle checkpoints, DNA metabolic process and meiotic cell cycle (Figure 4B). The top five enriched items in the low CLS samples were monocarboxylic acid metabolic process, metabolism of lipids, drug ADME, fatty acid omega-oxidation and small molecule catabolic process (Figure 4C). Furthermore, we performed GSEA to detect the pathways enriched in samples (Figures 4D, E). Cell cycle, homologous recombination, oocyte meiosis, RNA degradation and spliceosome were significantly enriched in high CLS samples. Complement and coagulation cascades, drug metabolism cytochrome P450, fatty acid metabolism, oxidative phosphorylation and primary bile acid biosynthesis were significantly enriched in low CLS samples. Moreover, we built a heatmap to explore the expression and correlation of some pathways of interest (Figure 4F). We discovered that myeloid inflammation and MHC class I were upregulated in high CLS samples, while cytolytic activity, type I and II IFN responses were upregulated in low CLS samples. The type II IFN response, however, was negatively correlated with CLS with the most significant.

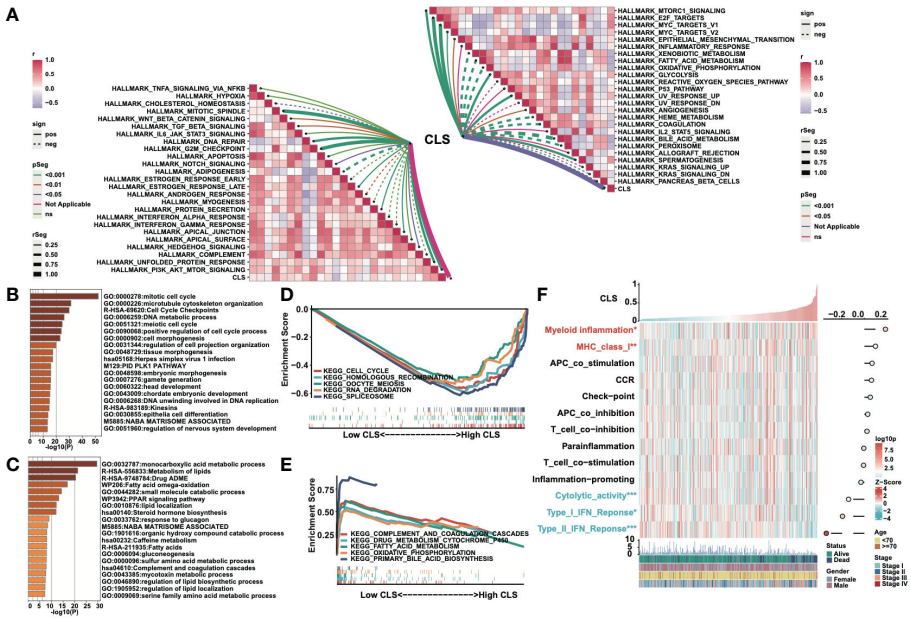


FIGURE 4
Functional analyses of the CLS model. **(A)** The correlation between CLS and the Hallmark cancer-related pathways. **(B)** The enriched items in high CLS samples in Metascape. **(C)** The enriched items in low CLS samples in Metascape. **(D)** The top five enriched items in high CLS samples by GSEA. **(E)** The top five enriched items in low CLS samples by GSEA. **(F)** The expression of the interested pathways in each sample and the correlation between interested pathways and CLS.

Immune analysis of the CLS model

First, we detected the enrichment and the correlation of the 22 TICs in samples. By generating a heatmap, we revealed that M2 macrophages, B memory cells, T regulatory cells, neutrophils, T follicular helper cells and CD4 memory activated T cells were significantly highly expressed in high CLS samples, while T gamma delta cells, NK resting cells, monocytes and M0 macrophages were upregulated in low CLS samples. Among them, M2 macrophages had the most significant positive correlation with CLS (Figure 5A). Then we calculated the immune and stromal scores and tumor purity (Figure 5B). We found that the tumor purity was higher in high CLS samples, while the immune and stromal scores were higher in low CLS samples. The results uncovered that high CLS could easily lead to tumorigenesis. In addition, we detected the relative expression of six checkpoints between high and low CLS samples (Figure 5C). CLTA-4, LAG-3, PD-1, PD-L1 and TIM-3 were highly expressed in low CLS samples, which indicated that low CLS patients had a better response to immunotherapy. Furthermore, a correlation between CLS and ESTIMATE/checkpoints was detected (Figure 5D). CLS was negatively correlated with stromal score and positively correlated with tumor purity. Nevertheless, CLS and checkpoints had a significantly negative correlation. Finally, the CTA score, HRD score and intratumor heterogeneity were evaluated. The expression of CTA was normal in the adult testis, but aberrant in several types of carcinoma (25). CTA score was associated with tumorigenesis and proliferation and was positively correlated with CLS. The CTA score was much higher in patients with high CLS (Figure 5E). The definition of HRD was that cells were uncapable to repair DNA double-strand breaks *via* homologous recombination repair pathway (26). As a characteristic of tumor tissue, HRD was positively correlated with CLS, and patients with high CLS had higher HRD score than patients with low CLS (Figure 5F). Intratumor heterogeneity, one of the reasons for the failure of cancer treatment and the determinative factor of the tumor microenvironment (27), was positively correlated with CLS. Intratumor heterogeneity was higher in high CLS patients (Figure 5G). Above all, patients with high CLS may have higher invasive and treatment resistance.

Mutational analysis of the CLS model

We detected the correlation and mutation counts in high and low CLS samples. However, we did not find any significance in all mutation counts (Figure 6A) and non-synonymous mutation counts (Figure 6B). Then, we exhibited a mutational waterfall plot in high and low CLS samples, and the top 20 genes with the most mutational frequency are listed (Figure 6C). The most frequently mutated gene was *TP53* in all samples (26%), followed by *TTN* (22%) and *CTNNB1* (23%). In addition, we compared the mutants between high and low CLS samples (Figure 6D). The results revealed that *TP53*, *CSMD1* and *RB1* had more mutants in high CLS samples. Since *TP53* was found to be the most significantly mutated gene between the two groups, we illustrated the mutational types of *TP53* in high and low CLS samples by generating a lollipop chart (Figure 6E). We found that 25.2% of mutations in high CLS samples were missense mutations, which was only 9% in low CLS samples. The percentage of other mutational types was higher in high CLS samples. Subsequently, we analyzed the mutation signatures in the two groups. By comparing five mutational signatures, we found that a difference existed between high and low CLS samples (Figures 6F, G). For instance, in signature B, many mutations occurred in the low CLS group but not in the high

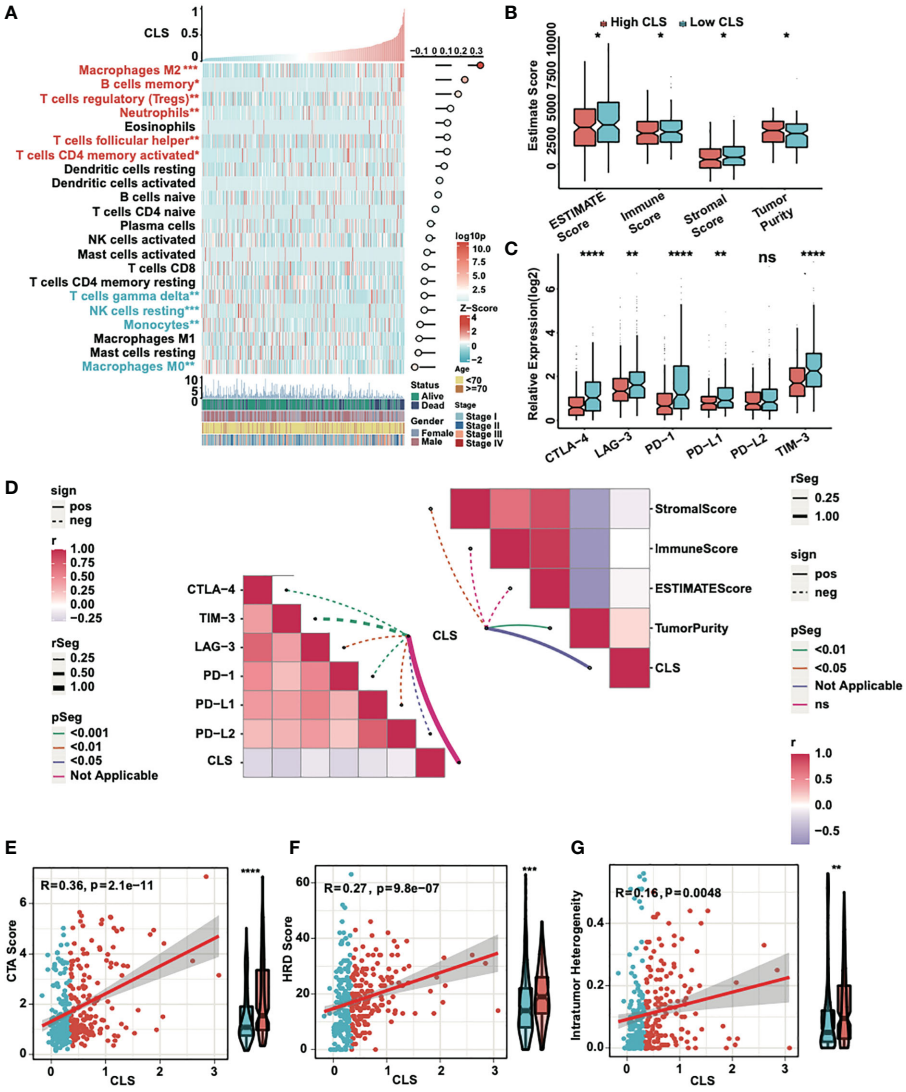


FIGURE 5
Immune analyses of the CLS model. **(A)** The expression and correlation between the TICs and CLS. **(B)** The ESTIMATE score (including immune and stromal score) and tumor purity in high and low CLS samples. **(C)** The relative expression of six immune checkpoints in high and low CLS samples. **(D)** The correlation between CLS and immune checkpoints/ESTIMATE. **(E)** The correlation between the CTA score and the CLS, and the level of CTA score in high and low CLS samples. **(F)** The correlation between the HRD score and the CLS, and the level of HRD score in high and low CLS samples. **(G)** The correlation between the intratumor heterogeneity and the CLS, and the level of the intratumor heterogeneity in high and low CLS samples. *P-value < 0.05, **P-value < 0.01, ***P-value < 0.001, ****P-value < 0.0001. ns, not significant

CLS group. In addition, we detected the frequency of amplification and deletion in each arm (Figure 6H). The results indicated that many deletions were existed in high CLS samples. In arms 3q, 12p, 12q and 22q, the mutational frequency of amplification was higher in high CLS samples but lower in the 5q and Xq arms. By detecting the total frequency of amplification (Figure 6I) and deletion (Figure 6J), we revealed that samples with high CLS showed higher amplification and deletion frequencies.

Application of the CLS model in clinical treatment

Neoantigens, which are specifically expressed in tumor tissue, have been proved to be the vital T cell-mediated immunotherapy targets for tumor patients (28). The expression of neoantigens were detected in high and low CLS samples (Figure 7A). We observed a negative correlation between CLS and neoantigens; moreover, the neoantigens was upregulated in low CLS samples. The results demonstrated that the patients with low CLS may have a satisfactory response to immunotherapy. By detecting the proliferation score, we concluded that the correlation was significantly positive between CLS and proliferation, and the proliferation score was higher in high CLS samples (Figure 7B), which indicated that high CLS patients had a higher capacity of proliferation. Next, we detected the estimated IC50 of four chemotherapeutic drugs, which are normally used in HCC treatment (Figure 7C). The results showed that patients with high CLS had better sensitivity to 5-fluorourcil, gemcitabine and doxorubicin in the TCGA dataset. The same result was obtained in

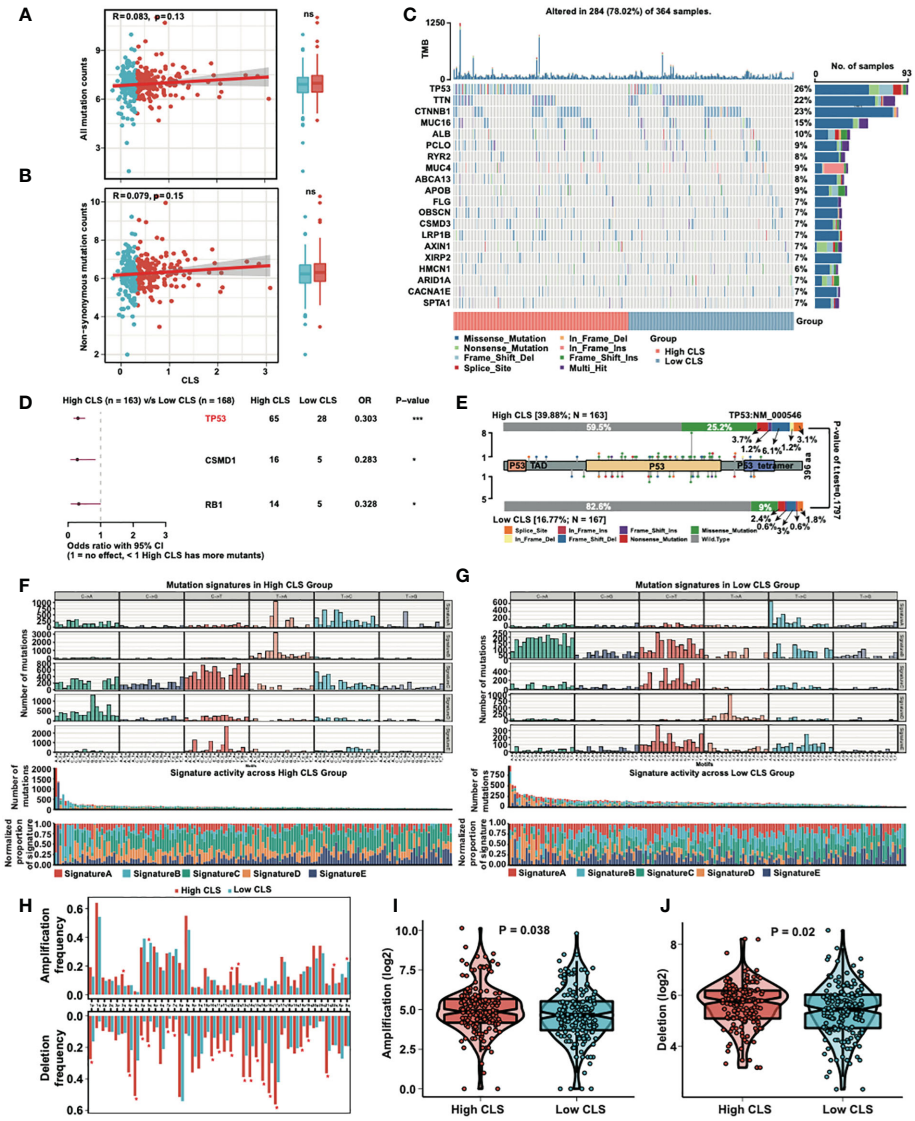


FIGURE 6
Mutational analyses of the CLS model. **(A)** The correlation between the all mutation counts and the CLS, and the number of all mutation counts in high and low CLS samples. **(B)** The correlation between the non-synonymous mutation counts and the CLS, and the number of all mutation counts in high and low CLS samples. **(C)** The waterfall plot of the top 20 altered mutation in high and low CLS samples. **(D)** The differentially mutated genes between high and low CLS samples. **(E)** The proportion and the types of the TP53 mutation in high and low CLS samples. **(F)** The number of mutations in five mutational signatures in high CLS samples. **(G)** The number of mutations in five mutational signatures in low CLS samples. **(H)** The amplification and deletion frequency in each arms between high and low CLS samples. **(I)** The total frequency of amplification in high and low CLS samples. **(J)** The total frequency of deletion in high and low CLS samples. TMB, Tumor mutational burden.

the ICGC dataset (Figure S2A). In addition, we calculated the IPS in each patient in two datasets (Figures 7D, S2B). The results showed that the low CLS patients had a higher IPS, which indicated that patients with low CLS may have a better response against immunotherapy. Moreover, the subclass mapping displayed that patients with low CLS had a better PD-1 response (Figure 7E), and a similar result was found in the ICGC dataset (Figure S2C). Furthermore, we used the TIDE algorithm to predict the immunotherapeutic sensitivity, and we detected the response rate in two subgroups in the TCGA dataset (Figure 7F). Patients with a low CLS had a better percentage of response than those with a high CLS. In the ICGC cohort, however, the response rate was higher in low CLS patients, with a P-value = 0.05 (Figure S2D). Finally, we performed the ROC analysis to compare our CLS model to five widely utilized biomarkers in the TCGA (Figure 7G) and ICGC databases (Figure S2E). The results uncovered that the CLS model had great accuracy for immunotherapeutic prediction and may act as a novel biomarker for HCC patients. Moreover, we predicted some potential small molecule drugs with related mechanisms by using MoA analysis (Figure S3), and the results may lead us to identify possible therapeutic methods for HCC patients.

Discussion

In our study, we constructed and validated a novel prognostic signature based on CLS for HCC patients. We assessed our CLS model systematically. In the functional assessment, we confirmed that

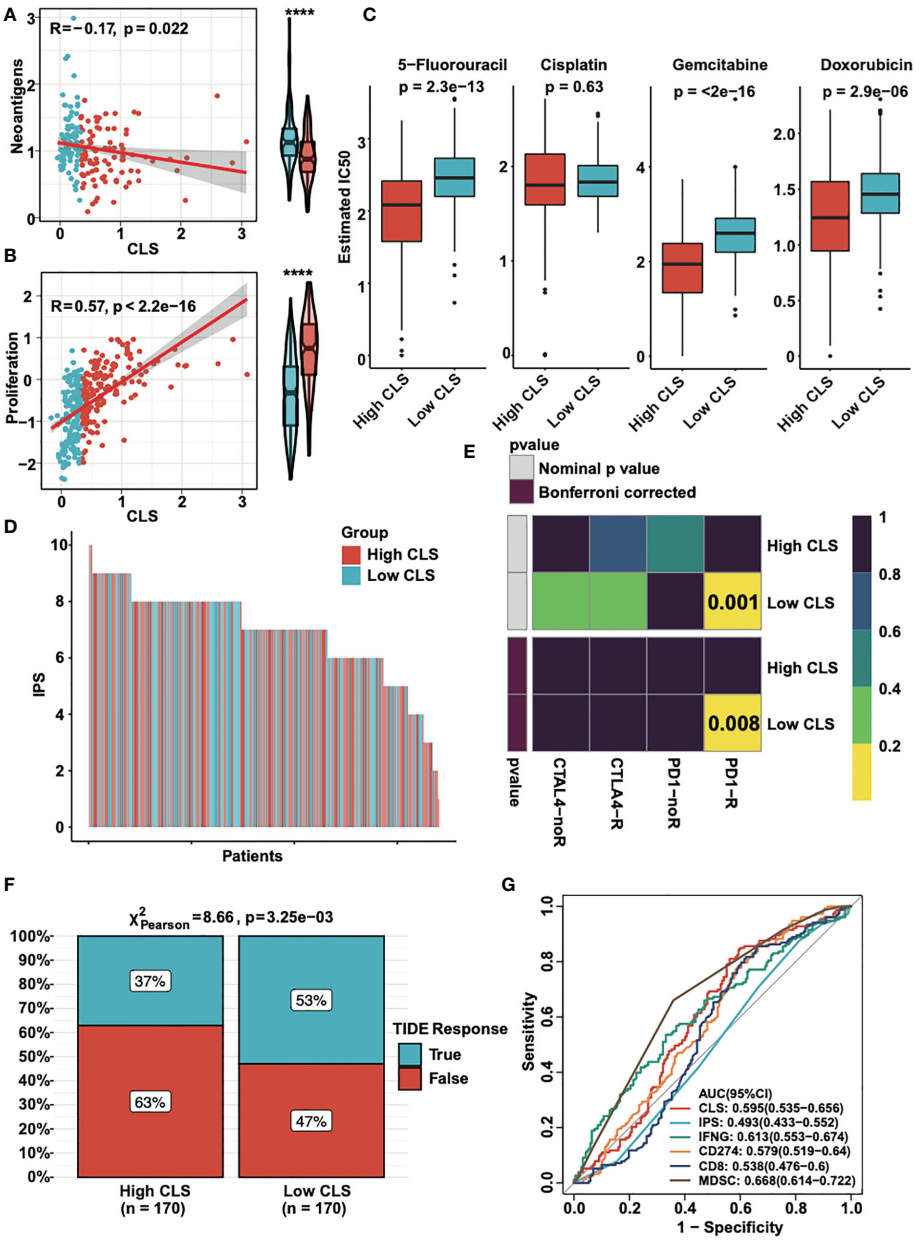


FIGURE 7
The clinical application of CLS model. **(A)** The expression and the correlation of the neoantigens in high and low CLS samples. **(B)** The expression and correlation of the proliferation score in high and low CLS samples. **(C)** The estimated IC50 of 5-fluorouracil, cisplatin, gemcitabine and doxorubicin in high and low CLS samples. **(D)** The IPS of each patients with high or low CLS. **(E)** TIDE analysis of the PD1 and CTLA4 response in patients with high and low CLS. **(F)** The proportion of the TIDE response in high and low CLS patients. **(G)** The AUC analysis of the CLS and biomarkers. IPS, Immunophenoscore.

CLS had a high correlation with cancer-related pathways. In addition, cell cycle and immune related pathways were enriched. By performing immune analysis, we announced that the tumor characteristics were more obvious in high CLS samples, which was related to invasion and resistance to the treatment. In mutational evaluation, more mutational frequency was found in high CLS samples, and the same went for amplification and deletion. We utilized our CLS model for predicting the clinical treatment response. We revealed that 5-florouracil, gemcitabine and doxorubicin had more sensitivity in high CLS patients. Nevertheless, patients with low CLS showed a better response to immunotherapy.

Thirteen lncRNAs were selected and was verified to be highly expressed in hepatocellular carcinoma. A previous article also reported that C10orf91 was upregulated in HCC and correlated with poor prognosis (29). One published article demonstrated that the lncRNA CECR7 was upregulated in HCC and related to OS (30). Other published research uncovered that lncRNA SNHG4 was highly expressed in liver cancer tissues compared to normal liver tissues; moreover, the expression of lncRNA SNHG4 was associated with OS (31). LncRNA BPESC1 was also reported to correlate with OS, and HCC patients with high expression of BPESC1 had worse OS (32).

By performing correlation analysis, we revealed that CLS was highly correlated with some cancer-related pathways, such as mitotic spindle, DNA repair, G2/M checkpoint, PI3K-AKT-MTOR signaling, MTORC1 signaling, E2F targets and MYC targets. The source of our CLS model was the cuproptosis-related lncRNAs, which had a high correlation with the level of copper. Currently, studies have proven that the level of copper correlates with various biochemical processes. One published article pointed out that a high level of copper enhanced the drug resistance and was involved in DNA damage repair in cancer cells (33). One previous article demonstrated that copper accumulation reduced the proportion of cells in G2/M phase *via* Ras/PI3K/Akt signaling (34). In addition, another article reported that a novel copper nanocomplex inhibited cell proliferation and caused the cell death *via* the PI3K/AKT/mTOR signaling pathway in cervical cancer cells (35). These results were consistent with our findings.

We analyzed the tumor microenvironment and the enrichment of TICs in each sample. Many tumor immune cells were enriched. M2 macrophages, for example, were reported to have tumor-promoting activities promoting cell proliferation, migration, angiogenesis and immunosuppression, subsequently resulting in poor outcome of HCC (36). This result coincided with our findings that M2 macrophages were significantly enriched in high CLS samples, which had unfavorable outcomes of HCC. Previous research illustrated that infiltration of regulatory T cells inhibit the anti-tumor immune response and is correlated to unsatisfactory prognosis (37). Neutrophils have been proved to promote the progress of tumorigenesis and associated to poor prognosis (38). Our result showed that regulatory T cells and neutrophils were enriched in high CLS patients, which was a good explanation of high CLS patients with a poor overall survival. According to the analysis of the tumor microenvironment, the tumor purity was higher and the immune and stromal scores were lower in the high CLS samples. The result was corresponded to the findings that the high CLS patients had higher progression of HCC and worse survival status. By detecting the relative expression, we revealed that immune checkpoints were highly expressed in low CLS samples except PD-L2. The results indicated that the checkpoint inhibitors may have a better response in low CLS patients. In addition, in the analysis of clinical application of this article, we predicted the effect of chemotherapy and immunotherapy in high and low CLS patients. The results demonstrated that the chemotherapy was sensitive in high CLS patients, while immunotherapy was better in low CLS patients. The reversed result can be explained by the treatment chosen according to the progress of the HCC. Low CLS patients may be in the early stage of the HCC, patients may get more benefits from immunotherapy because of the easier mobilization of the immune system. However, the effect of immunotherapy may decreased in advanced HCC patients because of the immune escape and T cell exhaustion. Moreover, in TIDE analysis, the response of PD-1 and CTLA4 was better in low CLS patients with HCC. Currently, some immunotherapy trials have been performed, which have shown similar results. One of them demonstrated that anti-CTLA-4 monoclonal antibody had promising outcomes in HCC patients (39). Another study reported that an antibody against PD-1 was well tolerated and had an acceptable objective response rate (40). In addition, the combination of an anti-CTLA-4 monoclonal antibody (tremelimumab) and an anti-PD-L1 monoclonal antibody (durvalumab) was found to be tolerable and enhanced the antitumor effect (41). Overall, the immunotherapy is a potential method for HCC patients, especially for the patients with low CLS.

We know that instability of the gene is one of the characteristic of most carcinomas. Mutation drives the occurrence and development of the most type of cancers (42). In our study, we revealed three genes that had more mutations in high CLS samples. *TP53*, which is the one of the most frequently mutated genes in HCC, plays a vital role in apoptosis and cell cycle regulation (43). Studies have indicated that *TP53* mutation may cause cancer progression (44). Moreover, patients with mutated *TP53* had worse OS and relapse-free survival times (45). *CSMD1* is considered to be a tumor suppressor gene in many types of cancer, such as breast cancer (46), colorectal cancer (47), gastric cancer (48) and HCC (49); thus, the mutation of the *CSMD1* may cause the proliferation of the cancer. One published article revealed that the mutation of *CSMD1* may promote the progression of esophageal cancer (50). Interestingly, one article demonstrated that *CSMD1* mutation co-occurred with *TP53* mutation (51). In our research, we also detected the concurrent mutation of *TP53* and *CSMD1* in high CLS samples. As a tumor suppressor gene, *RB1* is a negative regulator in the progression of the cell cycle *via* the regulation of the E2F transcription factors (52, 53). Mutation of *RB1* may cause cancer genesis (54). Together, the result was sensible that the patients with a high frequency of mutated genes *TP53*, *CSMD1* and *RB1* may have a worse survival status.

We were aware of the study having some limitations. First, our results were obtained based on the online databases, and clinical trials with large samples are necessary. Second, we could not find the immunotherapy information for HCC; instead, we verified the results in a melanoma cohort. Thus, a novel HCC cohort is needed for the further analyses.

In this article, we established and verified a novel prognostic CLS model by machine learning. Meanwhile, We performed systematic analyses, including function, mutation, immunity and clinical application, to ensure the stability and value of the constructed model for the purpose of utilization of our model in the clinical assessment and treatment.

Data availability statement

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

Author contributions

DL, JL: designed the study, analyzed the data and wrote the manuscript. HC, QM, FW, FX: assisted with the data analyses. YH: Reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank the researchers who published the data online for use. We thank the creators who built and developed the R package.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1097075/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Prognostic signature based on CLS in ICGC database. A. The AUC of CLS, age, gender and stage in ICGC database. B. The survival status and the expression of the 13 cuproptosis-related lncRNAs of each sample ranked from high to low CLS in ICGC database. C. Kaplan-Meier analysis of the high and low CLS patients in ICGC database. D. The 1-, 3- and 5-year AUC of the CLS-based prognostic signature in ICGC.

SUPPLEMENTARY FIGURE 2
The clinical application of CLS model in ICGC database. A. The estimated IC50 of 5-fluorouracil, cisplatin, gemcitabine and doxorubicin in high and low CLS samples in ICGC. B. The IPS of each patients with high or low CLS in ICGC. C. TIDE analysis of the PD1 and CTLA4 response in patients with high and low CLS in ICGC. D. The proportion of the TIDE response in high and low CLS patients in ICGC. E. The AUC analysis of the CLS and biomarkers in ICGC.

SUPPLEMENTARY FIGURE 3
MoA analysis in HCC.

References

1. Degasperi E, Colombo M. Distinctive features of hepatocellular carcinoma in non-alcoholic fatty liver disease. *Lancet Gastroenterol Hepatol* (2016) 1:156–64. doi: 10.1016/S2468-1253(16)30018-8

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

3. Marasco G, Colecchia A, Colli A, Ravaioli F, Casazza G, Bacchi Reggiani ML, et al. Role of liver and spleen stiffness in predicting the recurrence of hepatocellular carcinoma after resection. *J Hepatol* (2019) 70:440–8. doi: 10.1016/j.jhep.2018.10.022

4. Bruix J, Qin S, Merle P, Granito A, Huang YH, Bodoky G, et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* (2017) 389:56–66. doi: 10.1016/S0140-6736(16)32453-9

5. Liu Z, Lin Y, Zhang J, Zhang Y, Li Y, Liu Z, et al. Molecular targeted and immune checkpoint therapy for advanced hepatocellular carcinoma. *J Exp Clin Cancer Res: CR* (2019) 38:447. doi: 10.1186/s13046-019-1412-8

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

7. Kahlson MA, Dixon SJ. Copper-induced cell death. *science* (2022) 375:1231–2. doi: 10.1126/science.abo3959

8. Babak MV, Ahn D. Modulation of intracellular copper levels as the mechanism of action of anticancer copper complexes: Clinical relevance. *Biomedicines* (2021) 9(8):852. doi: 10.3390/biomedicines9080852

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

10. Zhang S, Qin C, Cao G, Xin W, Feng C, Zhang W. Systematic analysis of long noncoding RNAs in the senescence-accelerated mouse prone 8 brain using RNA sequencing, molecular therapy. *Nucleic Acids* (2016) 5:e343. doi: 10.1038/mtna.2016.57

11. Bhan A, Mandal SS, LncRNA HOTAIR. A master regulator of chromatin dynamics and cancer. *Biochim Biophys Acta* (2015) 1856:151–64. doi: 10.1016/j.bbcan.2015.07.001

12. Bhan A, Mandal SS. Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. *ChemMedChem* (2014) 9:1932–56. doi: 10.1002/cmdc.201300534

13. Goyal B, Yadav SRM, Awasthee N, Gupta S, Kunnumakkara AB, Gupta SC. Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. *Biochim Biophys Acta Rev Cancer* (2021) 1875:188502. doi: 10.1016/j.bbcan.2021.188502

14. Sarfi M, Abbastabar M, Khalili E. Long noncoding RNAs biomarker-based cancer assessment. *J Cell Physiol* (2019) 234:16971–86. doi: 10.1002/jcp.28417

15. Wang G, Zhang ZJ, Jian WG, Liu PH, Xue W, Wang TD, et al. Novel long noncoding RNA OTUD6B-AS1 indicates poor prognosis and inhibits clear cell renal cell carcinoma proliferation *via* the wnt/β-catenin signaling pathway. *Mol Cancer* (2019) 18:15. doi: 10.1186/s12943-019-0942-1

16. Liang JY, Wang DS, Lin HC, Chen XX, Yang H, Zheng Y, et al. A novel ferroptosis-related gene signature for overall survival prediction in patients with hepatocellular carcinoma. *Int J Biol Sci* (2020) 16:2430–41. doi: 10.7150/ijbs.45050

17. McDermott DF, Huseni MA, Atkins MB, Motzer RJ, Rini BI, Escudier B, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med* (2018) 24:749–57. doi: 10.1038/s41591-018-0053-3

18. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity* (2018) 48:812–830.e814. doi: 10.1016/j.immuni.2018.03.023

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

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

21. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance, sci transl med. *Sci Transl Med* (2017) 9(379): eaah3560. doi: 10.1126/scitranslmed.aah3560

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

23. Liu C, Wang T, Yang J, Zhang J, Wei S, Guo Y, et al. Distant metastasis pattern and prognostic prediction model of colorectal cancer patients based on big data mining. *Front Oncol* (2022) 12:878805. doi: 10.3389/fonc.2022.878805

24. Hijazi Z, Oldgren J, Lindbäck J, Alexander JH, Connolly SJ, Eikelboom JW, et al. The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. *Lancet* (2016) 387:2302–11. doi: 10.1016/S0140-6736(16)00741-8

25. Kulkarni P, Shiraishi T, Rajagopalan K, Kim R, Mooney SM, Getzenberg RH. Cancer/testis antigens and urological malignancies. *Nat Rev Urol* (2012) 9:386–96. doi: 10.1038/nrurol.2012.117

26. Stewart MD, Merino Vega D, Arend RC, Baden JF, Barbash O, Beaubier N, et al. Homologous recombination deficiency: Concepts, definitions, and assays. *Oncol* (2022) 27:167–74. doi: 10.1093/oncolo/oyab053

27. Wolf Y, Samuels Y. Intratumor heterogeneity and antitumor immunity shape one another bidirectionally. *Clin Cancer Res: Off J Am Assoc Cancer Res* (2022) 28:2994–3001. doi: 10.1158/1078-0432.CCR-21-1355

28. Ma W, Pham B, Li T. Cancer neoantigens as potential targets for immunotherapy. *Clin Exp Metastasis* (2022) 39:51–60. doi: 10.1007/s10585-021-10091-1

29. Gong A, Luo X, Tan Y, Chen H, Luo G. High expression of C10orf91 and LINC01224 in hepatocellular carcinoma and poor prognosis. *Am J Transl Res* (2022) 14:2567–79.

30. Zhang J, Fan D, Jian Z, Chen GG, Lai PB. Cancer specific long noncoding RNAs show differential expression patterns and competing endogenous RNA potential in hepatocellular carcinoma. *PloS One* (2015) 10:e0141042. doi: 10.1371/journal.pone.0141042

31. Jiao Y, Li Y, Jia B, Chen Q, Pan G, Hua F, et al. The prognostic value of lncRNA SNHG4 and its potential mechanism in liver cancer, bioscience reports. *Biosci Rep.* (2020) 40(1):BSR20130729. doi: 10.1042/BSR20190729

32. Gu X, Li H, Sha L, Zhao W. Construction and comprehensive analyses of a competing endogenous RNA network in tumor-Node-Metastasis stage I hepatocellular carcinoma. *BioMed Res Int* (2020) 2020:5831064. doi: 10.1155/2020/5831064

33. Jin J, Ma M, Shi S, Wang J, Xiao P, Yu HF, et al. Copper enhances genotoxic drug resistance *via* ATOX1 activated DNA damage repair. *Cancer Lett* (2022) 536:215651. doi: 10.1016/j.canlet.2022.215651

34. Liu H, Deng H, Jian Z, Cui H, Guo H, Fang J, et al. Copper exposure induces hepatic G0/G1 cell-cycle arrest through suppressing the Ras/PI3K/Akt signaling pathway in mice. *Ecotoxicol Environ Saf* (2021) 222:112518. doi: 10.1016/j.ecoenv.2021.112518

35. Chen H, Feng X, Gao L, Mickymaray S, Paramasivam A, Abdulaziz Alfaiz F, et al. Inhibiting the PI3K/AKT/mTOR signalling pathway with copper oxide nanoparticles from houttuynia cordata plant: attenuating the proliferation of cervical cancer cells. *Artif Cells Nanomed Biotechnol* (2021) 49:240–9. doi: 10.1080/21691401.2021.1890101

36. Capece D, Fischietti M, Verzella D, Gaggiano A, Cicciarelli G, Tessitore A, et al. The inflammatory microenvironment in hepatocellular carcinoma: a pivotal role for tumor-associated macrophages. *BioMed Res Int* (2013) 2013:187204. doi: 10.1155/2013/187204

37. Tanaka A, Sakaguchi S. Regulatory T cells in cancer immunotherapy. *Cell Res* (2017) 27:109–18. doi: 10.1038/cr.2016.151

38. Teijeira A, Garasa S, Ochoa MC, Villalba M, Olivera I, Cirella A, et al. IL8, neutrophils, and NETs in a collusion against cancer immunity and immunotherapy. *Clin Cancer Res: Off J Am Assoc Cancer Res* (2021) 27:2383–93. doi: 10.1158/1078-0432.CCR-20-1319

39. Sangro B, Gomez-Martin C, de la Mata M, Iñarrairaegui M, Garralda E, Barrera P, et al. A clinical trial of CTLA-4 blockade with tremelimumab in patients with hepatocellular carcinoma and chronic hepatitis c. *J Hepatol* (2013) 59:81–8. doi: 10.1016/j.jhep.2013.02.022

40. El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* (2017) 389:2492–502. doi: 10.1016/S0140-6736(17)31046-2

41. Kelley RK, Sangro B, Harris W, Ikeda M, Okusaka T, Kang YK, et al. Safety, efficacy, and pharmacodynamics of tremelimumab plus durvalumab for patients with unresectable hepatocellular carcinoma: Randomized expansion of a phase I/II study. *J Clin Oncol: Off J Am Soc Clin Oncol* (2021) 39:2991–3001. doi: 10.1200/JCO.20.03555

42. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability–an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* (2010) 11:220–8. doi: 10.1038/nrm2858

43. Lai PB, Chi TY, Chen GG. Different levels of p53 induced either apoptosis or cell cycle arrest in a doxycycline-regulated hepatocellular carcinoma cell line *in vitro*. *Apoptosis* (2007) 12:387–93. doi: 10.1007/s10495-006-0571-1

44. Brosh R, Rotter V. When mutants gain new powers: news from the mutant p53 field, nature reviews. *Cancer* (2009) 9:701–13. doi: 10.1038/nrc2693

45. Liu J, Ma Q, Zhang M, Wang X, Zhang D, Li W, et al. Alterations of TP53 are associated with a poor outcome for patients with hepatocellular carcinoma: evidence from a systematic review and meta-analysis. *Eur J Cancer* (2012) 48:2328–38. doi: 10.1016/j.ejca.2012.03.001

46. Gialeli C, Tuysuz EC, Staaf J, Guleed S, Paciorek V, Mörgelin M, et al. Complement inhibitor CSMD1 modulates epidermal growth factor receptor oncogenic signaling and sensitizes breast cancer cells to chemotherapy. *J Exp Clin Cancer Res: CR* (2021) 40:258. doi: 10.1186/s13046-021-02042-1

47. Shull AY, Clendenning ML, Ghoshal-Gupta S, Farrell CL, Vangapandu HV, Dudas L, et al. Somatic mutations, allele loss, and DNA methylation of the cub and sushi multiple domains 1 (CSMD1) gene reveals association with early age of diagnosis in colorectal cancer patients. *PloS One* (2013) 8:e58731. doi: 10.1371/journal.pone.0058731

48. Chen XL, Hong LL, Wang KL, Liu X, Wang JL, Lei L, et al. Deregulation of CSMD1 targeted by microRNA-10b drives gastric cancer progression through the NF-κB pathway. *Int J Biol Sci* (2019) 15:2075–86. doi: 10.7150/ijbs.23802

49. Zhao Y, Yang B, Chen D, Zhou X, Wang M, Jiang J, et al. Combined identification of ARID1A, CSMD1, and SENP3 as effective prognostic biomarkers for hepatocellular carcinoma. *Aging* (2021) 13:4696–712. doi: 10.18632/aging.202586

50. Fan X, Song J, Fan Y, Li J, Chen Y, Zhu H, et al. CSMD1 mutation related to immunity can be used as a marker to evaluate the clinical therapeutic effect and prognosis of patients with esophageal cancer. *Int J Gen Med* (2021) 14:8689–710. doi: 10.2147/IJGM.S338284

51. Di Nardo L, Pellegrini C, Di Stefani A, Ricci F, Fossati B, Del Regno L, et al. Molecular alterations in basal cell carcinoma subtypes. *Sci Rep* (2021) 11:13206. doi: 10.1038/s41598-021-92592-3

52. Nevins JR. The Rb/E2F pathway and cancer. *Hum Mol Genet* (2001) 10:699–703. doi: 10.1093/hmg/10.7.699

53. Sherr CJ. Principles of tumor suppression. *Cell* (2004) 116:235–46. doi: 10.1016/S0092-8674(03)01075-4

54. Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer, nature reviews. *Cancer* (2019) 19:326–38. doi: 10.1038/s41568-019-0143-7

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/fimmu.2023.1097472

OPEN ACCESS

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

REVIEWED BY
Shengjie Chai,
Uber Technologies, Inc., United States
Yutao Wang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Shaobo Jiang
jiangshaobo@sdu.edu.cn
Hanbo Wang
wanghanbo0709@163.com

†These authors have contributed equally to this work

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

RECEIVED 13 November 2022
ACCEPTED 09 January 2023
PUBLISHED 25 January 2023

CITATION
Sun Z, Jing C, Zhan H, Guo X, Suo N, Kong F, Tao W, Xiao C, Hu D, Wang H and Jiang S (2023) Identification of tumor antigens and immune landscapes for bladder urothelial carcinoma mRNA vaccine.
Front. Immunol. 14:1097472.
doi: 10.3389/fimmu.2023.1097472

Identification of tumor antigens and immune landscapes for bladder urothelial carcinoma mRNA vaccine

Zhuolun Sun[1†], Changying Jing[2,3†], Hailun Zhan[1†], Xudong Guo[4†], Ning Suo[4], Feng Kong[4], Wen Tao[5], Chutian Xiao[6], Daoyuan Hu[6], Hanbo Wang[4*] and Shaobo Jiang[4*]

[1]Department of Urology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [2]Faculty of Medicine, Ludwig Maximilian University of Munich (LMU), Munich, Germany, [3]Institute of Diabetes and Regeneration, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [4]Department of Urology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China, [5]Department of Urology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [6]Department of Urology, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Background: Bladder urothelial carcinoma (BLCA) is associated with high mortality and recurrence. Although mRNA-based vaccines are promising treatment strategies for combating multiple solid cancers, their efficacy against BLCA remains unclear. We aimed to identify potential effective antigens of BLCA for the development of mRNA-based vaccines and screen for immune clusters to select appropriate candidates for vaccination.

Methods: Gene expression microarray data and clinical information were retrieved from The Cancer Genome Atlas and GSE32894, respectively. The mRNA splicing patterns were obtained from the SpliceSeq portal. The cBioPortal for Cancer Genomics was used to visualize genetic alteration profiles. Furthermore, nonsense-mediated mRNA decay (NMD) analysis, correlation analysis, consensus clustering analysis, immune cell infiltration analysis, and weighted co-expression network analysis were conducted.

Results: Six upregulated and mutated tumor antigens related to NMD, and infiltration of APCs were identified in patients with BLCA, including HP1BP3, OSBPL9, SSH3, ZCCHC8, FANCI, and EIF4A2. The patients were subdivided into two immune clusters (IC1 and IC2) with distinct clinical, cellular and molecular features. Patients in IC1 represented immunologically 'hot' phenotypes, whereas those in IC2 represented immunologically 'cold' phenotypes. Moreover, the survival rate was better in IC2 than in IC1, and the immune landscape of BLCA indicated significant inter-patient heterogeneity. Finally, CALD1, TGFB3, and ANXA6 were identified as key genes of BLCA through WGCNA analysis, and their mRNA expression levels were measured using qRT-PCR.

Conclusion: HP1BP3, OSBPL9, SSH3, ZCCHC8, FANCI, and EIF4A2 were identified as potential antigens for developing mRNA-based vaccines against BLCA, and patients in IC2 might benefit more from vaccination.

KEYWORDS

bladder urothelial carcinoma, immune clusters, immune landscape, mRNA vaccine, tumor antigens

Introduction

Bladder cancer (BC) is one of the most prevalent cancers worldwide. An estimated 83,730 new BC cases and 17,200 BC-related deaths were reported in the United States of America in 2021 (1). Increasing evidence implicates that BC is a clinically and genetically heterogeneous disease that is characterized by poor therapeutic efficacy and rapid tumor progression (2–4). More than 90% of BC cases are histologically categorized as bladder urothelial carcinoma (BLCA), which can present as non-muscle-invasive (75%) and muscle-invasive (25%) BC (NMIBC and MIBC, respectively) (4). Although the 5-year survival rate is as high as 90%, patients with NMIBC often relapse and progress to MIBC. Patients with MIBC usually have a poor prognosis because of aggressive metastasis and delayed diagnosis (5, 6). In addition to surgery, platinum-based chemotherapy is the first-line treatment for advanced or metastatic BLCA, which may extend median overall survival (OS) by approximately 1 year with a limited response rate (7, 8). However, non-responsive patients may lose the opportunity to receive additional therapeutic intervention for tumor development. Immune checkpoint blockade has recently emerged as a valuable treatment option for MIBC; however, its clinical benefits are observed only in a small proportion of patients (9, 10). These studies highlight the need for novel therapeutic strategies that may improve the clinical outcomes of patients with BLCA.

In the context of the ongoing coronavirus disease 2019 (COVID-19) pandemic, development of vaccines has been recognized as the top priority of pharmaceutical and biotechnology industries worldwide (11, 12). Therapeutic cancer vaccines are designed to reprogram the immune system of patients, specifically cytotoxic T lymphocytes, to safely and efficiently eliminate cancer cells (13). Antigens used for developing cancer vaccines include whole tumor cells, peptides, viral vectors, dendritic cells, DNA or RNA (14). The significant technological innovation and development investment in the last decade have made mRNA an optimal vehicle to carry tumor-specific antigens (15). Furthermore, mRNA-based vaccines are promising strategies for cancer therapy owing to their high efficacy, rapid development capabilities, safe administration and low-cost manufacturing as compared with other vaccine types (15–17). Recent preclinical and clinical trials have verified the viability of mRNA vaccines encoding tumor-specific antigens to combat multiple cancers, including lung cancer (18), prostate cancer (19), melanoma (20) and other cancers (15). However, tumor-specific or tumor-associated antigens (TSAs or TAAs, respectively) vary greatly among individuals. Recognizing immunogenic tumor neoantigens and relieving inhibitory tumor microenvironment (TME) are the main obstacles to developing mRNA vaccines against BLCA (16).

Several studies have shown that disruption of transcriptional regulation at different stages can lead to the accumulation of a large number of abnormal transcripts in cancer cells (21). These aberrant transcripts usually harbor premature termination codon; even if they are transcribed, they may be subsequently degraded by an mRNA surveillance pathway termed nonsense-mediated mRNA decay (NMD) (22). A relationship between NMD and tumor immunity is frequently observed and recognized as an attractive target for cancer therapy in some cases (22, 23). Recent studies have demonstrated that transcripts that harbor aberrant splicing patterns and frameshift mutations express antigenic peptides, with the disruption of normal NMD functionality (24). Therefore, it is important to perform a comprehensive analysis of alternative splicing (AS) patterns and NMD for developing individualized mRNA vaccines against tumors.

In this study, we investigated the potential BLCA antigens for developing mRNA vaccines and elucidated the immune landscape to identify eligible patients for vaccination. We confirmed six tumor antigens relevant to NMD, AS and antigen-presenting cell infiltration and defined two immune clusters of patients with BLCA. The two immune clusters presented distinct clinical, molecular and tumor immune microenvironment (TIME) characteristics, which were consistent in TCGA and GSE32894 cohorts. In addition, we assessed the immune landscape of BLCA by analyzing the expression profile of immune-related genes in individual patients. Finally, we identified CALD1, TGFB3, and ANXA6 as key genes of BLCA through WGCNA analysis and measured their mRNA expression levels using qRT-PCR. Therefore, the present study provides information regarding the complicated TIME in patients with BLCA and offers a reliable reference for developing and administering cancer vaccines.

Materials and methods

Identification of tumor antigens

Data extraction

The RNA-sequencing data and clinical information of patients with BLCA were retrieved from The Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/tcga/) and Gene Expression Omnibus (GEO) (GSE32894, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32894). The inclusion criteria were as follows: (1) RNAs that

were detectable in >30% of the samples and (2) OS time > 30 days. The detailed clinical characteristics of the patients enrolled in this study are summarized in Supplementary Table 1. The original gene IDs of the respective datasets were transformed into the corresponding gene symbols based on annotation information on the platform. In addition, the expression profiles were indicated as transcripts per millions for subsequent analyzes. The batch effects between different datasets were corrected using the 'ComBat' method.

Profiling of AS events

The mRNA splicing patterns of 18 healthy patients and 399 patients with BLCA were retrieved from TCGA SpliceSeq portal (https://bioinformatics.mdanderson.org/TCGA SpliceSeq/). The percent spliced-in index (PSI) value, ranging from 0 to 1, is the ratio between reads including or excluding the designated exons and indicates the efficiency of certain splicing events (25). To improve the reliability of the results, the primary PSI data that contained vacancy values were removed. The overlapping sets between different AS events were visualized using UpSet plots drawn using the Upset R package (26). To determine cancer-associated AS events (CASEs) in BLCA, we compared the PSI values of AS events between normal and BLCA tissues, and the P-value was adjusted using the Benjamini–Hochberg (BH) method. AS events with an absolute log2 (fold change) ≥ 1 and an adjusted P-value < 0.05 were considered statistically significant.

cBioPortal analysis

The 'maftools' R package and cBioPortal for Cancer Genomics (cBioPortal, https://www.cbioportal.org/) were used to retrieve the mutation data from TCGA database to compare and visualize potential genetic variations in each sample (27). Statistical significance was defined as P-value < 0.05.

NMD analysis

We identified genes with abnormally upregulated AS events and frameshift mutations as candidate antigens against BLCA. Many studies have highlighted the relationship between NMD and tumor immunity and revealed the potential of NMD as a therapeutic target for cancers in some cases (24). Further NMD analysis may assist in developing individualized tumor vaccines, such as for melanoma (23). Patients with BLCA were divided into the low- or high-expression groups according to the median expression of NMD factors (UPF1, UPF2, UPF3A and UPF3B). Subsequently, the expression levels and AS events of candidate genes between the two groups were analyzed. P-value was calculated using the 'ggpubr' package (stat_compare_means function), with P-value < 0.05 as the threshold.

TIMER analysis

The Tumor Immune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/) is a public online database that allows systematic evaluation of the immune infiltration data for different cancers from TCGA (28). In this study, TIMER was used to assess and demonstrate the Spearman correlation between the abundance of tumor-infiltrating immune cells (TIICs) and the expression of tumor antigens. Purity adjustments were performed using Spearman's correlation analysis. Statistical significance was defined as P-value < 0.05.

Prediction of the peptides of antigens for BLCA samples

The Cancer Immunome Atlas (TCIA, https://tcia.at/home) was used to screen for peptides of neoantigens for each BLCA sample with default parameters. A list of peptides was obtained by selecting the 'Neoantigens' tab after inputting candidate genes for antigens in the TCIA filter.

Identification of immune clusters

Immune-related gene data extraction

A total of 1894 immune-related genes (IRGs) were retrieved from The Immunology Database and Analysis Portal (ImmPort, https://www.immport.org/shared/home) (29) and a study of Charoentong et al. (30) for both discovery (TCGA) and validation (GSE32894) cohorts. We choose this data matrix as the validation cohort because it represents one of the most comprehensive datasets, including the most survival data as well as clinical stage and tumor grade. After filtering these candidate IRGs associated with prognosis, 233 prognostic genes in 399 BLCA samples and 371 prognostic genes in 224 BLCA samples were identified in the discovery and validations cohort, respectively.

Identification and validation of immune clusters

Consensus clustering was performed to determine robust immune clusters according to the expression profiles of 233 prognostic IRGs using the 'ConsensusClusterPlus' package. Specifically, the algorithm of partition around medoids was used for 500 bootstraps, with 80% patients being resampled and '1-Pearson correlation' as the distance metric in TCGA cohort. The cluster number was tested from 2 to 9, and the optimal one was identified to yield the least ambiguous cluster assignments across clustering permutations and the most stable consensus matrix. The immune clusters were further confirmed in the GSE32894 cohort using similar settings. The coherence of the identified immune clusters was quantified in the two cohorts *via* in-group proportion and Spearman's correlation analyzes. The prognostic significance of these immune clusters in the discovery cohort was estimated *via* Kaplan–Meier survival analysis and validated in the validation cohort. Clinical features of these immune clusters, including stage, grade, clinical T stage and sex, were assessed using the 'ggplot2' R package.

Molecular, cellular and immunological features of the BLCA immune clusters

Tumor mutation burden (TMB) and mutated gene counts were visualized among the BLCA immune subtypes using the 'maftools' R package. In addition, copy number variations (CNVs) were compared. The correlation of the immune clusters with immune checkpoints (ICPs) and NMD factors was analyzed using the Wilcoxon test. Multiple biomarkers have been identified to predict the prognosis of BLCA. Therefore, the association between different BLCA biomarkers from The Cancer Genome Interpreter (CGI, https://www.cancergenomeinterpreter.org/home) and the immune clusters was assessed (31). The anticancer immune activity of the immune clusters was estimated using the Tracking Tumor Immunophenotyping (TIP, http://biocc.hrbmu.edu.cn/TIP/) (32). Furthermore, the TME-based

ESTIMATE approach was used to compute the immune scores of the immune clusters, and the 'CIBERSORT' R package was used to compare the infiltration of immune cells.

Immune landscape analysis

To further reveal the distribution of immune clusters in each patient, graph learning-based dimensionality reduction analysis was performed using gene expression data. The maximum number of components was set to 4. Moreover, an approach used by Mao et al. (33) was adopted for dimensionality reduction using the Discriminative Dimension Reduction Tree algorithm and the reduceDimension function of the 'Monocle' package. The immune landscape was demonstrated using the function plot cell trajectory, and the plots corresponding to different immune clusters were represented in different colors. In addition, Pearson correlation analysis was used to examine the correlation among 22 TIICs in individual principal components, and differences in the abundance of TIICs between clusters were analyzed using the Wilcoxon test.

Weighted gene co–expression network analysis

Prognostic IRGs were used to perform weighted co-expression network analysis (WGCNA) to obtain gene co-expression modules using the 'WGCNA' package (34). The soft-thresholding power was selected according to the scale-free network topology criterion to construct a correlation adjacency matrix. The resulting modules were used to estimate module eigengenes (MEs) and quantify module similarity. Univariate Cox regression analysis was performed to identify modules that were remarkably associated with patient survival ($P < 0.05$). Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed for genes in each module to annotate gene functions and pathways using the 'clusterProfiler' package (35). Module membership (MM) shows the correlation between genes and modules, and genes with $MM > 0.85$ were defined as hub genes in the prognostic modules.

Quantitative validation of hub genes using quantitative real-time polymerase chain reaction (qRC-PCR)

To validate hub gene expression levels measured by the microarray, the qRT-PCR analyzes were applied using Applied Biosystems 7500 Fast Real-Time PCR System (Thermo Fisher Scientific) with SYBR Premix Ex TaqTM kit (Takara, Dalian, China). Total RNA was isolated from 40 pairs of BLCA and tumor-adjacent normal tissues using TRIzol® reagent (Invitrogen; Thermo Fisher Scientific, Inc.). Reactions were performed at 50 °C for 5 s (1 cycle) 95 °C for 15 min (1 cycle), followed by 95 °C for 15 s and 60 °C for 1 min (40 cycles). Each sample was run in triplicate. Relative mRNA levels were normalized against GAPDH. Data were analyzed using the $2^{-\Delta\Delta Cq}$ method. The primer sequences were listed in Supplementary Table 2.

Results

Identification of potential tumor antigens of BLCA

To identify potential antigens of BLCA, we first screened out aberrant AS events and overexpressed genes that could express TAAs. An integrated profile of AS events was established using the RNA-seq data of patients retrieved from TCGA database. Initially, 39,508 AS events were detected from 18,888 genes, accounting for approximately 92.78% of the potential protein-coding genes (36). The AS events are divided into seven types according to the splicing patterns, including alternate acceptor site (AA), alternate donor site (AD), retained intron (RI), exon skipping (ES), alternate promoter (AP), alternate terminator (AT) and mutually exclusive exons (ME) (Figure 1A). Among these splicing events, ES was the most predominant pattern identified, whereas ME was the least predominant (Figure 1B). Given that a single gene may have multiple AS events, an UpSet plot was generated to visualize the intersecting genes of each AS type. We found that ME always occurred in conjunction with other AS events in most cases, whereas PTK2 had all seven AS events (Figure 1C). To screen for BLCA-specific AS events, we conducted differential expression analysis by comparing 399 BLCA samples with 18 normal samples and identified 2736 CASEs (Figures 1D, E). Among these CASEs, 2352 were upregulated in 1776 genes, whereas 384 were downregulated in 340 genes (Figure 1F). Although ES was the predominant pattern, AP accounted for the highest proportion of CASEs, followed by AT (Figure 1G). The inconsonant distribution patterns among all AS events and CASEs suggested that each AS event played a distinct role in BLCA carcinogenesis.

Furthermore, we analyzed the mutation landscape of BLCA samples from TCGA and found that TP53 had the highest mutation rate (49%) (Supplementary Figure 1A). Tumor genomic mutations contribute to the appearance of neoantigens, and frameshift-mutation-derived peptides have been reported to have the highest immunogenicity (37). A total of 1451 genes with frameshift mutations encoding TSAs or TAAs were screened by evaluating fractional genomic alterations (Supplementary Figure 1B) and mutation counts (Supplementary Figure 1C) in patients, and TTN, TP53, MUC16, KMT2D, ARID1A, KDM6A, SYNE1, PIK3CA, KMT2C and RB1 were identified as the most frequent genetic mutations according to fractional genomic alterations (Supplementary Figure 1D) and mutation counts (Supplementary Figure 1E). This finding was consistent with the overall landscape of mutations. In addition, these 10 genes had the highest mutation count, suggesting underlying genomic interactions. Therefore, based on the combined analysis of the expression and mutation data of patients with BLCA, 153 overexpressed genes with frameshift mutations were identified as potential candidate antigens.

Identification of tumor antigens associated with NMD and antigen-presenting cells

Recently, transcripts harboring frameshift mutation and abnormal AS patterns have been reported to produce antigenic peptides by regulating NMD, which is a determinant of the efficacy of cancer immunotherapy (38). NMD-associated tumor antigens were selected from the identified genes as latent targets for mRNA vaccine development by analyzing the AS events and mutation landscape. We screened for differentially expressed genes in four groups: UPF1, UPF2, UPF3A and UPF3B (Figures 2A–D). The results revealed that most of the top 20 genes were significantly positively correlated with NMD expression in each group. In addition, we analyzed differences

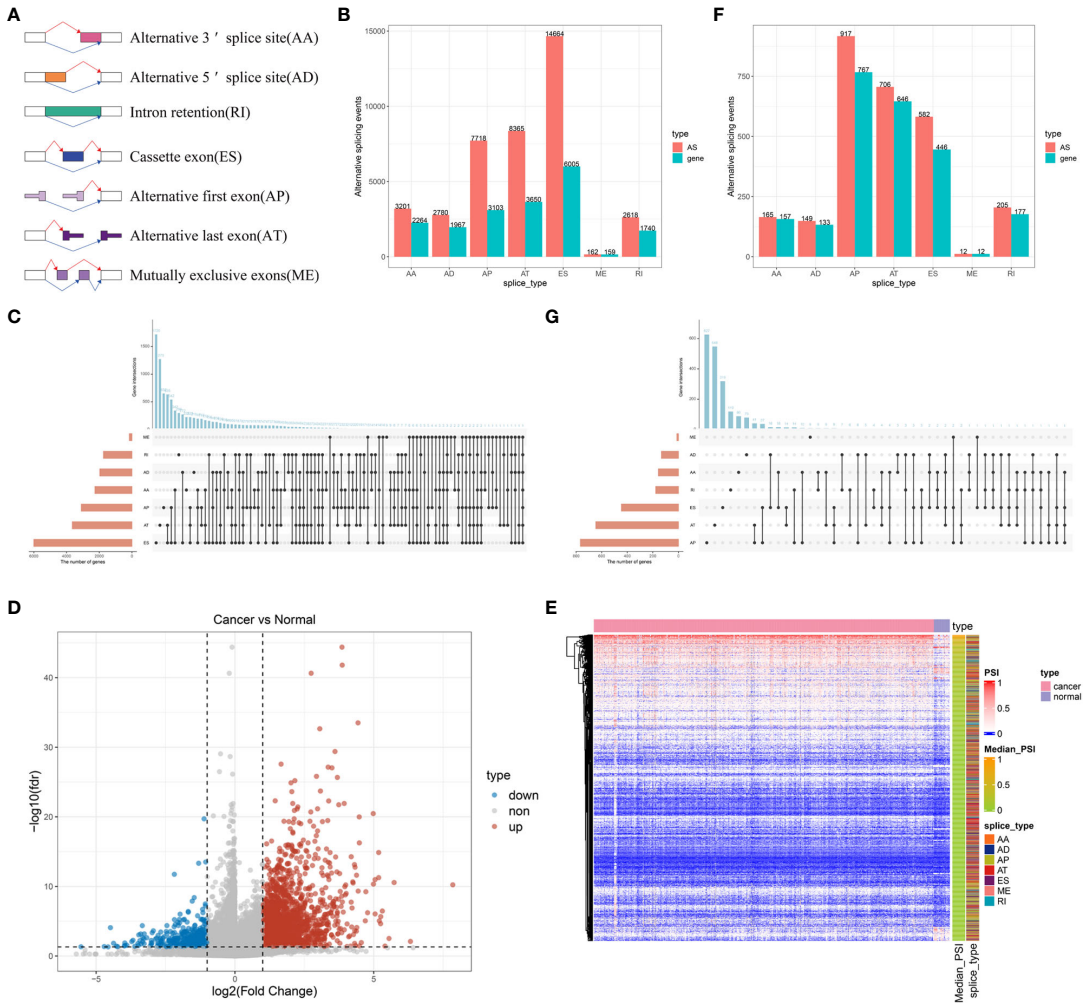


FIGURE 1
Profiling of integrated AS events detected in BLCA. **(A)** Schematic representation of seven different AS events. **(B)** The total number of AS events and the corresponding genes for each AS event in BLCA. **(C)** UpSet plot of interactive genes among seven different types of AS events. **(D)** Heatmap of CASEs between BLCA and normal tissues (|log2FC| ≥ 1, adjusted P < 0.05). **(E)** Volcano plot of CASEs identified in BLCA. **(F)** The total number of CASEs and the corresponding genes for each AS event in BLCA. **(G)** UpSet plot of interactive genes among seven different types of CASEs.

in the PSI value of 885 CASEs from 153 genes in the four NMD groups and found that the PSI values of a majority of CASEs were significantly higher in the high-NMD-expression group than in the low-NMD-expression group. The top 20 CASEs among four groups are shown in Figure 2E–H. Finally, six potential antigens, namely, HP1BP3, OSBPL9, SSH3, ZCCHC8, FANCI and EIF4A2, were identified through the intersection of overexpressed genes, genes with frameshift mutations and NMD-related genes (Supplementary Figure 2). Analysis of immune cell infiltration demonstrated that elevated expression of HP1BP3, OSBPL9, ZCCHC8, FANCI and EIF4A2 was associated with enhanced infiltration of B cells, macrophages and/or DCs (Supplementary Figures 3A–E). In addition, high SSH3 expression was associated with the infiltration of immune cells with some fluctuant (Supplementary Figure 3F). These results suggest that the six neoantigens produced during oncogenesis can be processed and presented by APCs, leading to the initiation of immune responses, and hence are promising targets for developing mRNA vaccines against BLCA with underlying immune activation functions. The peptides of six neoantigens predicted based on TCIA data are listed in Supplementary Table 3.

Identification of potential immune clusters of BLCA

The heterogeneity of TME poses a challenge to cancer immunotherapy, especially in BLCA (39). Therefore, systematic investigation of immunotyping is of great importance to differentiate among patients with BLCA with diverse TIME, which may help in selecting eligible patients for vaccination. In this study, the expression profiles of 1894 IRGs in patients with BLCA were retrieved from TCGA database, and 233 IRGs were identified to be associated with prognosis and used to perform consensus clustering analysis. Based on the consensus accumulative distribution function and delta area (Figures 3A, B), we determined k as 2 for stable clustering of IRGs and obtained two immune clusters designated as

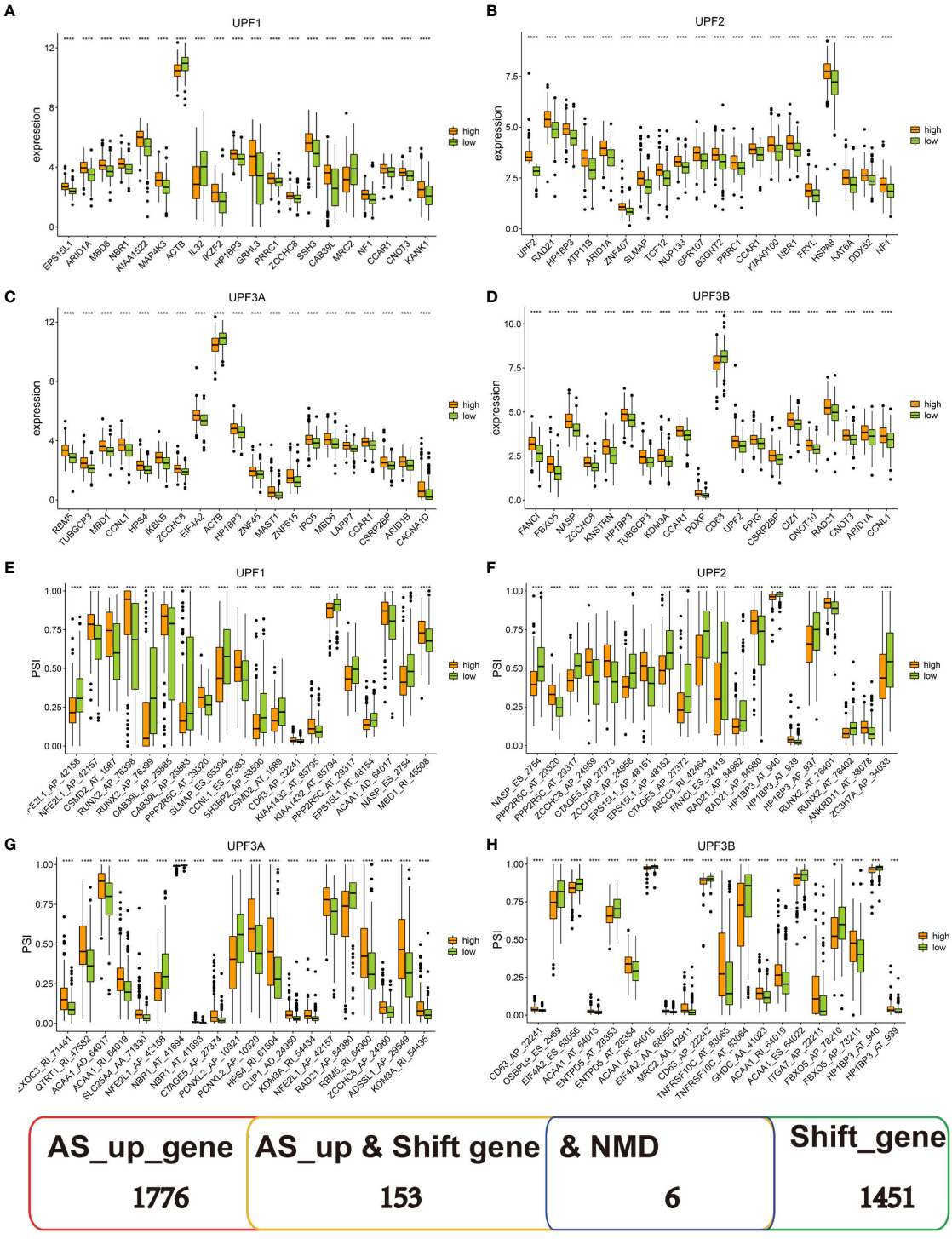


FIGURE 2
Identification of tumor antigens associated with nonsense-mediated mRNA decay factors. **(A–D)** The top 20 differentially expressed genes in four groups, including **(A)** UPF1, **(B)** UPF2, **(C)** UPF3A, and **(D)** UPF3B. **(E–H)** The top 20 CASEs in four groups, including **(E)** UPF1, **(F)** UPF2, **(G)** UPF3A and **(H)** UPF3B. *** P < 0.001, **** P < 0.0001.

IC1 and IC2 (Figure 3C). Principal component analysis revealed that patients in the two clusters were distributed in different directions (Figure 3D). In addition, survival was different between the two clusters; patients in IC1 had a poor prognosis (Figure 3E). Subtype distribution across different clinicopathological features revealed that patients with different stages, grades and clinical T stages were regularly clustered (Figures 3F–H). However, the sex of patients was unsuitable for further differentiation because sex distribution between the two clusters was similar (Figure 3I). The results obtained in TCGA cohort were validated in the GSE32894 cohort using the same approach, and 224 patients with BLCA were divided into two immune clusters (Supplementary Figures 4A–D). We then compared

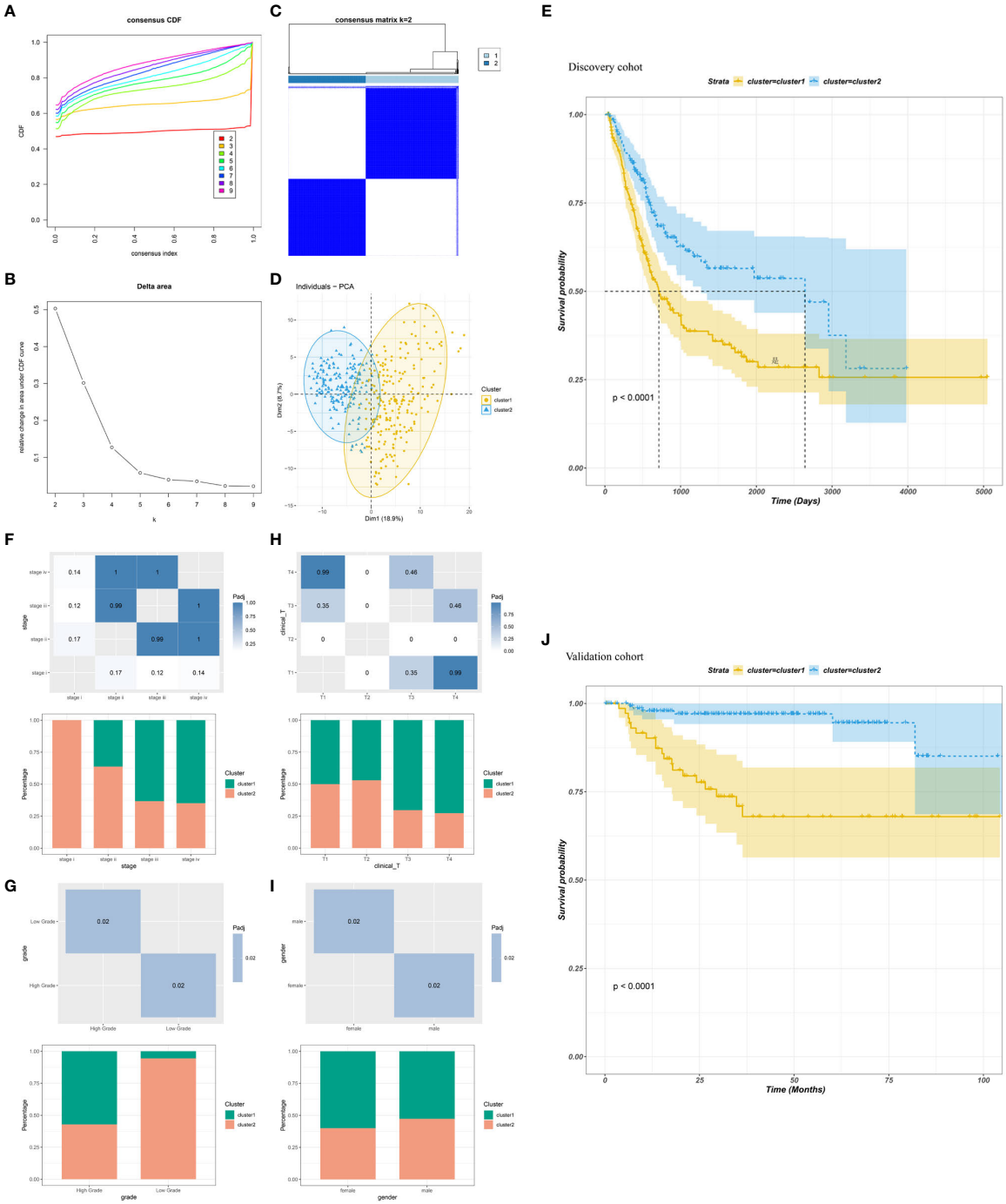


FIGURE 3
Identification of immune clusters of BLCA. **(A)** Cumulative distribution function curve and **(B)** delta area of immune-related genes in TCGA cohort. **(C)** Sample clustering heatmap. **(D)** Principal component analysis demonstrating two distinct clusters reflecting immune status. **(E)** Survival analysis of BLCA immune clusters in TCGA cohort. **(F–I)** Distribution of IC1–2 based on **(F)** stage, **(G)** grade, **(H)** clinical T stage and **(I)** sex in TCGA cohort. **(J)** Survival analysis of BLCA immune clusters in the GSE32894 cohort.

the distribution of different clinicopathological features in two clusters in the GSE32894 cohort (Supplementary Figures 4E–G). These immune clusters also had significant differences in survival, and patients in IC1 had a poorer OS (Figure 3J), suggesting the stability and reproducibility of the established immune clusters. Therefore, these immune clusters can be used as effective prognostic biomarkers for BLCA and are superior to conventional clinical indicators.

Association of immune clusters with mutation status

It has been reported that TMB and somatic mutation rates can be used to evaluate immunotherapeutic efficacy (40). In this study, TMB and mutations were calculated in the two clusters using the mutation data retrieved from TCGA database. No differences were observed in TMB and the number of mutated genes between the two clusters (Figures 4A, B).

10.3389/fimmu.2023.1097472

After analyzing the distribution of the top 20 mutations between two immune clusters, we found that TMB was less extensive in IC1 than in IC2 (93.95 versus 96.09%, respectively). TP53 mutation was significantly more frequent in CI1 than in CI2; however, contradictory results were observed regarding the mutation levels of TTN, MUC16 and KMT2D (Figure 4C, D). It has been reported that copy number alterations (CNAs) are one of the most important hallmarks of the progression of malignancies (41). We found that the frequency of somatic CNVs was significantly lower in patients in IC1 than in patients in IC2 (Figure 4E). In addition, the GISTIC score (G-score) of each patient was evaluated, with an absolute value greater than the threshold of 0.4 based on TCGA data. We found that the G-score varied markedly between the two clusters and was higher in IC2 (Figure 4F). The distribution of CNVs, with either deletions or gains, across all chromosomes was also assessed in the two clusters (Figures 4G, H). These results suggested that the immune clusters could assess the TMB, somatic mutation rates and CNAs of patients with BLCA to a certain extent, which may provide a basis for development of vaccines in the future.

Association of immune clusters with ICPs and NMD factors

ICPs and NMD factors play an important role in anti-tumor immunity, which may affect the response to mRNA vaccines (9, 23). Therefore, we further examined the expression patterns of ICPs and NMD factors in different clusters. A total of 43 ICP-related modulators were detected in TCGA cohort; of which 38 (88.4%) exhibited significant differences between immune clusters (Supplementary Figure 5A). Among these 38 differentially expressed ICP-related genes, only three (TNFRSF14, TNFRSF25 and TNFSF15) were downregulated, whereas almost all other genes were upregulated in IC1. In addition, 36 (90.0%) out of 40 ICP-related genes were differentially expressed in the GSE32894 cohort, and all of them were upregulated in IC1 (Supplementary Figure 4B). Furthermore, four NMD factors were identified in both TCGA and GSE32894 cohorts. Two factors, namely, UPF1 and UPF3A, were diversely expressed in the two clusters in TCGA cohort and were

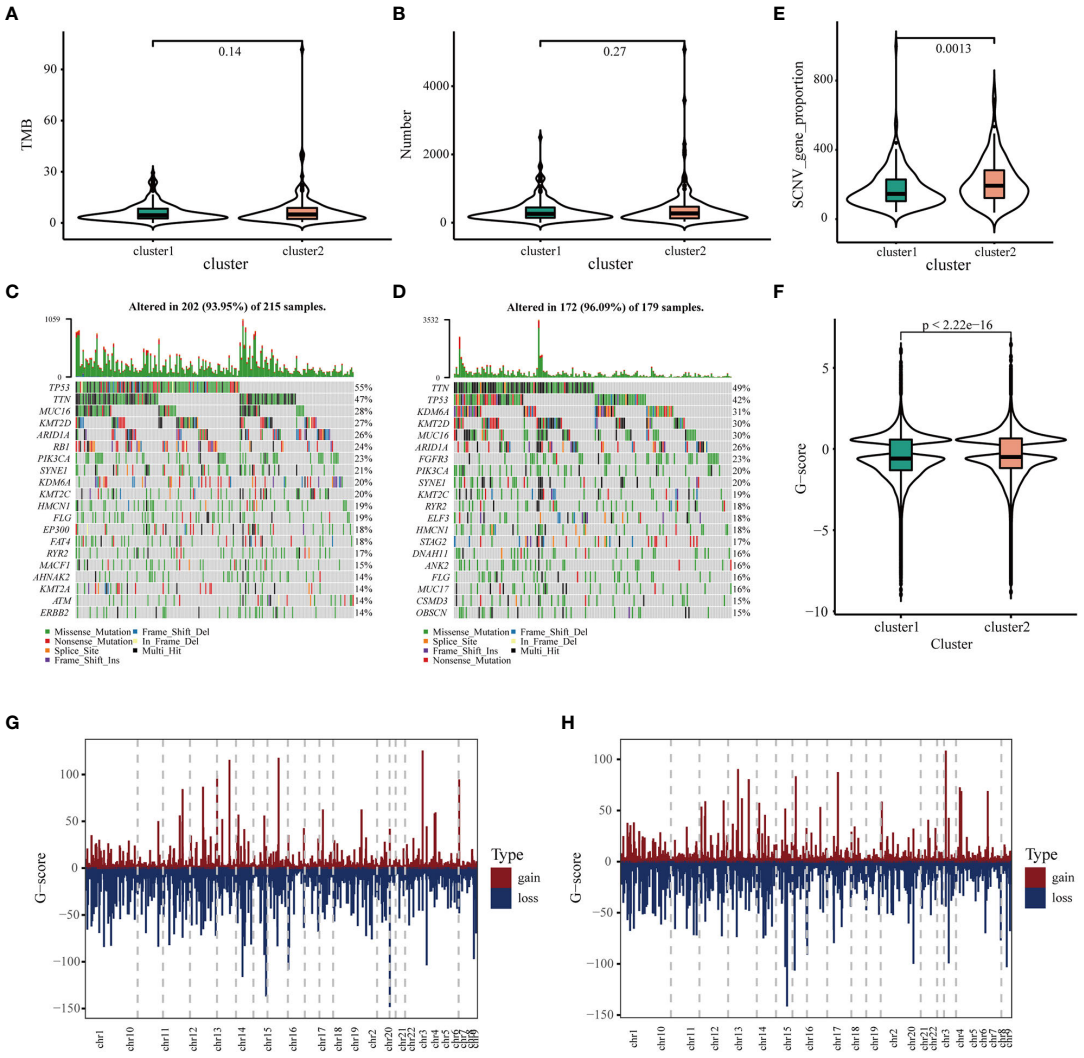


FIGURE 4
Association of immune clusters with mutation. **(A)** Tumor mutation burden and **(B)** mutation count across IC1 and IC2. Waterfall diagram of top 20 mutated genes in **(C)** IC1 and **(D)** IC2. **(E)** Association of immune clusters with somatic CNVs. **(E, F)** Association of immune clusters with G-score. **(G, H)** Gain or loss frequencies of CNVs across chromosomes in **(G)** IC1 and **(H)** IC2.

upregulated in IC2 (Supplementary Figure 5C). Moreover, UPF3A expression was significantly different in the GSE32894 cohort and had the same expression pattern as that of TCGA cohort (Supplementary Figure 5D). Overall, the immune clusters mimicked the expression levels of ICPs and NMD factors, thus serving as potential biomarkers for predicting the efficacy of mRNA vaccines. mRNA vaccines may function better in IC2 owing to the relatively low expression of ICPs and high expression of NMD factors.

Association of immune clusters with tumor markers

We systematically identified 16 prognostic and diagnostic markers of BLCA based on the CGI database. Of these genes, 10 had significantly different expressions between the two immune clusters in both TGCA and GSE32894 cohorts. The expression of CD274, FANCC and TUBB3 was significantly higher in IC1 than in IC2, whereas that of ERBB2, ERBB3, ERCC2, FGFR3, TP53, TSC1 and TSC2 was lower (Supplementary Figures 6A, B). However, these 10 markers have not been approved by the FDA and are either undergoing investigation in early trials or pre-clinical studies; therefore, their clinical applicability remains to be investigated. Currently, nuclear matrix protein 22 (NMP-22) is the most frequently used prognostic marker for BLCA, and patients with high expression have a significantly poorer prognosis (42). Therefore, we analyzed the expression of NMP-22 in patients with BLCA. Serum NMP-22 in IC2 in the TCGA cohort was significantly upregulated (Supplementary Figure 6C), while there was no significant difference in GSE32894 between the two clusters (Supplementary Figure 6D). Overall, the results revealed that the immune clusters were superior to other currently available cancer biomarkers in predicting patient outcomes.

Association of immune clusters with immune microenvironment characteristics

Considering that the efficacy of mRNA vaccines is greatly associated with the immune status tumors, immune activity scores were first assessed using the TIP approach for analyzing and visualizing the status of anti-cancer immunity in the two immune clusters using RNA-seq data of patients with BLCA retrieved from TCGA. The overall score differed significantly between the two clusters, with patients in IC1 having a higher abundance of antitumor immune cells (Figure 5A). To further confirm the feasibility of clustering, we used the ESTIMATE algorithm to assess the immune features of BLCA in both TCGA and GSE32894 cohorts according to the expression of immune cell components. We found that patients in IC1 had higher stromal, immune and ESTIMATE scores but lower tumor purity and cytolytic activity (CYT) (Figure 5B). These results are consistent with those of previous studies, which have reported that low tumor purity (43) and CYT (44) serve as robust indicators for unfavorable prognosis. Furthermore, we examined differences in the abundance of 22 TIICs between the two clusters and found higher enrichment scores in IC1 (Figure 5C). For example, patients in IC1 had higher infiltration of naive B cells, activated memory CD4 T cells, macrophages and neutrophils (Figure 5D). Subsequent analyzes in the GSE32894 cohort yielded similar results (Figures 5E–G). Moreover, patients in IC1 had significantly higher infiltration of a majority of immune cells, including but not limited to memory B cells, plasma cells, CD8 T cells and activated memory CD4 T cells. Therefore, IC1 was considered an immunologically 'hot' phenotype, whereas IC2 was considered an immunologically 'cold' phenotype. Based on the abovementioned analyzes, we speculated that immune clusters can evaluate the immune status of BLCA and may help in selecting eligible patients for mRNA vaccination. These vaccines may be involved in the activation of various TIICs in immunologically 'cold' IC2. The six pan-cancer immune categories (C1–C6) defined by Thorsson et al. were closely related to prognosis and immunoregulation in tumors (45). As shown in Figure 5H, a distinct distribution of C1–C6 was observed in IC1 and IC2. In addition, there was a large degree of overlap in the proportion of C1–2 between the two immune clusters. The proportion of C1, C3 and C4 increased significantly, whereas that of IC2 decreased significantly in IC2 as compared to IC1. Patients in IC2 with a longer survival duration may be associated with the high proportion of C3 samples in IC2. These results facilitated a deeper understanding of the characteristics of the immune microenvironment in BLCA while further complementing previous studies.

Immune landscape of BLCA

The immune gene expression profiles were integrated to assess the immune landscape of BLCA (Figure 6A). We found that the overall pattern of IC1 and IC2 distribution was reversed in the immune landscape. Principal component 1 (horizontal axis) had a positive correlation with activated memory CD4 T cells and M1 macrophages but a negative correlation with activated dendritic cells and naive CD4 T cells. In addition, principal component 2 (vertical axis) was most positively correlated with plasma cells and Tregs but most negatively correlated with activated memory CD4 T cells, resting NK cells and M1 macrophages (Figure 6B). The correlations among different immune cells between two principal components further demonstrated the accuracy of our approach. Heterogeneity can be found within the same cluster, and IC1 presented opposing distribution. Therefore, we stratified IC1 into two subclusters (IC1A and IC1B) based on the distribution of immune cell populations (Figure 6C). Considerable differences in the proportion of certain immune cells were observed. IC1B had a higher enrichment score of M0 macrophages and activated dendritic cells and a lower enrichment score of CD8 T cells (Figure 6D), suggesting that mRNA vaccines may be more effective in IC1B. In addition, we conducted the survival analysis of extremely distributed samples in the immune landscape and found that the survival rate of state 5 was significantly higher than that of state 1, suggesting that immune cluster-based immune landscape can be used to assess patient outcomes (Figures 6E, F). Collectively, the immune landscape based on immune clusters precisely identified immune components in each patient and predicted their outcomes, thus facilitating individualized mRNA vaccination.

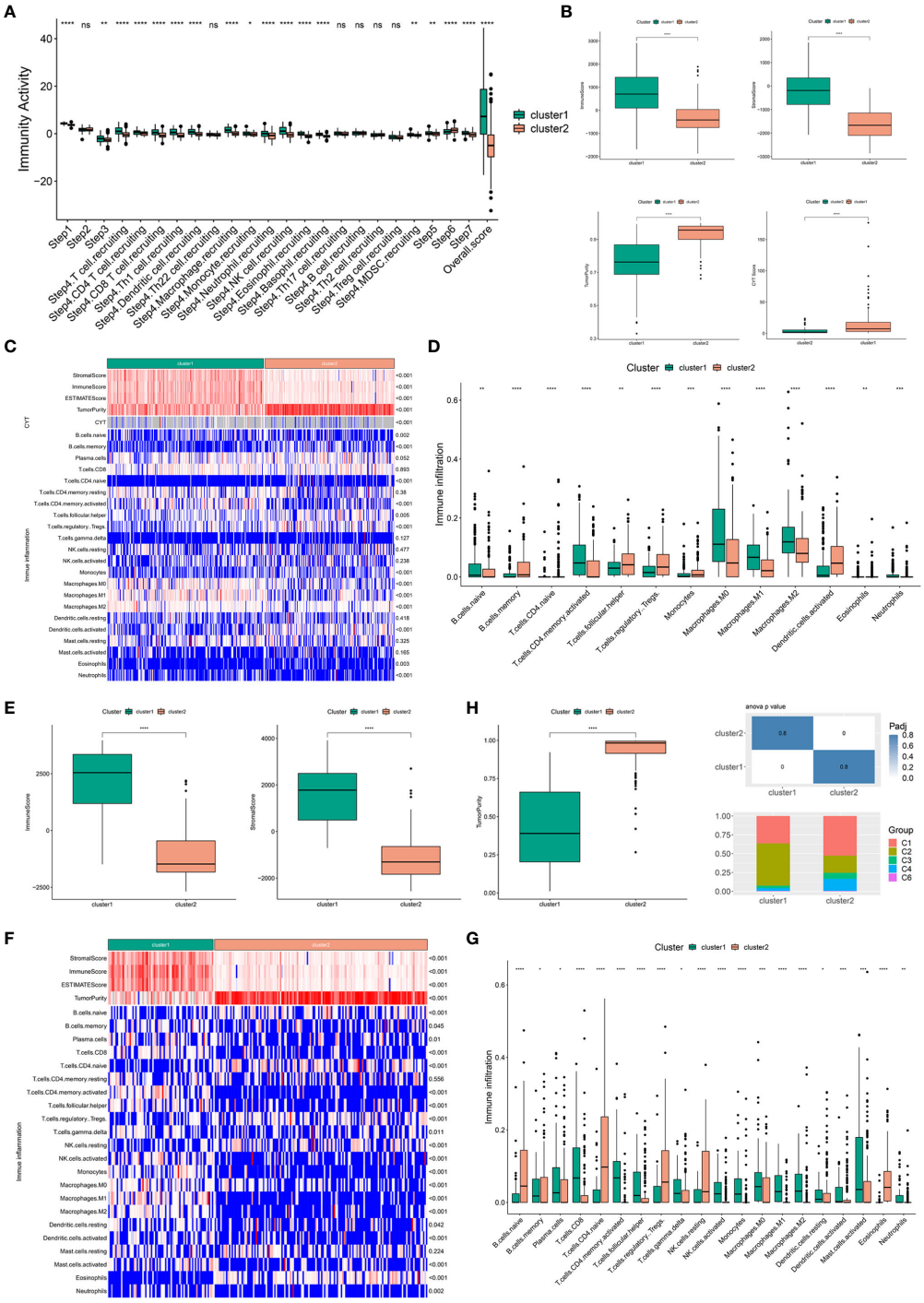


FIGURE 5
Association of immune clusters with immune microenvironment characteristics. **(A)** Distribution of immune activity scores in IC1 and IC2. **(B)** Association of immune subtypes with immune scores, stromal scores, tumor purity and CYT in TCGA cohort. Heatmap **(C)** and bar plot **(D)** of the relationship between immune clusters and immune cell subpopulations in TCGA cohort. **(E)** Association of immune subtypes with immune scores, stromal scores and tumor purity in the GSE32894 cohort. Heatmap **(F)** and bar plot **(G)** of the relationship between immune clusters and immune cell subpopulations in the GSE32894 cohort. **(H)** Distribution of individual immune categories in the two immune clusters. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; ns, not significant.

Identification of immune gene co-expression modules and hub genes of BLCA

WGCNA was performed to identify immune gene co-expression modules containing immune genes associated with the effectiveness of mRNA vaccines. No outliers were found in the sample clustering (Figure 7A), and a soft-threshold power of β = 3 (scale-free R2 = 0.85) was selected to ensure a scale-free network (Figures 7B, C). Subsequently, the representation matrix was converted to an adjacency matrix and then to a topological matrix. Considering the minimum module size of 30 genes as a criterion, a dendrogram was constructed using the average-linkage hierarchy clustering method.

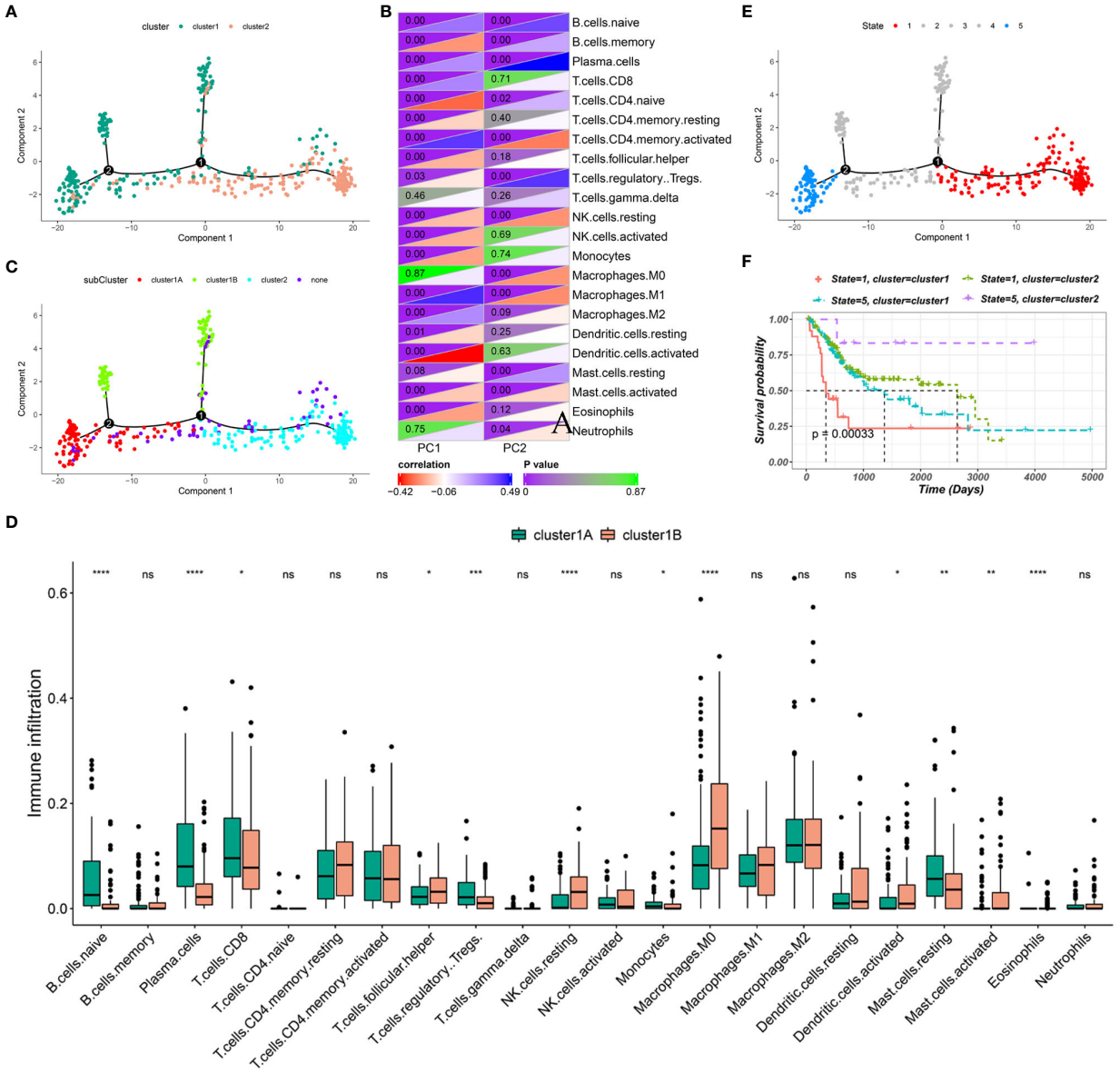


FIGURE 6
Immune landscape of BLCA. **(A)** Immune landscape of BLCA. Each point represents a patient, and the immune clusters are color-coded. **(B)** Correlation between two principal components and immune cells. **(C)** Immune landscape of the subclusters of BLCA immune clusters. **(D)** The proportion of certain immune cells in the IC1A–B subclusters. **(E)** Immune landscape of samples from five extreme locations and **(F)** their prognoses. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001; ns, not significant.

MEs were calculated by merging the closed modules with a deep split of 5 and a height of 0.2 (Figure 7D). Eventually, four modules that contained similar gene patterns were identified, and the grey module included genes that were not present in any module (Figure 7E). We further examined MEs in the two immune clusters and noticed significantly different distribution of all three modules (except the grey module). IC1 had a higher number of MEs in the brown and turquoise modules, whereas IC2 showed higher eigengenes in the blue module (Figure 7F). Further prognostic correlation results suggested that the blue, turquoise and brown modules were distinctly associated with the prognosis of BLCA (Figure 7G). Moreover, functional enrichment analysis indicated that genes in the blue module were relevant to BC, those in the turquoise module were associated with pathways in cancer and those in the brown module were associated with natural killer cell-mediated cytotoxicity (Figures 7H–J). Eventually, three hub genes with relevance > 85% to MEs of three modules were identified, including CALD1, TGFB3 and ANXA6. These hub genes can be used as predictive and prognostic biomarkers and for identifying eligible patients with BLCA for mRNA vaccination. qRT-PCR was applied to examine the relative mRNA levels of CALD1, TGFB3 and ANXA6 in 40 pairs of BLCA and adjacent normal tissues. Results demonstrated that the expression levels of these 3 genes were higher in BLCA tissues than in tumor-adjacent normal tissues (Supplementary Figure 7).

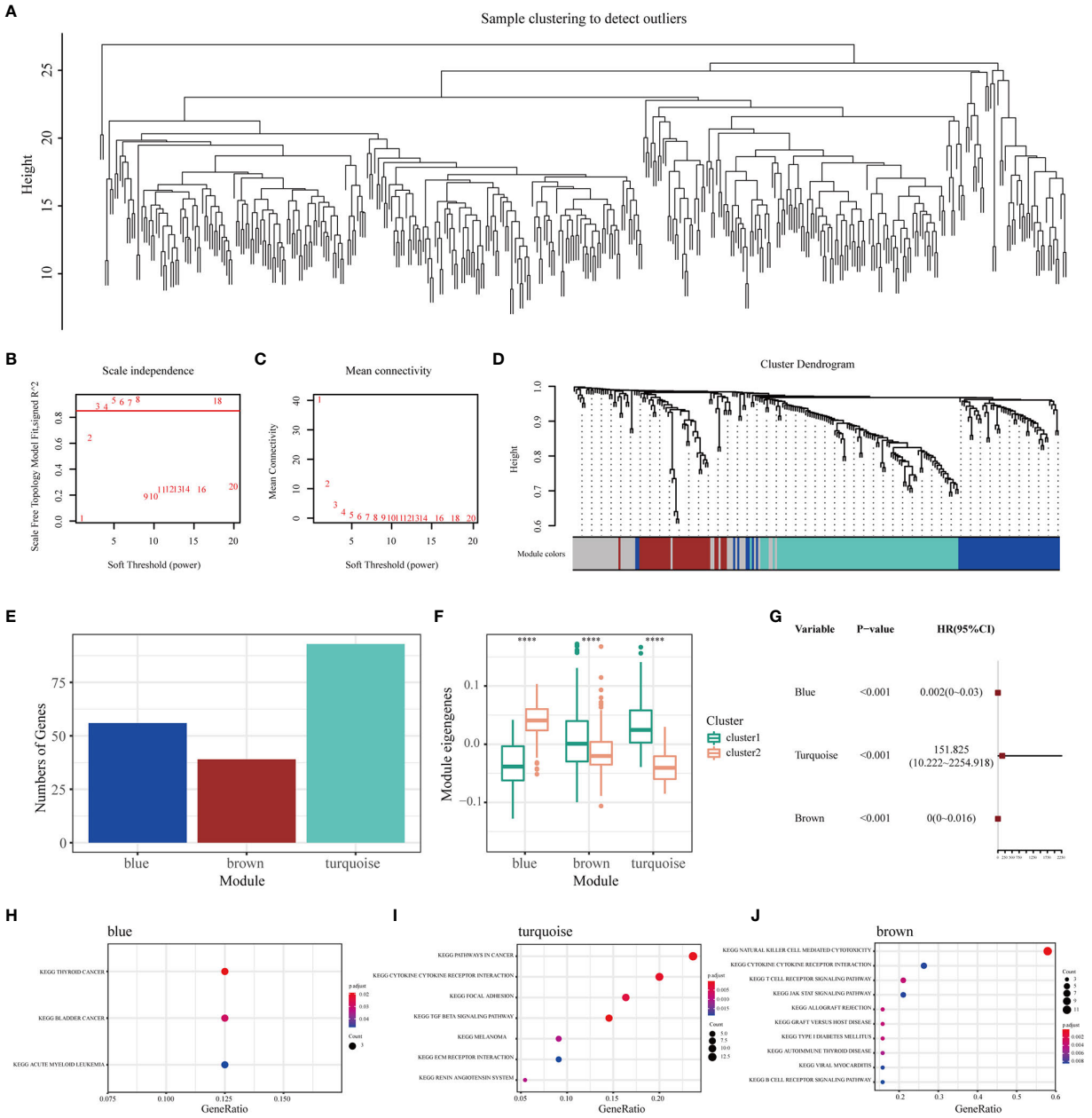


FIGURE 7
Identification of immune gene co-expression modules and hub genes of BLCA. **(A)** Sample clustering. **(B)** Scale-free fit index and **(C)** mean connectivity for various soft-thresholding powers. **(D)** Dendrogram of immune genes clustered based on the average-linkage hierarchy clustering method. **(E)** The number of genes in each module. **(F)** Differential distribution of MEs of each module in BLCA immune clusters. **(G)** Prognostic analysis of the blue, turquoise and brown modules. The top KEGG terms enriched in the **(H)** blue, **(I)** turquoise and **(J)** brown modules. ****P < 0.0001.

Discussion

BLCA is one of the most aggressive malignancies with significant tumor heterogeneity, multidrug resistance and uncontrolled metastasis (3, 5). Cisplatin combined with gemcitabine has been established as a standard therapeutic strategy for the management of patients with advanced BLCA; however, clinical benefits are limited (6, 7). Immunotherapy has revolutionized treatment paradigms in oncology, especially with the clinical success of immune checkpoint inhibitors; however, its effectiveness in BLCA remains unknown (10). mRNA-based vaccines with cancer antigens represent a promising alternative immunotherapeutic strategy, with multiple ongoing human clinical trials (15). A recent study demonstrated that a combination of mRNA vaccine and immune checkpoint inhibitor can enhance the immune response against melanoma and inhibit tumor progression (20). However, the efficacy of mRNA vaccines in patients with BLCA remains unknown.

In this study, we systematically profiled the aberrant AS events and mutational landscape of BLCA for future development of individualized mRNA-based cancer vaccines. To elucidate the clinical relevance of the selected antigens, their correlation with NMD factors and immune cell infiltration was examined. Six tumor antigens (HP1BP3, OSBPL9, SSH3, ZCCHC8, FANCI and EIF4A2) were correlated with the expression of NMD factors and infiltration of APCs, which may be promising candidates for mRNA vaccines. These findings revealed the

importance of these candidates in the development of BLCA, which can be recognized and presented directly to T cell receptors to eradicate tumor cells and induce antitumor immunity. Although functional validation and clinical evaluation of these candidate genes require further investigation, previous studies have demonstrated their potential for developing mRNA vaccines against tumors. In recent years, several studies have focused on the role of EIF4A2 in regulating immune responses and numerous cellular and pathophysiological processes, which serves as a prognostic biomarker and is correlated with immune infiltration in multiple cancers, including BLCA (46, 47).

Because therapeutic effects of mRNA vaccines vary between individuals, patients with BLCA were divided into two immune clusters (IC1 and IC2) based on their immune-related gene profiles to identify eligible patients for vaccination. The two immune clusters possessed distinct clinical, molecular and TIME features. For example, patients in IC2 had a better prognosis in both TCGA and GSE32894 cohorts, suggesting that the immune cluster could serve as a prognostic biomarker for BLCA. In addition, we observed that the predictive power of IC2 was superior to conventional tumor biomarkers such as NMP-22. In addition, this cluster can be used to predict the response to vaccine therapy. Elevated rates of somatic CNVs in IC2 are suggestive of greater responsiveness to mRNA vaccines. With regard to the expression of ICPs and NMD factors, mRNA vaccines might have better efficacy in IC2. The immune activity of the two immune subtypes was assessed *via* TIP analysis. IC2 had strikingly lower immune activity, suggesting that mRNA vaccines targeting IC1 might reinforce its immune response.

Given that the tumor immune status is critical for the efficacy of mRNA vaccines, we further investigated the immune cell components and found a strikingly distinct TIME in the two immune clusters. This finding suggested that the two clusters might have varying mechanisms for regulating immune escape in tumors, which may require individualized therapeutic strategies. In addition, we found that IC2 had an immunologically cold phenotype with less infiltration of immune cells (an 'immune desert') and immunologically inactive 'non-inflamed' tumors. This phenotype may be related to the lack of APCs and tumor antigens, resulting in T cell anergy and insensitivity to antigen activation. To reinvigorate the immune system of such patients against tumor cells, mRNA-based vaccines that trigger immune cell infiltration may be an appropriate option. However, IC1 had a favorable immunologically hot phenotype, characterized by the increased infiltration of immune cells and immunologically active 'inflamed' tumors. Therefore, ICBs are especially advantageous for patients in IC1, which may further regulate the production of CD8+ T lymphocytes and the suppression of Tregs, inducing antitumor immunity (48, 49). Recent studies have highlighted the role of inflammation in tumorigenesis and tumor progression, revealing a close relationship between inflammation and BLCA (50, 51). An inflammatory phenotype with a high density of macrophages might, at least partially, lead to poor outcomes in patients in IC1 (51). Another important consideration in determining the prognosis is the preponderance of the immune-suppressive or -stimulatory environment. In a study, patients with BLCA were divided into C1-C6 subtypes, except for the C5 subtype, based on previous immunotyping studies among 33 cancer types (45). The C3 subtype was associated with the best prognosis, followed by the C1, C2, C4 and C6 subtypes. Our data showed substantial variations in the proportion of five categories in IC1 and IC2. Patients in IC2 with a longer survival duration may be associated with the high proportion of C3 samples in IC2. Moreover, the proportion of C2 (IFN-g dominant) in IC1 was significantly higher than that in IC2, whereas the proportion of C4 (lymphocyte failure) showed the opposite trend. This finding provides further verified the 'hot' phenotype of IC1 and the 'cold' phenotype of IC2. Therefore, mRNA vaccine administration in IC2 might stimulate the immune response, thus converting the 'cold' TME to 'hot' by increasing the infiltration of inflammatory immune cells (49). Therefore, our results are reliable and complement the classification schemes previously developed.

Furthermore, the complex immune landscape of BLCA demonstrated substantial heterogeneity among individuals and within the same immune subgroups, thus facilitating the accurate determination of immune cell components in each patient to aid in developing individualized mRNA vaccines. Intra-cluster heterogeneity observed in IC1 was based on the distribution of immune cell groups. The infiltration of M0 macrophages and activated dendritic cells was higher and that of CD8 T cells was lower in IC1B than in IC1A, suggesting that the therapeutic efficacy of mRNA vaccines may be better in patients in IC1B. In such patients, novel treatment strategies based on mRNA vaccines combined with chemotherapy or immunotherapy may modulate both TME and immune response of the host, which is considered more conducive to successful therapy (52).

Furthermore, we used WGCNA to construct co-expressed gene modules and identified three key modules (blue, turquoise and brown) significantly correlated with each immune cluster, which was of fundamental importance in investigating the underlying biological mechanisms of the clusters. Subsequent KEGG analysis suggested that the three modules had substantial disparity among the involved pathways, suggesting that the classification method had high discrimination power. In addition, CALD1, TGFB3 and ANXA6 were identified as immune hub genes (MM > 0.85), which may serve as biomarkers for predicting the outcomes of patients with BLCA and selecting eligible patients for mRNA vaccination.

Therefore, this study provides critical insights into developing mRNA vaccines for other diseases. The emergence of the COVID-19 pandemic made mRNA vaccines an innovative and promising platform (53). Although mRNA vaccines have protected millions of patients with COVID-19 and prevented many deaths worldwide, the evolving variants such as D614G require these vaccines to be updated periodically (54). Therefore, there are significant implications for improving the clinical treatment of COVID-19 by determining specific antigens and eligible patients for mRNA vaccine administration.

Conclusion

In conclusion, HP1BP3, OSBPL9, SSH3, ZCCHC8, FANCI and EIF4A2 were identified as potential antigens for developing mRNA vaccines against BLCA. In addition, patients in IC2 may benefit more from mRNA vaccination. These findings provide new sights into developing mRNA vaccines against BLCA and defining the eligible population for mRNA vaccination.

Data availability statement

Publicly available datasets were analyzed in this study, the names of the repositories/accession numbers are included within the article/Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by Shandong Provincial Hospital Affiliated to Shandong First Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

ZS, CJ, HZ and XG conceived the study, conducted bioinformatics analysis, and wrote the manuscript. NS and FK participated in drafting and revising the article. CX, WT, and DH refined the data analysis and prepared all the figures. HW and SJ revised and supervised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The research was supported by the Key Research and Development Program of Shandong, China (2019GSF108263), Natural Science Foundation of Shandong Province, China (ZR2020QH069). Natural Science Foundation of Shandong Province, China (ZR2022MH004).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1097472/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA: Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654

2. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* (2017) 171(3):540–56.e25. doi: 10.1016/j.cell.2017.09.007

3. Knowles MA, Hurst CD. Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat Rev Cancer* (2015) 15(1):25–41. doi: 10.1038/nrc3817

4. Cumberbatch MGK, Jubber I, Black PC, Esperto F, Figueroa JD, Kamat AM, et al. Epidemiology of bladder cancer: A systematic review and contemporary update of risk factors in 2018. *Eur Urol* (2018) 74(6):784–95. doi: 10.1016/j.eururo.2018.09.001

5. Dy GW, Gore JL, Forouzanfar MH, Naghavi M, Fitzmaurice C. Global burden of urologic cancers, 1990-2013. *Eur Urol* (2017) 71(3):437–46. doi: 10.1016/j.eururo.2016.10.008

6. Nadal R, Bellmunt J. Management of metastatic bladder cancer. *Cancer Treat Rev* (2019) 76:10–21. doi: 10.1016/j.ctrv.2019.04.002

7. Pietzak EJ, Zabor EC, Bagrodia A, Armenia J, Hu W, Zehir A, et al. Genomic differences between "Primary" and "Secondary" muscle-invasive bladder cancer as a basis for disparate outcomes to cisplatin-based neoadjuvant chemotherapy. *Eur Urol* (2019) 75(2):231–9. doi: 10.1016/j.eururo.2018.09.002

8. Huang HM, Li HX. Tumor heterogeneity and the potential role of liquid biopsy in bladder cancer. *Cancer Commun (London England)* (2021) 41(2):91–108. doi: 10.1002/cac2.12129

9. Bellmunt J, Powles T, Vogelzang NJ. A review on the evolution of pd-1/Pd-L1 immunotherapy for bladder cancer: The future is now. *Cancer Treat Rev* (2017) 54:58–67. doi: 10.1016/j.ctrv.2017.01.007

10. Chism DD. Urothelial carcinoma of the bladder and the rise of immunotherapy. *J Natl Compr Cancer Network JNCCN* (2017) 15(10):1277–84. doi: 10.6004/jnccn.2017.7036

11. Wu S, Zhong G, Zhang J, Shuai L, Zhang Z, Wen Z, et al. A single dose of an adenovirus-vectored vaccine provides protection against sars-Cov-2 challenge. *Nat Commun* (2020) 11(1):4081–. doi: 10.1038/s41467-020-17972-1

12. Pushparajah D, Jimenez S, Wong S, Alattas H, Nafissi N, Slavcev RA. Advances in gene-based vaccine platforms to address the covid-19 pandemic. *Advanced Drug Del Rev* (2021) 170:113–41. doi: 10.1016/j.addr.2021.01.003

13. Xia L, Schrump DS, Gildersleeve JC. Whole-cell cancer vaccines induce Large antibody responses to carbohydrates and glycoproteins. *Cell Chem Biol* (2016) 23(12):1515–25. doi: 10.1016/j.chembiol.2016.10.012

14. Huang X, Zhang G, Tang T, Liang T. Identification of tumor antigens and immune subtypes of pancreatic adenocarcinoma for mrna vaccine development. *Mol Cancer* (2021) 20(1):44. doi: 10.1186/s12943-021-01310-0

15. Pardi N, Hogan MJ, Porter FW, Weissman D. Mrna vaccines - a new era in vaccinology. *Nat Rev Drug Discovery* (2018) 17(4):261–79. doi: 10.1038/nrd.2017.243

16. Miao L, Zhang Y, Huang L. Mrna vaccine for cancer immunotherapy. *Mol Cancer* (2021) 20(1):41–. doi: 10.1186/s12943-021-01335-5

17. Van Hoecke L, Verbeke R, Dewitte H, Lentacker I, Vermaelen K, Breckpot K, et al. Mrna in cancer immunotherapy: Beyond a source of antigen. *Mol Cancer* (2021) 20(1):48–. doi: 10.1186/s12943-021-01329-3

18. Sebastian M, Schröder A, Scheel B, Hong HS, Muth A, von Boehmer L, et al. A phase I/Iia study of the mrna-based cancer immunotherapy Cv9201 in patients with stage Iiib/Iv non-small cell lung cancer. *Cancer Immunol Immunother* (2019) 68(5):799–812. doi: 10.1007/s00262-019-02315-x

19. Kübler H, Scheel B, Gnad-Vogt U, Miller K, Schultze-Seemann W, Vom Dorp F, et al. Self-adjuvanted mrna vaccination in advanced prostate cancer patients: A first-in-Man phase I/Iia study. *J immunotherapy Cancer* (2015) 3:26. doi: 10.1186/s40425-015-0068-y

20. Wang Y, Zhang L, Xu Z, Miao L, Huang L. Mrna vaccine with antigen-specific checkpoint blockade induces an enhanced immune response against established melanoma. *Mol Ther J Am Soc Gene Ther* (2018) 26(2):420–34. doi: 10.1016/j.ymthe.2017.11.009

21. Suzuki A, Makinoshima H, Wakaguri H, Esumi H, Sugano S, Kohno T, et al. Aberrant transcriptional regulations in cancers: Genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. *Nucleic Acids Res* (2014) 42(22):13557–72. doi: 10.1093/nar/gku885

22. Lindeboom RG, Supek F, Lehner B. The rules and impact of nonsense-mediated mrna decay in human cancers. *Nat Genet* (2016) 48(10):1112–8. doi: 10.1038/ng.3664

23. Litchfield K, Reading JL, Lim EL, Xu H, Liu P, Al-Bakir M, et al. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat Commun* (2020) 11(1):3800. doi: 10.1038/s41467-020-17526-5

24. Oka M, Xu L, Suzuki T, Yoshikawa T, Sakamoto H, Uemura H, et al. Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer. *Genome Biol* (2021) 22(1):9–. doi: 10.1186/s13059-020-02240-8

25. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative splicing signatures in rna-seq data: Percent spliced in (Psi). *Curr Protoc Hum Genet* (2015) 87:11. doi: 10.1002/0471142905.hg1116s87

26. Conway JR, Lex A, Gehlenborg N. Upsetr: An r package for the visualization of intersecting sets and their properties. *Bioinf (Oxford England)* (2017) 33(18):2938–40. doi: 10.1093/bioinformatics/btx364

27. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cbio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* (2012) 2(5):401–4. doi: 10.1158/2159-8290.cd-12-0095

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

29. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, et al. Immport: Disseminating data to the public for the future of immunology. *Immunologic Res* (2014) 58(2-3):234–9. doi: 10.1007/s12026-014-8516-1

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

31. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer genome interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* (2018) 10(1):25. doi: 10.1186/s13073-018-0531-8

32. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, et al. Tip: A web server for resolving tumor immunophenotype profiling. *Cancer Res* (2018) 78(23):6575–80. doi: 10.1158/0008-5472.can-18-0689

33. Qi M, Li W, Tsang IW, Yijun S. Principal graph and structure learning based on reversed graph embedding. *IEEE Trans Pattern Anal Mach Intell* (2017) 39(11):2227–41. doi: 10.1109/tpami.2016.2635657

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

35. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: An r package for comparing biological themes among gene clusters. *Omics J Integr Biol* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

36. Pertea M, Shumate A, Pertea G, Varabyou A, Breitwieser FP, Chang Y-C, et al. Chess: A new human gene catalog curated from thousands of Large-scale rna sequencing experiments reveals extensive transcriptional noise. *Genome Biol* (2018) 19(1):208–. doi: 10.1186/s13059-018-1590-2

37. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep* (2016) 15(4):857–65. doi: 10.1016/j.celrep.2016.03.075

38. Lindeboom RGH, Vermeulen M, Lehner B, Supek F. The impact of nonsense-mediated mrna decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet* (2019) 51(11):1645–51. doi: 10.1038/s41588-019-0517-5

39. Meeks JJ, Al-Ahmadie H, Faltas BM, Taylor JA3rd, Flaig TW, DeGraff DJ, et al. Genomic heterogeneity in bladder cancer: Challenges and possible solutions to improve outcomes. *Nat Rev Urol* (2020) 17(5):259–70. doi: 10.1038/s41585-020-0304-1

40. Sha D, Jin Z, Budczies J, Kluck K, Stenzinger A, Sinicrope FA. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discovery* (2020) 10(12):1808–25. doi: 10.1158/2159-8290.cd-20-0522

41. Chaluvally-Raghavan P, Zhang F, Pradeep S, Hamilton MP, Zhao X, Rupaimoole R, et al. Copy number gain of hsa-Mir-569 at 3q26.2 leads to loss of Tp53inp1 and aggressiveness of epithelial cancers. *Cancer Cell* (2014) 26(6):863–79. doi: 10.1016/j.ccell.2014.10.010

42. Chou R, Gore JL, Buckley D, Fu R, Gustafson K, Griffin JC, et al. Urinary biomarkers for diagnosis of bladder cancer: A systematic review and meta-analysis. *Ann Internal Med* (2015) 163(12):922–31. doi: 10.7326/m15-0997

43. Zhang C, Cheng W, Ren X, Wang Z, Liu X, Li G, et al. Tumor purity as an underlying key factor in glioma. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2017) 23 (20):6279–91. doi: 10.1158/1078-0432.ccr-16-2598

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

45. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023

46. De Benedetti A, Graff JR. Eif-4e expression and its role in malignancies and metastases. *Oncogene* (2004) 23(18):3189–99. doi: 10.1038/sj.onc.1207545

47. Crew JP, Fuggle S, Bicknell R, Cranston DW, de Benedetti A, Harris AL. Eukaryotic initiation factor-4e in superficial and muscle invasive bladder cancer and its correlation with vascular endothelial growth factor expression and tumour progression. *Br J Cancer* (2000) 82(1):161–6. doi: 10.1054/bjoc.1999.0894

48. Sharpe AH. Introduction to checkpoint inhibitors and cancer immunotherapy. *Immunol Rev* (2017) 276(1):5–8. doi: 10.1111/imr.12531

49. Newman JH, Chesson CB, Herzog NL, Bommareddy PK, Aspromonte SM, Pepe R, et al. Intratumoral injection of the seasonal flu shot converts immunologically cold tumors to hot and serves as an immunotherapy for cancer. *Proc Natl Acad Sci United States America* (2020) 117(2):1119–28. doi: 10.1073/pnas.1904022116

50. Wigner P, Grębowski R, Bijak M, Saluk-Bijak J, Szemraj J. The interplay between oxidative stress, inflammation and angiogenesis in bladder cancer development. *Int J Mol Sci* (2021) 22(9):4483. doi: 10.3390/ijms22094483

51. Crispen PL, Kusmartsev S. Mechanisms of immune evasion in bladder cancer. *Cancer Immunol Immunother* (2020) 69(1):3–14. doi: 10.1007/s00262-019-02443-4

52. Sahin U, Oehm P, Derhovanessian E, Jabulowsky RA, Vormehr M, Gold M, et al. An rna vaccine drives immunity in checkpoint-Inhibitor-Treated melanoma. *Nature* (2020) 585(7823):107–12. doi: 10.1038/s41586-020-2537-9

53. Lazarus R, Baos S, Cappel-Porter H, Carson-Stevens A, Clout M, Culliford L, et al. Safety and immunogenicity of concomitant administration of covid-19 vaccines (Chadox1 or Bnt162b2) with seasonal influenza vaccines in adults in the uk (Comflucov): A multicentre, randomised, controlled, phase 4 trial. *Lancet (London England)* (2021) 398(10318):2277–87. doi: 10.1016/S0140-6736(21)02329-1

54. Weissman D, Alameh MG, de Silva T, Collini P, Hornsby H, Brown R, et al. D614g spike mutation increases sars cov-2 susceptibility to neutralization. *Cell Host Microbe* (2021) 29(1):23–31.e4. doi: 10.1016/j.chom.2020.11.012

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 15 February 2023
DOI 10.3389/fimmu.2023.1112181

OPEN ACCESS

EDITED BY
Meng Zhou,
Wenzhou Medical University, China

REVIEWED BY
Mou Peng,
Department of Urology, Central South University, China
Denise Cecil,
University of Washington, United States

*CORRESPONDENCE
Kai Huang
kaiserhuang@csu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 30 November 2022
ACCEPTED 26 January 2023
PUBLISHED 15 February 2023

CITATION
Li M, Long X, Bu W, Zhang G, Deng G, Liu Y, Su J and Huang K (2023) Immune-related risk score: An immune-cell-pair-based prognostic model for cutaneous melanoma.
Front. Immunol. 14:1112181.
doi: 10.3389/fimmu.2023.1112181

Immune-related risk score: An immune-cell-pair-based prognostic model for cutaneous melanoma

Mingjia Li[1,2,3,4,5,6†], Xinrui Long[1,2,3,4,5†], Wenbo Bu[7†], Guanxiong Zhang[1,2,3,4,5], Guangtong Deng[1,2,3,4,5], Yuancheng Liu[1,2,3,4,5], Juan Su[1,2,3,4,5] and Kai Huang[1,2,3,4,5*]

[1]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China, [2]National Engineering Research Center of Personalized Diagnostic and Therapeutic Technology, Changsha, China, [3]Hunan Engineering Research Center of Skin Health and Disease, Central South University, Changsha, China, [4]Hunan Key Laboratory of Skin Cancer and Psoriasis, Xiangya Hospital, Central South University, Changsha, China, [5]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [6]Department of Dermatology, Peking University First Hospital, Peking University, Beijing, China, [7]Department of Dermatological Surgery, Hospital for Skin Diseases, Institute of Dermatology, Chinese Academy of Medical Sciences, Peking Union Medical College, Nanjing, China

Background: Melanoma is among the most malignant immunologic tumor types and is associated with high mortality. However, a considerable number of melanoma patients cannot benefit from immunotherapy owing to individual differences. This study attempts to build a novel prediction model of melanoma that fully considers individual differences in the tumor microenvironment.

Methods: An immune-related risk score (IRRS) was constructed based on cutaneous melanoma data from The Cancer Genome Atlas (TCGA). Single-sample gene set enrichment analysis (ssGSEA) was used to calculate immune enrichment scores of 28 immune cell signatures. We performed pairwise comparisons to obtain scores for cell pairs based on the difference in the abundance of immune cells within each sample. The resulting cell pair scores, in the form of a matrix of relative values of immune cells, formed the core of the IRRS.

Results: The area under the curve (AUC) for the IRRS was over 0.700, and when the IRRS was combined with clinical information, the AUC reached 0.785, 0.817, and 0.801 for the 1-, 3-, and 5-year survival, respectively. Differentially expressed genes between the two groups were enriched in staphylococcal infection and estrogen metabolism pathway. The low IRRS group showed a better immunotherapeutic response and exhibited more neoantigens, richer T-cell receptor and B-cell receptor diversity, and higher tumor mutation burden.

Conclusion: The IRRS enables a good prediction of prognosis and immunotherapy effect, based on the difference in the relative abundance of different types of infiltrating immune cells, and could provide support for further research in melanoma.

KEYWORDS

cutaneous melanoma, cell pair, tumor infiltrating immune cell, prognosis model, immunotherapy response

10.3389/fimmu.2023.1112181

Introduction

Cutaneous melanoma is a highly malignant tumor derived from melanocytes and is the most invasive and complex of all skin cancers (1). In 2020, the total number of new melanoma cases in the world was 325,000 with 57,000 deaths; these numbers are predicted to increase to 510,000 new cases with 96,000 deaths by 2040 (2). The occurrence of melanoma is caused by interactions between genetic susceptibility and environmental exposure (3), that is, an accumulation of genomic changes, including the mutation burden driven by high-intensity ultraviolet light and prolonged exposure to ultraviolet, which makes melanoma the most immunogenic tumor type with the ability to induce an immune response that can inhibit melanoma growth (4, 5). Immune checkpoint inhibitors, whose main targets are programmed cell death protein 1 (PD1), programmed cell death 1 ligand 1 (PDL1), and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), have been successfully used in the treatment of melanoma. The total effective rate of immune checkpoint inhibitors in patients with advanced melanoma is 32.9%–58.0% (6). However, only a third of melanoma patients show a durable response to immune checkpoint therapies (7). Biomarkers for the prediction of prognosis and immunotherapy effect in melanoma patients remain elusive. However, previous studies have shown that cytotoxic T lymphocyte (CTL) dysfunction and exhaustion result in lower response and sensitivity to immunotherapy (8). This means that the immune microenvironment is closely related to the effectiveness of immune checkpoint inhibitors.

Tumor-infiltrating immune cells (TIICs), including T cells, B cells, macrophages, and natural killer cells, form an important component of most solid tumors and have an essential role in the host antitumor immune response, which can affect tumor progression *via* antitumor activity or immunosuppression (9, 10). During the process of tumor development, including elimination, balance, and escape, the dual function and plasticity of TIICs lead to complexity and changes in the antitumor response (11, 12). For example, in many tumor types, patients with high levels of $CD8^+$ T-cell infiltration tend to have a better prognosis. On the contrary, patients with obvious infiltration of immunosuppressive cells, such as regulatory T cells, tend to have a worse prognosis. Therefore, the quantity and quality of TIICs are key determinants of prognosis (9). The value of TIICs in prognosis prediction and drug resistance analysis has been verified in a variety of tumors, including melanoma (7, 13–15). The American Joint Committee on Cancer (AJCC) guidelines are widely used to evaluate the prognosis of melanoma patients. However, TNM staging mainly describes the invasion and metastasis of tumor tissue at the pathological level, which cannot take into account the composition of tumor-infiltrating cells in the immune microenvironment. Although there are many prognostic models that incorporate immune gene expression, few studies have constructed prognostic models directly based on TIICs. This may be because of the different methods used for determining the specific content of infiltrating cells, which are affected by various measurement factors such that it is difficult to establish a unified standard.

In this study, we develop a prognostic prediction model for melanoma based on TIICs. We adopt the relative value of cell fraction to form a cell pair algorithm. In addition, we present an online nomogram, of which the IRRS is the core, including clinical indicators, to facilitate the use of the IRRS by clinicians.

Materials and methods

Study design and data collection

The integrated research design is presented in Figure 1. Transcription profiles and clinical data of cutaneous melanoma patients were obtained from The Cancer Genome Atlas (TCGA; https://portal.gdc.cancer.gov/; TCGA-SKCM cohort). After removing cases with duplication, lack of expression profiles, or lack of survival data, the data of 458 patients were included in the training group for the construction of the IRRS score. The GSE65904, GSE54467, GSE91061, and GSE115821 datasets from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and a cohort from Liu et al. were used as testing sets for validation (16–20). Missing values in the clinical or pathological data of patients were filled using the missForest package (21–28).

Establishment and validation of the cell pair algorithm

We carried out single-sample gene set enrichment analysis (ssGSEA) to analyze the expression of corresponding markers of 28 immune cell types (29), thereby obtaining the abundance of these cell types in patient tumor tissues. Then, the cells related to prognosis were screened by univariate Cox (uni-Cox) regression ($P< 0.05$).

The prognosis-related immune cells were termed A cells, and the A cells were paired with all 28 immune cells (termed B cells) in turn to form a set of A|B pairs. If the A-cell abundance exceeded the B-cell abundance for a given cell pair, the value of that pair was recorded as 1; otherwise, it was recorded as 0. This method enables the relative cell abundance to be considered without dependence on the absolute number; this avoids the variation caused by the use of different methods for gene measurement and annotation and differential cell abundance analysis. A matrix containing values of 0 or 1 was constructed, from which cell pairs with 0 or 1 accounting for more than 80% of the total were removed. In the human body, the content of some immune cells is much higher than that of other immune cells, such as neutrophils. The remaining cell pairs were screened by uni-Cox regression analysis ($P< 0.05$) to obtain those correlated with prognosis. We applied the least absolute shrinkage and selection operator (LASSO) Cox regression analysis (glmnet package) to avoid overfitting and obtain the remaining cell pairs. Then, each cell pair was assigned the optimal coefficient by multivariate Cox (multi-Cox), and the IRRS was generated as follows:

$$IRRS = \sum Score\ \ A|B \times Coef\_A|B$$

The receiver operating characteristic (ROC) curves, the Kaplan–Meier survival curves, the GEO datasets, and the cohort from Liu et al. were used to verify the effectiveness of the IRRS in predicting prognosis and immunotherapy effect.

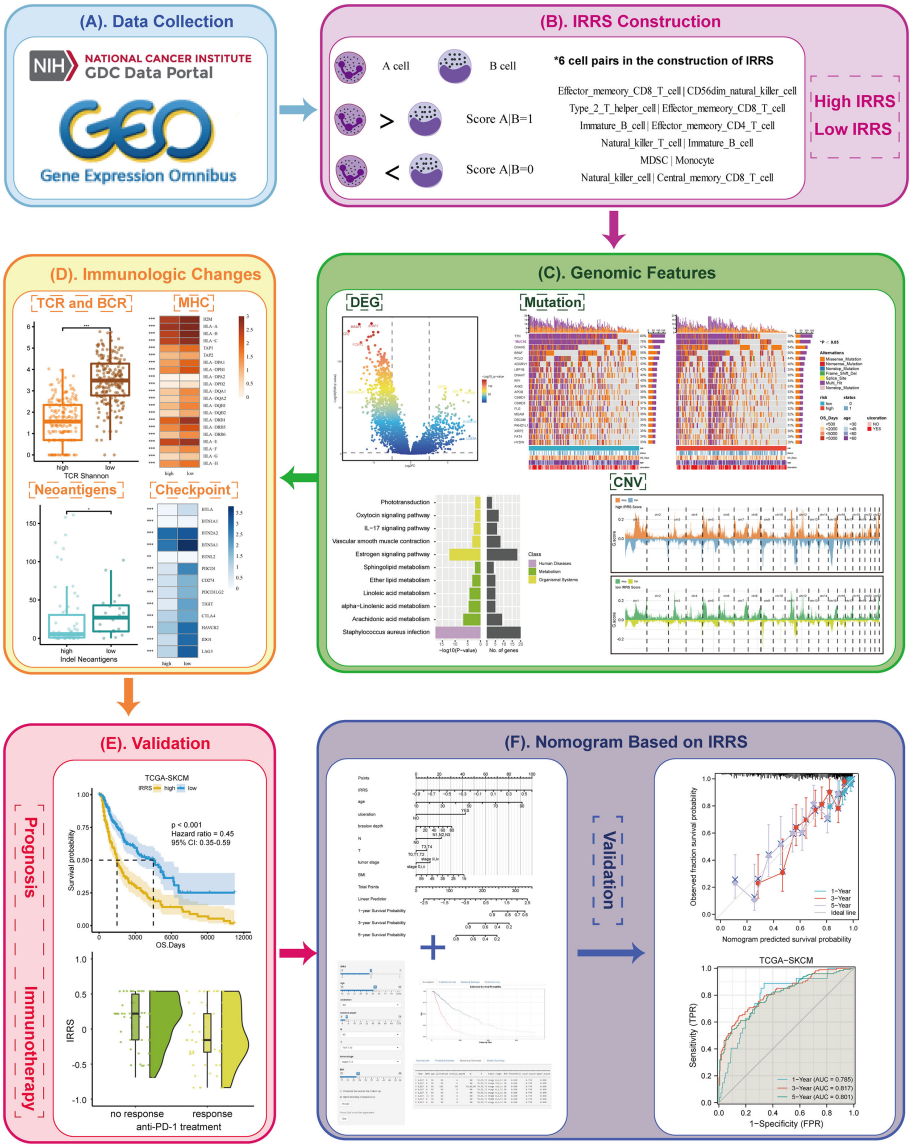


FIGURE 1
Overview of the workflow. **(A)** The data of this study were from The Cancer Genome Atlas (TCGA) and GEO cohorts. **(B)** The immune-related risk score (IRRS) was constructed by the relative value of cell abundance, that is, cell pairs. **(C)** Discovered the difference in the genomic features between the high and low IRRS through DEG, mutation, and CNV. **(D)** Immunologic changes based on the IRRS was analyzed by the TCR and BCR, MHC, neoantigens, and checkpoints. **(E)** The prognosis and immunotherapy predictive effects were confirmed. **(F)** A nomogram based on the IRRS was constructed and validated.

Differentially expressed genes and analysis

The differentially expressed genes between the high and low IRRS groups were analyzed using the DESeq2 package, with threshold |log2 fold change (FC)| ≥2 and Benjamini–Hochberg-adjusted *P*-value<0.05 (30). Functional enrichment analysis and clustering of the identified biological processes were conducted using the clusterProfiler R package (31).

The main regulator (MR) is a gene located at the hub of a regulatory network that controls a large number of target genes (termed as its regulon). We used the MR4Cancer tool (http://cis.hku.hk/MR4Cancer) to identify potential MRs that could explain the DEGs between the high and low IRRS groups (32). An MR network diagram was drawn using Cytoscape.

Genomic features

We used the maftools package to draw the OncoPrint, and the Fischer test was used to evaluate differences in gene mutation frequency between the two groups (33). The somaticInteractions function in the maftools package was used to accurately evaluate the exclusive occurrence and co-occurrence of mutations in pairwise comparisons of the 25 genes with the highest mutation frequency. The DeconstructSigs package was used to analyze the cosmic mutation signature of each patient (34).

Significant deletion or amplification events in the regions of the genome were investigated with GISTIC 2.0, a revised computational program used to identify somatic copy number alterations (35).

Immunologic changes

T-cell receptor (TCR), B-cell receptor (BCR), and neoantigen data were from the research of Thorsson et al. (36). Tumor immune dysfunction and exclusion (TIDE) score and microsatellite instability (MSI) score were obtained using the official TIDE website (http://tide.dfci.harvard.edu/) (37). The statistical significance of the MSI score was evaluated by Welch's *t*-test, and other indexes were evaluated by the Wilcoxon rank-sum test.

Construction and validation of the nomogram model

A nomogram was constructed to predict specific outcomes based on the IRRS and clinical variables using the rms package (38). ROC curves, calibration curves, and decision curve analysis (DCA) curves were drawn to verify the reliability of the nomogram. In addition, the nomogram was compared with the traditional TNM staging system by calculating the integrated discrimination improvement (IDI). Finally, the Dynnom package (cran.r-project.org/web/packages/rms) was used to generate an online version of the nomogram model with an interactive interface for clinical applications.

Results

Construction and validation of the IRRS

A total of 28 immune cell types from 458 melanoma patients (TCGA data) were analyzed. A total of 19 immune cell types related to prognosis were identified by uni-Cox analysis ($P< 0.05$) (Figure 2A). After pairing, 532 immune cell pairs were screened and entered into LASSO Cox regression analysis, and 11 immune cell pairs were retained (Figures S1, S2).

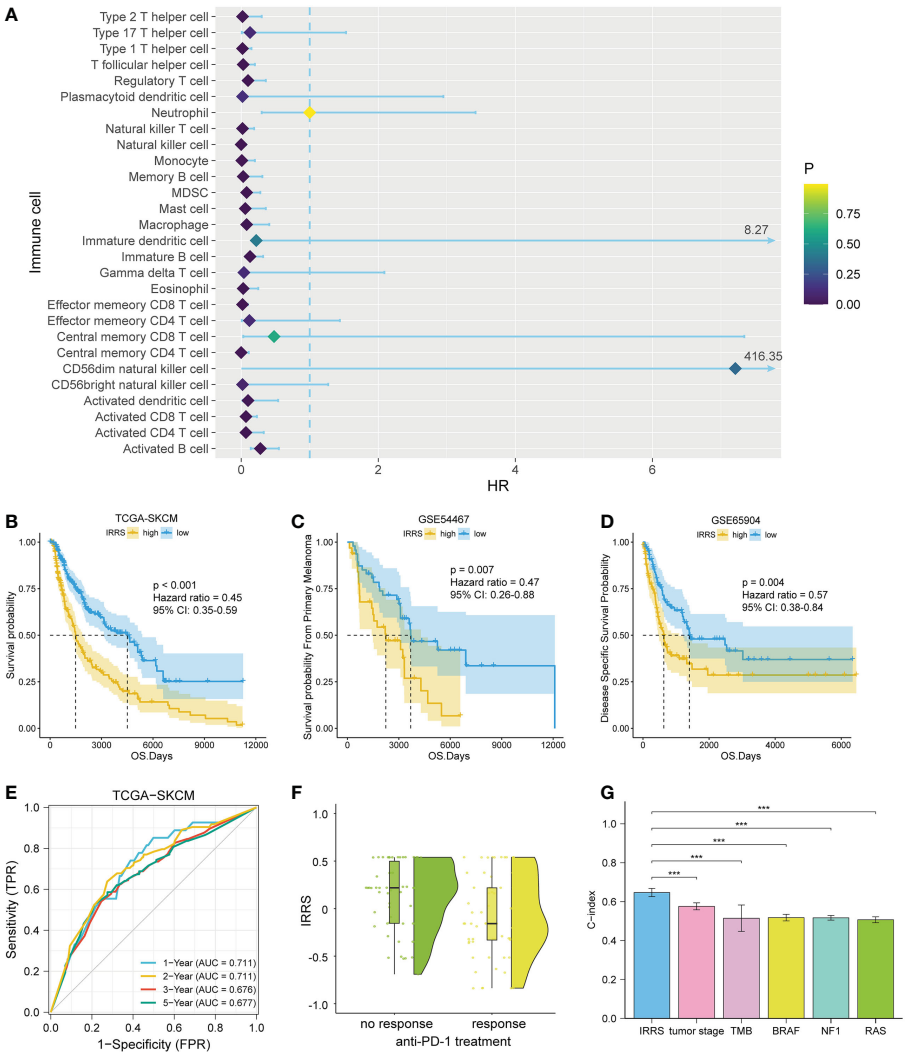


FIGURE 2
Construction and validation of the IRRS. **(A)** The 19 candidate cells screened based on uni-Cox analysis. **(B–D)** The Kaplan–Meier curves of survival probability for patients in the TCGA-SKCM, GSE54467, and GSE65904 cohorts. **(E)** The ROC curve for patients in the TCGA-SKCM cohort. **(F)** Comparison of immunotherapeutic responses ($P< 0.01$) for patients in the GSE91061, GSE115821, and Liu et al. cohorts. **(G)** Comparison of C-index between the IRRS and tumor stage, TMB, and driver mutations (BRAF, NF1, and RAS) in the TCGA.

We then carried out multi-Cox regression to generate the best coefficients for the corresponding immune cell pairs; only six immune cell pairs were included in the final model (Table 1). Patients with melanoma were classified into high IRRS or low IRRS groups based on the median IRRS score. Compared with patients in the high IRRS group, patients in the low IRRS group had longer overall survival (hazard ratio = 0.45, 95% confidence interval 0.35–0.59, log-rank test $P< 0.001$). To confirm the predictive effect of the IRRS, an ROC curve analysis was performed; the area under the ROC curve (AUC) values were 0.711, 0.711, 0.676, and 0.677 for the 1-, 2-, 3-, and 5-year survival, respectively (Figures 2B, E).

To further assess the reproducibility and validity of the IRRS, we used external datasets, including GSE65904 and GSE54467, to validate its prognostic value. We also used the median as a group point to plot the Kaplan–Meier curve. Notably, the patients in the high-risk group had shorter overall survival. In addition, in the three anti-PD1 treatment cohorts of GSE91061, GSE115821, and Liu et al., patients with low IRRS exhibited significantly better immunotherapeutic response ($P< 0.01$) (Figures 2C, D, F).

The independent predictive ability of the IRRS

To estimate whether the IRRS was independent of other clinical or pathological factors of melanoma patients, multi-Cox regression was performed, in which covariables included age, gender, body mass index, ulceration, Breslow depth, Clark level, T stage, N stage, M stage, tumor stage, and the IRRS. Multi-Cox analysis showed that the IRRS, age, and ulceration were independent predictive factors for the prognosis of melanoma patients (Table 2). The C-index of the IRRS was higher than those of the other independent predictive factors (0.647 for the IRRS *vs*. 0.600 and 0.626 for age and ulceration, respectively). To further confirm the predictive performance of the IRRS, we also compared the C-index values for the IRRS with those for tumor stage, tumor mutation burden (TMB), and driver mutations (BRAF, NF1, and RAS); the results showed that the IRRS had the best predictive effect with respect to prognosis (Figure 2G).

Furthermore, the high IRRS group had significantly worse overall survival than the low IRRS group, regardless of whether the patients were in the early or late TNM stages (Figure 3).

Enrichment analysis of differentially expressed genes

Screening identified 422 upregulated genes and 915 downregulated genes in the high-risk group compared with the low-risk group ($|\log2\ FC| > 2$, $P< 0.05$) (Figure 4A). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the differentially upregulated genes showed that these genes were mainly enriched in *Staphylococcus aureus* infection and estrogen signaling pathway (Figure 4B). In the low IRRS group, GSEA showed significant enrichment, with enrichment scores over 0.7 in 22 pathways, including 12 immune-related pathways. In addition to *S. aureus* infections, some pathways related to viral infection have also been enriched. Figure 4C shows the eight immune-related pathways with the highest enrichment scores.

We used MR4Cancer to identify the MRs, which were prioritized based on DEGs through overrepresentation analysis and GSEA. Among them, the eight transcriptional regulators with the most nodes were selected and used to plot a network of the MRs and DEGs from the MRs obtained by the analysis of upregulated DEGs in the high IRRS group (Figure 4D). Notably, FOXN1 was found to orchestrate the expression of 236 DEGs.

Genomic features of the IRRS

Based on the maftools analysis, the top 20 most frequently altered genes were identified (Figure 5A). Among these genes, MUC16 was more frequently altered in the low IRRS group, and NRAS–BRAF was the most mutually exclusive pair in the high IRRS group. On the other hand, the most frequently co-occurring gene changes in the low IRRS group were for the pair RP1–MUC16 (Figure 5C).

Then, we analyzed the cosmic mutation signatures and found that signature 7 was higher in the low-risk group, which was related to ultraviolet radiation. In addition, the low IRRS group showed higher TMB and MSI (Figure 5B).

The IRRS and immunological mechanism

Immune checkpoint blockade has become a promising strategy for the treatment of many cancers. Therefore, we studied the expression of key immune checkpoint molecules, including PDCD1,

TABLE 1 The final immune cell pairs and the corresponding coefficient generated by multi-Cox regression.

Immune cell pairs	Coefficient
Effector_memory_CD8+T_cell|CD56 dim_natural_killer_cell	−0.201895537
Type_2_T_helper_cell_|Effector_memory_CD8_T_cell	0.162775185
Immature_B_cell_|Effector_memory_CD4+T_cell	−0.167620476
Natural_killer_T_cell_|Immature_B_cell	0.376203094
MDSC_|_Monocyte	−0.146018787
Natural_killer_cell|Central_memory_CD8_T_cell	−0.32046159

TABLE 2 Univariable and multivariable Cox regression analyses of the IRRS and survival in the TCGA cohort.

Characteristics	Total (*N*)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*P*-value	Hazard ratio (95% CI)	*P*-value
IRRS	458	2.718 (2.030–3.639)	**<0.001**	2.899 (2.123–3.959)	**<0.001**
Age	458	1.025 (1.015–1.034)	**<0.001**	1.018 (1.008–1.028)	**<0.001**
Gender	458				
Male	284	Reference			
Female	174	0.878 (0.662–1.164)	0.365		
BMI	458	0.965 (0.931–1.000)	**0.048**	0.985 (0.952–1.020)	0.400
Ulceration	458				
No	214	Reference			
Yes	244	2.523 (1.907–3.338)	**<0.001**	1.970 (1.458–2.662)	**<0.001**
Breslow depth	458	1.026 (1.013–1.040)	**<0.001**	1.006 (0.989–1.024)	0.486
M	458				
M0	435	Reference			
M1	23	1.752 (0.926–3.316)	0.085		
N	458				
N0	277	Reference			
N1, N2, N3	181	1.710 (1.292–2.262)	**<0.001**	1.416 (0.615–3.263)	0.414
T	458				
T0, T1, T2	149	Reference			
T3, T4	309	1.738 (1.301–2.324)	**<0.001**	1.159 (0.843–1.593)	0.363
Tumor stage	458				
Stage 0, I, II	263	Reference			
Stage III, IV	195	1.654 (1.253–2.182)	**<0.001**	1.212 (0.530–2.770)	0.649

The bold values represents P-value < 0.05, that is, the relevant prognostic predictive factors are statistically significant.

CD247, PDCD1LG2, TIGIT, CTLA-4, HAVCR2, IDO1, and LAG3. Cutaneous melanoma patients with low IRRS scores had a higher expression of immune checkpoint molecules, indicating that patients in the low-risk group were more likely to exhibit better immunotherapeutic responses (Figure 5D).

We also analyzed adaptive receptors, including TCR and BCR Shannon diversity, and new antigens, including single nucleotide variant (SNV) and indel new antigens. The results showed that the group with low IRRS had higher TCR and BCR diversity and more new antigens (Figure 5B). HLA genes control the adaptive immune response by presenting antigens to T cells. The antigen-presenting genes that we analyzed all showed high expression in the low IRRS group (Figure 5D).

TIDE uses T-cell dysfunction and exclusion markers to simulate immune escape in tumors with different CTL levels, which can be used to predict the effects of immunotherapy. The high IRRS group had a higher TIDE score, indicating that the patients in the high-score group would have a poorer response to immunotherapy than those in the low-score group ($P< 0.05$).

Copy number variation

Significant differences in copy number variation were detected between the high IRRS and low IRRS groups (Figure 6A). Importantly, focal amplification peaks in some immune-related gene areas were observed in the low IRRS group, such as PDCD1LG2(9p24.1) (Figures 6B, C). We annotated specific amplified genes in the high and low IRRS groups through gene ontology biological processes and then clustered the top 10 biological processes. Compared with those in the high IRRS group, the genes amplified in the low IRRS group were more enriched in immune-related processes (Figures 6D, E).

Nomogram based on the IRRS

Four independent prognostic clinical characteristics associated with overall survival were identified by uni-Cox analysis ($P< 0.05$) and multi-Cox regression (Figure 7A). These factors, which comprised age,

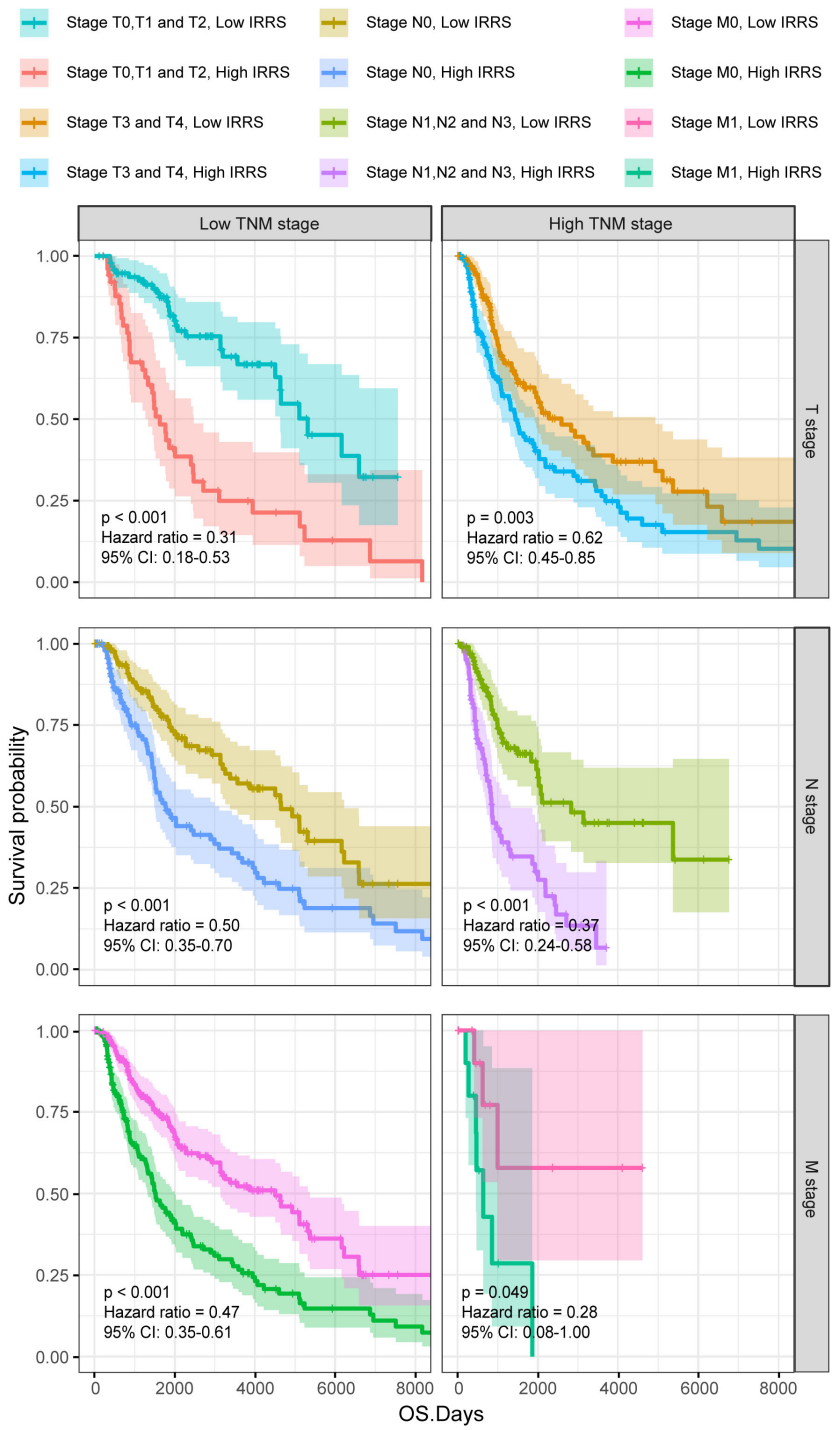


FIGURE 3
The Kaplan–Meier survival curves according to different TNM stages of patients from the TCGA-SKCM classified into high- and low-risk groups based on the IRRS score.

ulceration, Breslow depth, and N stage, were combined with the IRRS score and used to construct a nomogram to quantitatively estimate the survival rate of patients with cutaneous melanoma (Figure 7B).

Although a nomogram can enable visualization of a prognostic model to a great extent, it still requires a ruler to measure continuous variables, which may lead to error because of subjective judgment. Therefore, we developed an online website for clinicians (https://irrsmelanoma.shinyapps.io/IRRSSKCM/), where the predicted survival rate can be easily determined by inputting values of clinical factors (Figure 7C).

The calibration results of our nomogram were intuitively consistent with the actual calibration results (Figure 7F). The ROC curve analysis indicated that the nomogram had a good predictive effect on prognosis (the AUC values for the 1-, 3-, and 5-year survival were 0.785, 0817, and 0.801, respectively) (Figure 7D). According to the DCA curve, the nomogram had the highest overall net benefit

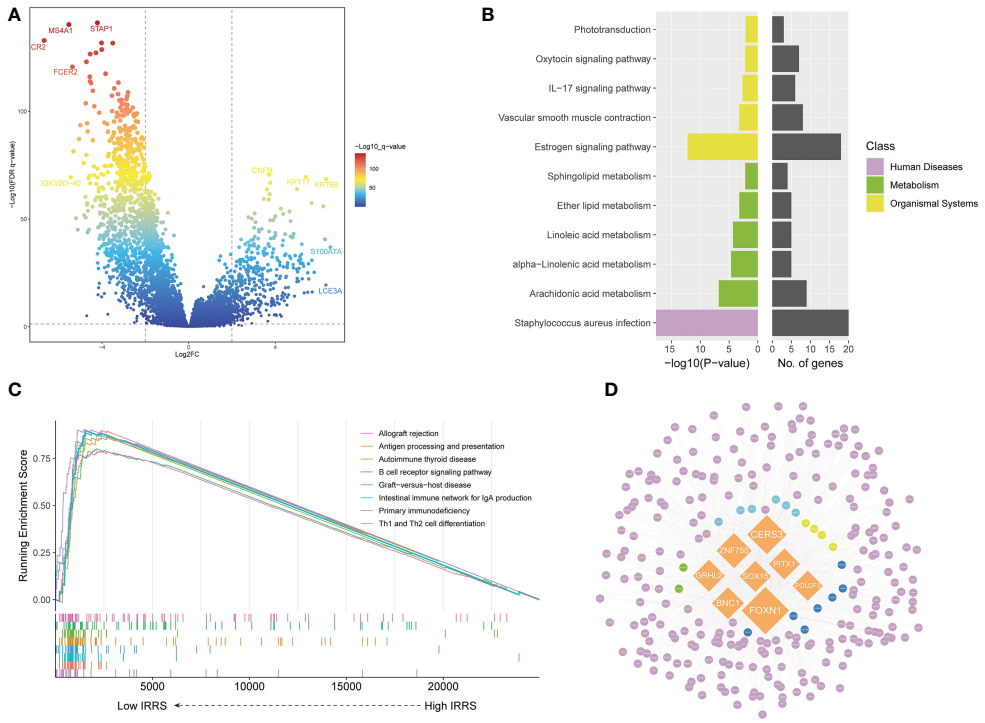


FIGURE 4
Screening of differentially expressed genes and the master regulator. **(A)** Volcano plot of differentially expressed genes between the low- and high-risk groups in the TCGA cohort. **(B)** KEGG enrichment of differentially expressed genes. **(C)** Gene set enrichment analysis of the high IRRS and low IRRS groups. **(D)** Network of the MRs and DEGs upregulated in the high IRRS group. Orange: eight MRs with the most nodes. Genes related to the KEGG enrichment pathway corresponding to each color: yellow, *Staphylococcus aureus* infection; green, estrogen signaling pathway; blue, both above; dark blue, arachidonic acid metabolism; purple, other DEGs.

within the threshold probability compared with a separate IRRS score or a separate clinical feature (Figure 7E). In addition, we compared the nomogram based on the IRRS with traditional AJCC TNM staging; the IDI values for 3, 5, and 10 years were 26.9% ($P< 0.001$), 28.8% ($P< 0.001$), and 32.5% ($P< 0.001$), respectively.

Discussion

In our study, first, we focused on the influence of TIICs on the prognosis of melanoma patients. Although there have been previous studies exploring the predictive value of immune genes, few prognosis models focusing on the components and content of tumor-infiltrating cells have been established (39–42). Then, we established a cell pair score matrix generated by comparing the abundance of immune cells in pairs, and the IRRS was constructed on the basis of this matrix. Finally, we not only verified the predictive ability of the IRRS but also analyzed its immune and genetic characteristics. An online nomogram integrating the IRRS and clinical information was constructed to avoid errors caused by the measurement process used by clinicians and for further visualization.

The advantages of our model are as follows. First, errors may be introduced into the models by the use of different methods of gene sequencing, continuous updating of annotations, different methods for infiltrating cell analysis, and batch effects; however, the proportion of TIICs exists in a relatively stable range. The adoption of relative values avoids the abovementioned issues. Second, the construction of cell pairs also enables the consideration of interactions between cells and personal immune factors. Therefore, this method improved the predictive ability of the IRRS. In the verification using cohorts from the GEO, the IRRS showed good prediction efficiency with respect to prognosis and immunotherapy response. Through multi-Cox analysis and stratification analysis of the IRRS, we could confirm that the IRRS was independent of other clinical or pathological factors.

In the differential gene expression analysis and KEGG enrichment analysis, the upregulated genes in the high IRRS group were found to mainly affect *S. aureus* infection, estrogen signaling pathway, and pathways related to lipid metabolism (arachidonic acid metabolism, linoleic acid metabolism, etc.). Previous studies have shown that increased colonization of *S. aureus* in squamous cell carcinoma might promote carcinogenesis by inducing chronic skin inflammation (43). Lutchminarian et al. reported a role of pathogenic bacteria in increasing the risk of postoperative complications (44). However, there have been few studies on the direct induction of melanoma carcinogenesis by epidermal microbiota, and whether the change in skin microbiota is the cause or result of melanoma remains to be studied (45). There are gender differences in the incidence of melanoma. The mortality, recurrence, and metastasis rates of melanoma in pregnant women have been found to be higher than those in a non-pregnant control group. Moreover, melanoma-related mortality and sentinel node positivity are higher in

10.3389/fimmu.2023.1112181

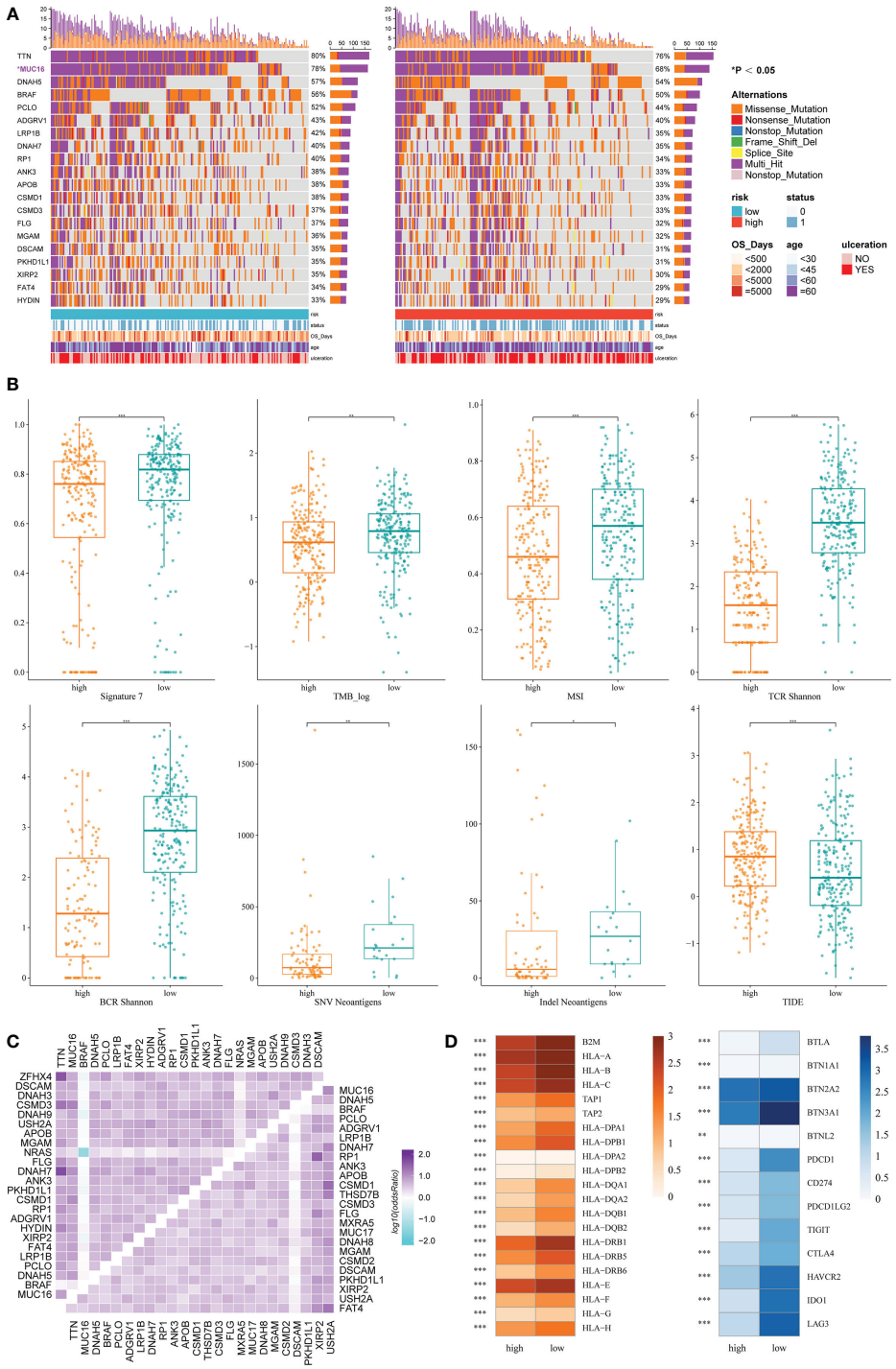


FIGURE 5
Genomic features and immunologic changes of the high- and low-score groups. **(A)** Mutation of top 20 most frequently altered genes in melanoma patients with high and low IRRS. **(B)** Cosmic mutation signature 7, tumor mutation burden, microsatellite instability, neoantigens, and TIDE score in the high- and low-score groups. **(C)** Heatmap depicting the co-occurrence or exclusivity of the top 25 most mutated genes in the high IRRS group (left upper corner) and the low IRRS group (lower right corner). **(D)** Association between HLA and immune checkpoint molecules and the IRRS. *P<0.05 **P<0.01 ***P<0.001.

women aged 40 to 49 (46). These results suggest that increased estrogen is closely related to the occurrence of melanoma (47). In addition, Conforti et al. confirmed that estrogen could resist the effects of immune checkpoint inhibitors by promoting macrophage polarization (48). A variety of fatty acids are related to the occurrence and development of cancer. An abnormal arachidonic acid metabolic pathway is mainly due to the activation of the COX and LOX pathways, which further affects the occurrence of inflammation and cancer (49). COX-1, COX-2, and LOX are the main drug inhibitor targets of this pathway (50). With the increasing use of immunotherapy, there are excellent prospects for combination treatments involving inhibitors of this pathway acting on specific alkyl receptors (51). Linoleic acid and α-linoleic acid reduce the

10.3389/fimmu.2023.1112181

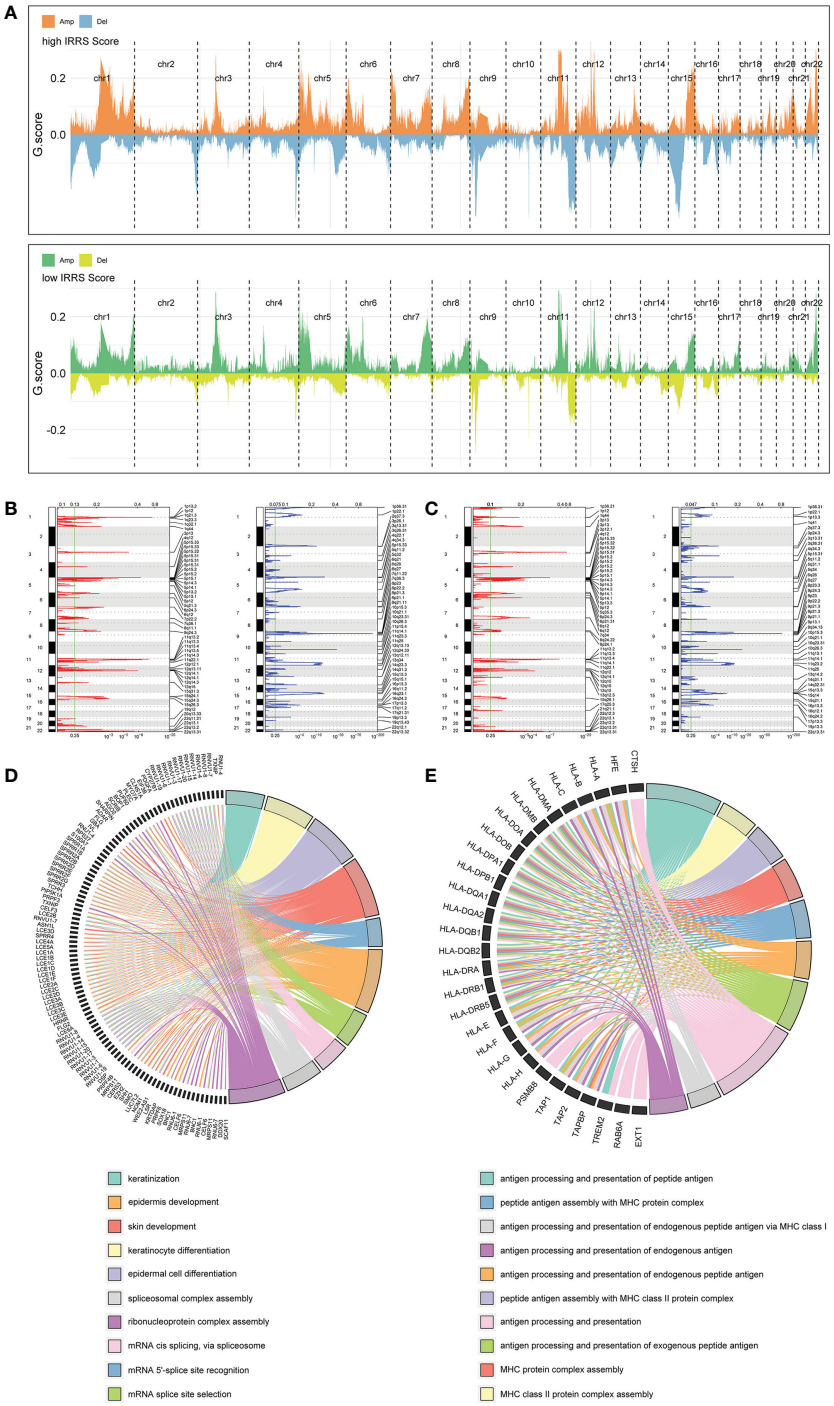


FIGURE 6
Copy number alterations in the high- and low-score groups. **(A)** Copy number profiles of the high IRRS score (above) and low IRRS score (below) groups. **(B)** Detailed cytobands with focal amplification (red) and deletion (blue) peaks identified in the high IRRS group. **(C)** Detailed cytobands with focal amplification (red) and deletion (blue) peaks identified in the low IRRS group. **(D)** Circular plot of the top 10 biological processes and corresponding enriched genes in the high IRRS. **(E)** Circular plot of the top 10 biological processes and corresponding enriched genes in the low IRRS.

production of melanin by melanocytes (52). Thus, lipid-related metabolic pathways may represent therapeutic targets in malignant melanoma. In addition, in the high IRRS score group, 236 upregulated DEGs were regulated by FOXN1 as a master regulator. FOXN1 plays an important part in wound healing (53). A possible reason for this upregulation of FOXN1 is that melanoma patients in the high-risk group tend to have worse tumor progression and often develop skin ulceration. Our findings about the MRs may provide new therapeutic targets and potential approaches to treat patients with malignant melanoma.

Antigen presentation ability, tumor immunogenicity, and gene changes can all affect the immune activity of tumors and influence the effectiveness of immunotherapy (54). The high immunogenicity of melanoma makes tumor immunotherapy with checkpoint inhibitors an important treatment option for advanced melanoma patients. The

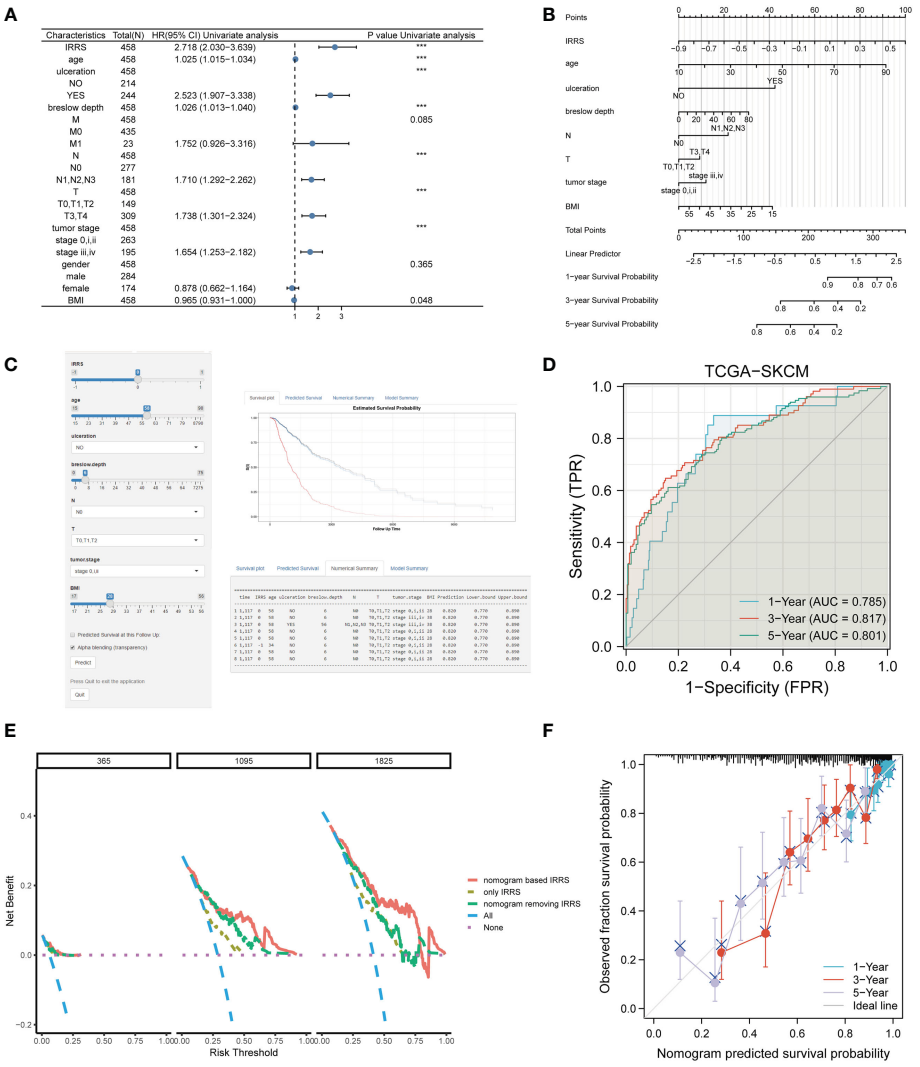


FIGURE 7
Construction of the nomogram. **(A)** The prognostic clinical factors screened based on uni-Cox regression. **(B)** The nomogram for predicting the survival rate of melanoma patients, including four independent clinical prognostic factors and the IRRS. **(C)** The online version of the nomogram. **(D)** The ROC analysis of the nomogram. **(E)** DCA of the nomogram. **(F)** The calibration curve of the nomogram. ***P<0.001.

higher TCR, BCR, and HLA diversity in the low IRRS group suggested higher antigen presentation ability in this group. Moreover, the higher levels of SNV or indel neoantigens in the low IRRS were the result of tumor-specific mutations, which determine tumor immunogenicity and increase responsiveness to checkpoint inhibitors (54, 55). In many solid tumors, MSI-H and high TMB are biomarkers of therapeutic benefit (56–58). The low IRRS group had a higher median value for both of these indicators, demonstrating a higher frequency of gene mutation, especially in genes related to ultraviolet exposure (mutation signature 7), which is related to increased sensitivity to checkpoint inhibitor drugs (59, 60). We also analyzed several important immune checkpoints that are related to tumor cell apoptosis (61), T-cell co-inhibition signal, lymphocyte activation (62), and T-cell immunoglobulin mucin (63). The expression levels of immune checkpoints in the high IRRS group were significantly lower than those in the low IRRS group, indicating that the low IRRS group may show a better response to immunotherapy.

In addition, the high IRRS group showed mutual exclusion of NRAS and BRAF. Previous studies have suggested a low incidence of NRAS–BRAF combined mutation, especially in soft tissue malignant melanoma (64). Kumar et al. reported exclusivity between BRAF and NRAS mutations in melanoma, and SPRY4 was a potential mediator of this synthetic response to dual oncogene inhibition (65). Petti et al. showed that the forced expression of NRAS in a single BRAF melanoma line led to growth arrest, that is, when the two mutations coexisted, the viability of cancer cells was impaired (66). On the one hand, this is consistent with our results in the high IRRS score group; that is, there was a higher degree of NRAS–BRAF mutual exclusion in the group with a poor prognosis. On the other hand, the coexistence of double mutations indicates a potential new approach to the treatment of melanoma.

In conclusion, we have introduced the use of relative values, established the IRRS as a prognostic indicator for melanoma, and provided insight into the role of TIICs in the occurrence and development of melanoma and the effects of immunotherapy.

Conclusion

The IRRS shows a good ability to predict prognosis and immunotherapy effect in melanoma, based on differences in the relative abundance of different types of TIICs, and could provide support for further research in melanoma.

Data availability statement

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

Author contributions

Conceptualization: ML and XL. Methodology: ML, XL and GZ. Acquisition of data: ML, XL, WB, and GD. Data analysis: ML, XL and WB. Validation: YL. Revision of the article: JS, WB, KH and GZ. Writing of the original draft: ML and XL. Visualization: YL, GD and KH. All authors contributed to the article and approved the submitted version.

Funding

The work was supported by the Innovation and Entrepreneurship Funding Project of Grade 2020 of Central South University (Innovation Project:2022105330036); The Project of Intelligent Management Software for Multimodal Medical Big Data for New Generation Information Technology, Ministry of Industry and Information Technology of People's Republic of China (TC210804V); and The Science and Technology Innovation Program of Hunan Province (2020RC3059).

Acknowledgments

We acknowledge the TCGA and GEO databases for providing their platforms and the contributors for uploading their meaningful datasets.

Conflict of interest

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

The reviewer MP declared a shared affiliation with the authors ML, XL, GZ, GD, YL, JS and KH to the handling editor at time of review.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1112181/full#supplementary-material

References

1. Ali Z, Yousaf N, Larkin J. Melanoma epidemiology, biology and prognosis. *EJC Suppl.* (2013) 11(2):81–91. doi: 10.1016/j.ejcsup.2013.07.012

2. Arnold M, Singh D, Laversanne M, Vignat J, Vaccarella S, Meheus F, et al. Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA Dermatol* (2022) 158(5):495–503. doi: 10.1001/jamadermatol.2022.0160

3. Dzwierzynski WW. Melanoma risk factors and prevention. *Clinics Plast Surg* (2021) 48(4):543–50. doi: 10.1016/j.cps.2021.05.001

4. Jespersen H, Lindberg MF, Donia M, Söderberg EMV, Andersen R, Keller U, et al. Clinical responses to adoptive T-cell transfer can be modeled in an autologous immune-humanized mouse model. *Nat Commun* (2017) 8(1):707. doi: 10.1038/s41467-017-00786-z

5. Vogelsang M, Martinez CN, Rendleman J, Bapodra A, Malecek K, Romanchuk A, et al. The expression quantitative trait loci in immune pathways and their effect on cutaneous melanoma prognosis. *Clin Cancer Res* (2016) 22(13):3268–80. doi: 10.1158/1078-0432.CCR-15-2066

6. Furue M, Ito T, Wada N, Wada M, Kadono T, Uchi H. Melanoma and immune checkpoint inhibitors. *Curr Oncol Rep* (2018) 20(3):29. doi: 10.1007/s11912-018-0676-z

7. Somasundaram R, Zhang G, Fukunaga-Kalabis M, Perego M, Krepler C, Xu X, et al. Tumor-associated b-cells induce tumor heterogeneity and therapy resistance. *Nat Commun* (2017) 8(1):607. doi: 10.1038/s41467-017-00452-4

8. Farhood B, Najafi M, Mortezaee K. CD8+ cytotoxic T lymphocytes in cancer immunotherapy: A review. *J Cell Physiol* (2019) 234(6):8509–21. doi: 10.1002/jcp.27782

9. Hiraoka N. Tumor-infiltrating lymphocytes and hepatocellular carcinoma: Molecular biology. *Int J Clin Oncol* (2010) 15(6):544–51. doi: 10.1007/s10147-010-0130-1

10. Veltman JD, Lambers ME, van Nimwegen M, Hendriks RW, Hoogsteden HC, Hegmans JP, et al. Zoledronic acid impairs myeloid differentiation to tumour-associated macrophages in mesothelioma. *Br J Cancer.* (2010) 103(5):629–41. doi: 10.1038/sj.bjc.6605814

11. Twitty CG, Huppert LA, Daud AI. Prognostic biomarkers for melanoma immunotherapy. *Curr Oncol Rep* (2020) 22(3):25. doi: 10.1007/s11912-020-0886-z

12. Pinton L, Solito S, Masetto E, Vettore M, Canè S, Puppa AD, et al. Immunosuppressive activity of tumor-infiltrating myeloid cells in patients with meningioma. *Oncoimmunology.* (2018) 7(7):e1440931. doi: 10.1080/2162402X.2018.1440931

13. Pagès F, Berger A, Camus M, Sanchez-Cabo F, Costes A, Molidor R, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med* (2005) 353(25):2654–66. doi: 10.1056/NEJMoa051424

14. Kadara H, Choi M, Zhang J, Parra ER, Rodriguez-Canales J, Gaffney SG, et al. Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. *Ann Oncol* (2017) 28(1):75–82. doi: 10.1093/annonc/mdw436

15. Ruiz de Galarreta M, Bresnahan E, Molina-Sánchez P, Lindblad KE, Maier B, Sia D, et al. β-catenin activation promotes immune escape and resistance to anti-PD-1 therapy in hepatocellular carcinoma. *Cancer Discovery* (2019) 9(8):1124–41. doi: 10.1158/2159-8290.CD-19-0074

16. Cirenajwis H, Ekedahl H, Lauss M, Harbst K, Carneiro A, Enoksson J, et al. Molecular stratification of metastatic melanoma using gene expression profiling: Prediction of survival outcome and benefit from molecular targeted therapy. *Oncotarget.* (2015) 6(14):12297–309. doi: 10.18632/oncotarget.3655

17. Jayawardana K, Schramm SJ, Haydu L, Thompson JF, Scolyer RA, Mann GJ, et al. Determination of prognosis in metastatic melanoma through integration of clinico-pathologic, mutation, mRNA, microRNA, and protein information. *Int J Cancer.* (2015) 136(4):863–74. doi: 10.1002/ijc.29047

18. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell.* (2017) 171 (4):934–49. doi: 10.1016/j.cell.2017.09.028

19. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med* (2018) 24(10):1545–9. doi: 10.1038/s41591-018-0157-9

20. Liu D, Schilling B, Liu D, Sucker A, Livingstone E, Jerby-Arnon L, et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat Med* (2019) 25(12):1916–27. doi: 10.1038/s41591-019-0654-5

21. Carmona CP, Bueno CG, Toussaint A, Träger S, Díaz S, Moora M, et al. Fine-root traits in the global spectrum of plant form and function. *Nature.* (2021) 597(7878):683–7. doi: 10.1038/s41586-021-03871-y

22. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med* (2021) 27 (10):1735–43. doi: 10.1038/s41591-021-01506-3

23. Stekhoven DJ, Bühlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics.* (2012) 28(1):112–8. doi: 10.1093/bioinformatics/btr597

24. Pommier JD, Gorman C, Crabol Y, Bleakley K, Sothy H, Santy K, et al. Childhood encephalitis in the greater Mekong region (the SouthEast Asia encephalitis project): A multicentre prospective study. *Lancet Glob Health* (2022) 10(7):e989–e1002. doi: 10.1016/S2214-109X(22)00174-7

25. Patel KV, Segar MW, Lavie CJ, Kondamudi N, Neeland IJ, Almandoz JP, et al. Diabetes status modifies the association between different measures of obesity and heart failure risk among older adults: A pooled analysis of community-based NHLBI cohorts. *Circulation.* (2022) 145(4):268–78. doi: 10.1161/CIRCULATIONAHA.121.055830

26. Kartal E, Schmidt TSB, Molina-Montes E, Rodríguez-Perales S, Wirbel J, Maistrenko OM, et al. A faecal microbiota signature with high specificity for pancreatic cancer. *Gut.* (2022) 71(7):1359–72. doi: 10.1136/gutjnl-2021-324755

27. Petrazzini BO, Chaudhary K, Márquez-Luna C, Forrest IS, Rocheleau G, Cho J, et al. Coronary risk estimation based on clinical data in electronic health records. *J Am Coll Cardiol* (2022) 79(12):1155–66. doi: 10.1016/j.jacc.2022.01.021

28. Auer TO, Álvarez-Ocaña R, Cruchet S, Benton R, Arguello JR. Copy number changes in co-expressed odorant receptor genes enable selection for sensory differences in drosophilid species [published online ahead of print, 2022 jul 21]. *Nat Ecol Evol* (2022) 6 (9):1343–53. doi: 10.1038/s41559-022-01830-y

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

30. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

31. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* (2010) 11(10):R106. doi: 10.1186/gb-2010-11-10-r106

32. Ru B, Tong Y, Zhang J. MR4Cancer: A web server prioritizing master regulators for cancer. *Bioinformatics.* (2019) 35(4):636–42. doi: 10.1093/bioinformatics/bty658

33. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118

34. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* (2013) 500(7463):415–21. doi: 10.1038/nature12477

35. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 12:R41. doi: 10.1186/gb-2011-12-4-r41

36. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity.* (2018) 48(4):812–830.e14. doi: 10.1016/j.immuni.2018.03.023

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

38. Harrell FE Jr. *rms: Regression Modeling Strategies*. R package version 51-4. 2019. https://CRAN.R-project.org/package=rms

39. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: From immunosurveillance to tumor escape. *Nat Immunol* (2002) 3(11):991–8. doi: 10.1038/ni1102-991

40. Watson RA, Tong O, Cooper R, Taylor CA, Sharma PK, de Los Aires AV, et al. Immune checkpoint blockade sensitivity and progression-free survival associates with baseline CD8+ T cell clone size and cytotoxicity. *Sci Immunol* (2021) 6(64):eabj8825. doi: 10.1126/sciimmunol.abj8825

41. Tian Q, Gao H, Zhao W, Zhou Y, Yang J. Development and validation of an immune gene set-based prognostic signature in cutaneous melanoma. *Future Oncol* (2021) 17(31):4115–29. doi: 10.2217/fon-2021-0104

42. Yang Y, Long X, Li G, Yu X, Liu Y, Li K, et al. Prediction of clinical prognosis in cutaneous melanoma using an immune-related gene pair signature. *Bioengineered.* (2021) 12(1):1803–12. doi: 10.1080/21655979.2021.1924556

43. Squarzanti DF, Zavattaro E, Pizzimenti S, Amoruso A, Savoia P, Azzimonti B. Non-melanoma skin cancer: news from microbiota research. *Crit Rev Microbiol* (2020) 46 (4):433–49. doi: 10.1080/1040841X.2020.1794792

44. Lutchminarian K, Clarke DL. The microbiology of ulcerative skin cancers: does the presence of pathogenic bacteria increase the risk of postoperative complications? *S Afr J Surg* (2021) 59(1):25a–e.

45. Woo YR, Cho SH, Lee JD, Kim HS. The human microbiota and skin cancer. *Int J Mol Sci* (2022) 23(3):1813. doi: 10.3390/ijms23031813

46. Gori A, Savarese I, D'Errico A, Grazzini M, Papi F, Maio V, et al. Estrogen receptor (ER)β expression and worse outcome from melanoma in pregnant and perimenopausal women. *J Am Acad Dermatol* (2016) 75(3):e117. doi: 10.1016/j.jaad.2016.02.1245

47. Dika E, Patrizi A, Lambertini M, Manuelpillai N, Fiorentino M, Altimari A, et al. Estrogen receptors and melanoma: A review. *Cells.* (2019) 8(11):1463. doi: 10.3390/cells8111463

48. Conforti F, Pala L, Bagnardi V, De Pas T, Martinetti M, Viale G, et al. Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol* (2018) 19(6):737–46. doi: 10.1016/S1470-2045(18)30261-4

49. Wang D, Dubois RN. Eicosanoids and cancer. *Nat Rev Cancer.* (2010) 10(3):181–93. doi: 10.1038/nrc2809

50. Yarla NS, Bishayee A, Sethi G, Reddanna P, Kalle AM, Dhananjaya BL, et al. Targeting arachidonic acid pathway by natural products for cancer prevention and therapy. *Semin Cancer Biol* (2016) 40-41:48–81. doi: 10.1016/j.semcancer.2016.02.001

51. Johnson AM, Kleczko EK, Nemenoff RA. Eicosanoids in cancer: New roles in immunoregulation. *Front Pharmacol* (2020) 11:595498. doi: 10.3389/fphar.2020.595498

52. Ando H, Ryu A, Hashimoto A, Oka M, Ichihashi M. Linoleic acid and alpha-linolenic acid lightens ultraviolet-induced hyperpigmentation of the skin. *Arch Dermatol Res* (1998) 290(7):375–81. doi: 10.1007/s004030050320

53. Bukowska J, Kopcewicz M, Walendzik K, Gawronska-Kozak B. Foxn1 in skin development, homeostasis and wound healing. *Int J Mol Sci* (2018) 19(7):1956. doi: 10.3390/ijms19071956

54. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science.* (2015) 348(6230):69–74. doi: 10.1126/science.aaa4971

55. Layer JP, Kronmüller MT, Quast T, van den Boorn-Konijnenberg D, Effern M, Hinze D, et al. Amplification of n-myc is associated with a T-cell-poor microenvironment in metastatic neuroblastoma restraining interferon pathway activity and chemokine expression. *Oncoimmunology.* (2017) 6(6):e1320626. doi: 10.1080/2162402X.2017.1320626

56. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, et al. Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol* (2017) 2017: PO.17.00073. doi: 10.1200/PO.17.00073

57. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* (2019) 30(1):44–56. doi: 10.1093/annonc/mdy495

58. Zhang Z, Lu M, Qin Y, Gao W, Tao L, Su W, et al. Neoantigen: A new breakthrough in tumor immunotherapy. *Front Immunol* (2021) 12:672356. doi: 10.3389/fimmu.2021.672356

59. Hollern DP, Xu N, Thennavan A, Glodowski C, Garcia-Recio S, Mott KR, et al. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell.* (2019) 179(5):1191–1206.e21. doi: 10.1016/j.cell.2019.10.028

60. Chang L, Chang M, Chang HM, Chang F. Microsatellite instability: A predictive biomarker for cancer immunotherapy. *Appl Immunohistochem Mol Morphol.* (2018) 26 (2):e15–21. doi: 10.1097/PAI.0000000000000575

61. Zhai L, Ladomersky E, Lenzen A, Nguyen B, Patel R, Lauing KL, et al. IDO1 in cancer: a Gemini of immune checkpoints. *Cell Mol Immunol* (2018) 15(5):447–57. doi: 10.1038/cmi.2017.143

62. Elia G, Ferrari SM, Galdiero MR, Ragusa F, Paparo SR, Ruffilli I, et al. New insight in endocrine-related adverse events associated to immune checkpoint blockade. *Best Pract Res Clin Endocrinol Metab* (2020) 34(1):101370. doi: 10.1016/j.beem.2019.101370

63. Monney L, Sabatos CA, Gaglia JL, Ryu A, Waldner H, Chernova T, et al. Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature.* (2002) 415(6871):536–41. doi: 10.1038/415536a

64. Wong CW, Fan YS, Chan TL, Chan AS, Ho LC, Ma TK, et al. Cancer genome project. BRAF and NRAS mutations are uncommon in melanomas arising in diverse internal organs. *J Clin Pathol* (2005) 58(6):640–4. doi: 10.1136/jcp.2004.022509

65. Kumar R, Njauw CN, Reddy BY, Ji Z, Rajadurai A, Klebanov N, et al. Growth suppression by dual BRAF(V600E) and NRAS(Q61) oncogene expression is mediated by SPRY4 in melanoma. *Oncogene.* (2019) 38(18):3504–20. doi: 10.1038/s41388-018-0632-2

66. Petti C, Molla A, Vegetti C, Ferrone S, Anichini A, Sensi M. Coexpression of NRASQ61R and BRAFV600E in human melanoma cells activates senescence and increases susceptibility to cell-mediated cytotoxicity. *Cancer Res* (2006) 66:6503–11. doi: 10.1158/0008-5472.CAN-05-4671

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fimmu.2023.1086342

OPEN ACCESS

EDITED BY
Hao Zhang,
Chongqing Medical University, China

REVIEWED BY
Jie Zhang,
Tongji University, China
Qianqian Song,
Wake Forest University, United States

*CORRESPONDENCE
Yifeng Cui
cui88963342@hrbmu.edu.cn
Zhaoyang Lu
lzy76772005@hrbmu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 01 November 2022
ACCEPTED 20 February 2023
PUBLISHED 03 March 2023

CITATION
Wen F, Meng F, Li X, Li Q, Liu J, Zhang R, Zhao Y, Zhang Y, Wang X, Ju S, Cui Y and Lu Z (2023) Characterization of prognostic value and immunological roles of RAB22A in hepatocellular carcinoma. *Front. Immunol.* 14:1086342. doi: 10.3389/fimmu.2023.1086342

Characterization of prognostic value and immunological roles of RAB22A in hepatocellular carcinoma

Fukai Wen[1,2†], Fanshuai Meng[1,2†], Xuewen Li[3†], Qingyu Li[1,2†], Jiaming Liu[1†], Rui Zhang[1,2], Yunzheng Zhao[1,2], Yu Zhang[1,2], Xin Wang[1,2], Shuai Ju[4], Yifeng Cui[1,2*] and Zhaoyang Lu[1,2*]

[1]Department of Hepatic Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Key Laboratory of Hepatosplenic Surgery, Ministry of Education, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [3]The Department of Inpatient Central Operating Room, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [4]Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Background: The protein-coding gene *RAB22A*, a member of the RAS oncogene family, is amplified or overexpressed in certain cancers. However, its action mechanism in hepatocellular carcinoma (HCC) remains unclear. Here, we aimed to examine the connection between *RAB22A* and survival prognosis in HCC and explore the biological significance of RAB22A.

Methods: A database-based pan-cancer expression analysis of RAB22A was performed. Kaplan–Meier analysis and Cox regression were performed to evaluate the association between *RAB22A* expression and survival prognosis in HCC. Using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA), various potential biological functions and regulatory pathways of RAB22A in HCC were discovered. Tumor immune infiltration was studied using the single sample gene set enrichment analysis (ssGSEA) method. N6-methyladenosine modifications and the regulatory network of competitive endogenous RNA (ceRNA) were verified in the TCGA cohort.

Results: *RAB22A* was upregulated in HCC samples and cell lines. A high *RAB22A* expression in HCC was strongly correlated with sex, race, age, weight, TNM stage, pathological stage, tumor status, histologic grade, TP53 mutation status, and alpha fetal protein (AFP) levels. Overexpression of *RAB22A* indicated a poor prognosis was related to overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI). GO and KEGG analyses revealed that the differentially expressed genes related to RAB22A might be involved in the proteasomal protein catabolic process, ncRNA processing, ribosome ribosomal subunit, protein serine/threonine kinase activity, protein serine kinase activity, Endocytosis, and non-alcoholic fatty liver disease. GSEA analyses revealed that the differentially expressed genes related to RAB22A might be involved in the T cell receptor, a co-translational protein, that binds to the membrane, axon guidance, ribosome, phagocytosis, and Eukaryotic translation initiation. *RAB22A* was correlated with N6-methyladenosine expression in HCC and

established *RAB22A*-related ceRNA regulatory networks. Finally,*RAB22A* expression was positively connected the levels of infiltrating with T helper cells, Tcm cells, and Th2 cells,In contrast, we observed negatively correlations with cytotoxic cells, DCs, and pDCs cells.Moreover,*RAB22A* expression showed a strong correlation with various immunomarkergroups in HCC.

Conclusions: RAB22A is a potential therapeutic target for improving HCC prognosis and is closely related to immune cell infiltration.

KEYWORDS

RAB22A, hepatocellular carcinoma, cancer immune infiltrates, prognosis, biomarker, bioinformatics analysis

1 Introduction

Hepatocellular carcinoma (HCC) is the sixth most diagnosed cancer and the fourth leading cause of cancer death worldwide, with approximately 841,000 new cases and 782,000 deaths annually (1). Many key factors, including infection with hepatitis B or C and contact with foods contaminated with aflatoxin, contribute to HCC development (2). Surgery is the typical treatment for HCC; however, the disease is prone to relapse and metastasis, making it difficult to cure (3). Therefore, there is an urgent need to identify new relevant biomarkers to improve the early diagnosis, prognostic assessment, and treatment of HCC.

RAB22A is a small GTPase that belongs to the RAB protein family, specifically, the RAB5 subfamily (4). This protein is mainly located in early endosomes, Golgi bodies, and late endosomes. RAB proteins are involved in the regulation of vesicular traffic and exosome formation (5). Studies have found that the RAB5 subfamily (including RAB5, RAB21, RAB22A, and RAB22B) is primarily involved in the endocytosis, transport, and metabolism of growth factor receptors and may thus be associated with cancer progression (6–8). *RAB22A* expression is elevated in several malignancies, including breast, colorectal, and osteosarcoma cancer (9–11). It accelerates the progression of malignant tumors *via* various mechanisms, for instance, miRNA downregulation (11), recycling of extracellular matrix metalloproteinase inducer (EMMPRIN) (12), and hypoxia-inducible factor (13). Nevertheless, the function of *RAB22A* in HCC remains unclear.

Furthermore, RAB22A has multiple immune functions and is a novel immunomodulatory factor. Accurate intracellular transport of MHC-I molecules in dendritic cells (DCs) and T lymphocytes depends on RAB22A function (14). RAB22A is also part of the accommodative immune response and is absorbed by a process that separates it from the envelope proteins and spreads it throughout the body (15). Previous research has identified RAB22A as the main endosomal target in pathogen infection and a critical regulator of microbial infection and intracellular transport (16). In summary, *RAB22A* may have a significant prognostic and immunological significance in HCC.

In the current study, we analyzed the expression of *RAB22A* in HCC and paracancerous tissues using multiple datasets and *in vitro* experiments. Additionally, we examined the connection between *RAB22A* and survival prognosis in HCC and explored the biological significance of RAB22A by performing enrichment and protein-protein interaction (PPI) network analyses and determining the correlation with immune cell infiltration. Furthermore, we constructed ceRNA regulatory networks involving RAB22A in HCC. Our study proposes a possible connection between *RAB22A* expression and the presence of immune infiltrates in HCC.

2 Materials and methods

2.1 Database source and processing

Gene expression and clinical data were extracted from multiple databases (Supplementary Table 1) and RAB22A expression levels from RNA-seq data (TPM) of patients with HCC were analyzed. The Supplementary Materials and Methods (17–19) presents detailed information on the included data.

2.2 Patients and clinical samples

The First Affiliated Hospital of Harbin Medical University provided 30 matched sets of HCC and nearby non-tumor liver samples from patients undergoing hepatectomy between February 2020 and June 2022. This project was approved by the First Affiliated Hospital of Harbin Medical University's Ethics Committee.

2.3 Western blotting and quantitative real-time PCR

Total proteins and total RNA were extracted from HCC samples. Details of the experimental procedures are provided in the Supplementary Materials and Methods.

2.4 Tumor immune infiltration analysis

We used the single sample gene set enrichment analysis (ssGSEA) method (20, 21) and TIMER database (22, 23) to investigate the relationships between *RAB22A* expression and immune cell infiltration, as detailed in the Supplementary Materials and Methods.

2.5 Gene set enrichment analysis

Enrichment analyses of relevant functional pathways were performed using the GO and KEGG databases (Supplementary Tables 3, 4) and GSEA (Supplementary Tables 5, 6), as detailed in the Supplementary Materials and Methods (24, 25).

2.6 Prediction and construction of ceRNA networks

Multiple databases were used to predict and screen the lncRNA-miRNA-mRNA (RAB22A) ceRNA network online. Details are provided in the Supplementary Materials and Methods.

2.7 Protein interaction network and module analysis

We created the protein–protein interaction (PPI) network using the Search Tool for the Retrieval of Interacting Genes (STRING) database (Supplementary Table 6) (26, 27), as detailed in the Supplementary Materials and Methods.

2.8 Statistical analysis

The R package (version 3.6.3) was used for statistical analyses and plotting. RAB22A expression in unpaired and paired samples was analyzed using the Wilcoxon rank sum test, and Wilcoxon signed rank test, respectively, with the pROC (1.17.0.1) package for ROC analysis. The *RAB22A* expression level was analyzed by querying the GEO, TIMER, and UALCAN databases (18). Using the KM method and log-rank test, we compared the differences in 10-year OS, DSS, and PFI between patients with high RAB22A expression and those with low RAB22A expression in TCGA. Cox analysis was used to determine the correlation between *RAB22A* expression and clinical features. $p < 0.05$ was considered to indicate significance.

3 Results

3.1 *RAB22A* is upregulated in HCC

First, we examined the *RAB22A* expression levels in different malignancies by assessing TCGA databases. RAB22A was highly expressed in 33 malignant tumors, including HCC (Figure 1A). In addition, *RAB22A* was highly expressed in the GEO datasets GSE121248, GSE87630, GSE76427, GSE84005, GSE57957, and GSE39791 HCC samples ($p < 0.001$) (Figures 1B–G). Western blot analysis of human normal liver cells (L02) and HCC cells (Hep G2, SK-Hep1, Huh7, HCCLM3, and MHCC97-H) validated the high expression of *RAB22A* in HCC cell lines (Figure 1H). The same results were obtained through qRT-PCR ($p < 0.001$) (Figure 1K). Next, we extracted 30 pairs of proteins from HCC and adjacent tissues and analyzed them using western blotting, which revealed that *RAB22A* was highly expressed in the former (Figure 1I). Results of western blot analysis of the liver and adjacent tissues are shown in Supplementary Figure 1. The high *RAB22A* mRNA expression levels in HCC tissues were further substantiated using qRT-PCR ($p < 0.001$) (Figure 1L). Immunohistochemistry (IHC) results also verified that *RAB22A* was upregulated in HCC tissues (Figure 1J). Finally, a receiver operating characteristic (ROC) curve was created. The ROC curve enclosed by the axes is the area below the curve (AUC). The AUC for *RAB22A* was 0.891, suggesting its remarkable diagnostic value for HCC (Figure 1M).

3.2 Association of RAB22A expression with clinical characteristics

Using the UALCAN database to perform subgroup analysis of numerous pathological characteristics, we found that *RAB22A* transcript levels were elevated in patients with HCC. (Figure 2A). The sub-group analysis of cancer stage, ethnicity, sex, age, weight, tumor grade, and TP53 mutation showed that the expression of RAB22A in HCC patients was significantly higher than that in the normal group (Figures 2B–H).

Logistic regression analysis showed that the increased expression of RAB22A in HCC was significantly correlated with sex (OR = 0.627 for male vs. female, $p = 0.036$), weight (OR = 0.567 for weight > 70kg vs. ≤ 70kg, $p = 0.009$), histological grades (OR=1.611 for G3 and G4 vs G1 and G2, $p = 0.028$), and tumor status (OR = 1.619 for with tumors vs. tumor free, $p = 0.026$). Conversely, RAB22A expression was not associated with age, M stage, T stage, N stage, height, BMI, AFP, or vascular invasion (Table 1).

Next, we collected data from TCGA database to determine the clinicopathological parameters of *RAB22A* in different patients with HCC. Detailed information on the clinical data is provided in Table 2. After excluding cases without the necessary clinical data, 374 cases with a median age of 61.5 (range: 49.25–70.00) years and male preponderance of 67% were included. High expression of RAB22A in HCC was positively associated with tumor status (tumor-free vs. with tumor, $p = 0.033$), sex (female vs. male, $p = 0.047$), weight (≤ 70 vs. > 70, $p = 0.012$), and histological grade (grades 3 and 4 vs. grades 1 and 2, $p = 0.031$). These results indicate that the overexpression of RAB22A in HCC is closely related to the clinicopathological characteristics.

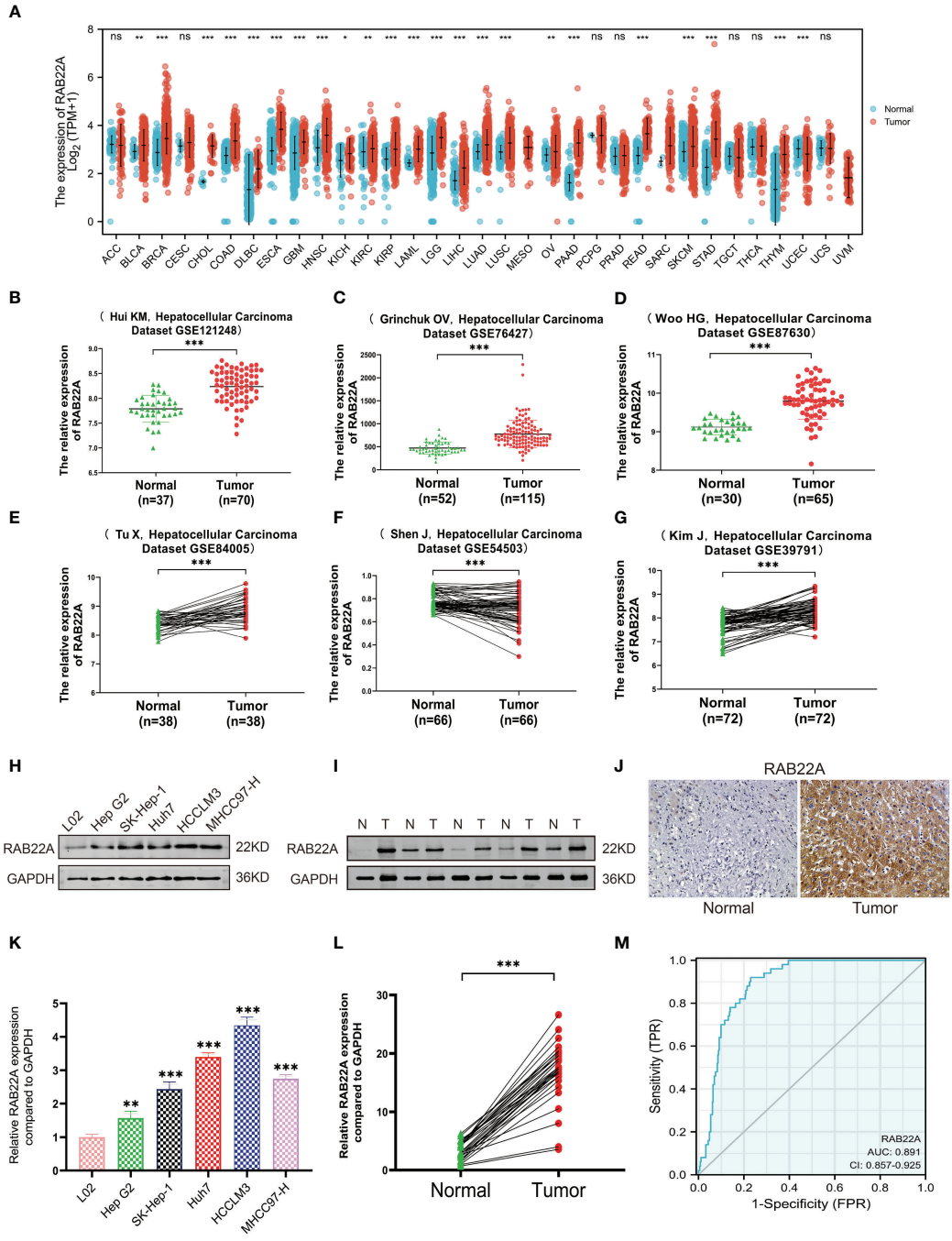


FIGURE 1
Expression level of *RAB22A* in HCC was verified in TCGA and GEO databases and *in vitro* experiments. **(A)** Comparison of the expression levels of *RAB22A* in different cancerous and normal tissues. **(B–G)** CEO database analysis of *RAB22A* expression in HCC tissues. **(H)** Western blotting assay of RAB22A protein expression levels in L02, Hep G2, SK-Hep1, Huh7, HCCLM3, and MHCC97-H cell lines. **(I)** Western blotting assay of *RAB22A* protein expression levels in HCC and adjacent tissues. **(J)** *RAB22A* protein levels in normal liver and HCC were measured using IHC. **(K)** qRT-PCR assay of RAB22A mRNA expression levels in L02, Hep G2, SK-Hep1, Huh7, HCCLM3, and HCCH97-H cell lines. **(L)** qRT-PCR assay of *RAB22A* mRNA expression levels in 30 pairs of HCC and adjacent tissues. **(M)** ROC curves were created to investigate the value of *RAB22A* in identifying HCC tissues. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, NS, no significance.

3.3 Prognostic value of *RAB22A* in HCC

Kaplan–Meier survival curves were analyzed to determine the connection between *RAB22A* expression and overall survival (OS), disease-free survival (DSS), and progression-free interval (PFI) in the prognosis of patients with HCC. Increased levels of *RAB22A* expression were inversely related to prognosis (Figures 3A–C). Additionally, subgroup analysis was performed on patients with low RAB22A expression and AFP < 400, and these patients had better OS, DSS, and PFI prognosis (Figures 3D–F). However, the groups with AFP (ng/mL) > 400 showed no significant differences (Supplementary Figures 2A–C). The high expression of RAB22A in

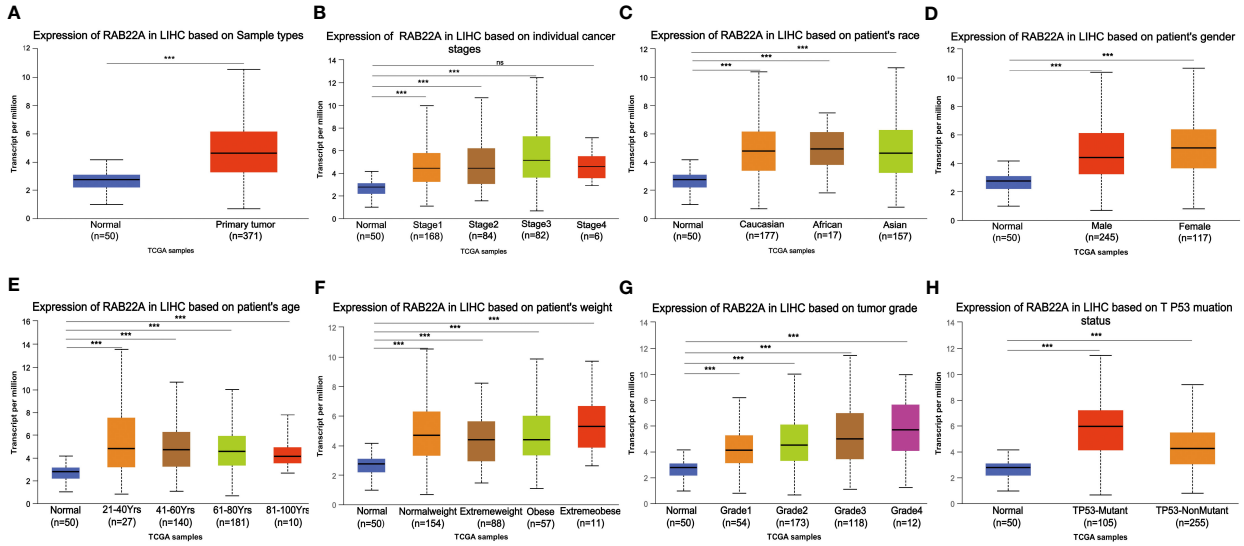


FIGURE 2
Box plot showing the relative transcription of RAB22A in individual cancer stages, race, gender, age, weight, tumor grade, and TP53 mutation status in a subgroup of patients with HCC. **(A)** *RAB22A* in normal and HCC tissues. **(B)** *RAB22A* in normal individuals or in patients with stages 1–4 liver cancer. **(C)** *RAB22A* in normal and LIHC samples based on patient ethnicity. **(D)** *RAB22A* in normal individuals and males and females with HCC. **(E)** *RAB22A* in healthy subjects of any age and patients aged 21–40, 41–60, 61–80, and 81–100 years with HCC. **(F)** *RAB22A* in healthy subjects of any weight and normal weight patients, extreme weight patients, obese patients, and extremely obese patients. **(G)** *RAB22A* in normal subjects and patients with different liver cancer tumor grades. **(H)** *RAB22A* in normal and TP53-mutant or TP53-non mutant patients. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, NS, no significance.

stage M0 liver cancer was associated with poor OS, DSS, and PFI in a subgroup of patients (Figures 3G–I). The subgroups of T3 versus T4, stages III vs. IV, and tumor versus tumor-free status had significantly worse OS (Supplementary Figures 1D–F). Finally, we compared predictive variables in patients with HCC obtained by univariate regression analysis to those obtained *via* multivariate survival analysis (OS) (Supplementary Table 7). Pathologic stage (stages I and II compared with stages III and IV; p <0.001), tumor size (T stages 1 and 2 versus T stages 3 and 4; $p < 0.001$), metastatic spread (M stages 0 and 1; $p = 0.017$), and tumor status (without or with tumor; $p < 0.001$) were highly significant in the univariate analysis. The multivariate analysis showed that with tumor ($p = 0.014$) was significant, suggesting that it is an independent risk factor.

TABLE 1 Association between *RAB22A* expression and clinicopathologic parameters by Logistic regression.

Characteristics	Total (N)	Odds Ratio (OR)	P value
Age (>60 vs. <=60)	373	0.851 (0.566-1.279)	0.438
M stage (M1 vs. M0)	272	2.912 (0.368-59.271)	0.357
Gender (Male vs. Female)	**374**	**0.627 (0.403-0.969)**	**0.036**
T stage (T3&T4 vs. T1&T2)	371	1.510 (0.942-2.437)	0.089
N stage (N1 vs. N0)	258	2.953 (0.373-60.136)	0.351
Weight (>70 vs. <=70)	**346**	**0.567 (0.369-0.867)**	**0.009**
Height (>=170 vs. < 170)	341	0.748 (0.484-1.152)	0.189
BMI (>25 vs. <=25)	337	0.758 (0.493-1.163)	0.205
AFP(ng/ml) (>400 vs. <=400)	280	1.608 (0.921-2.831)	0.096
Vascular invasion (No vs. Yes)	318	0.960 (0.604-1.526)	0.863
Histologic grade (G3 & G4 vs. G1 & G2)	**369**	**1.611 (1.053-2.475)**	**0.028**
Tumor status (With tumor vs. Tumor free)	**355**	**1.619 (1.061-2.478)**	**0.026**

The bold values indicates that the correlation analysis between RAB22A and clinicopathological parameters are statistically significant.

10.3389/fimmu.2023.1086342

TABLE 2 Correlation between clinicopathological variables and *RAB22A* expression.

Characteristic	Low expression of *RAB22A*	High expression of *RAB22A*	P value
n	187	187	
T stage, n (%)			0.166
T1	96 (25.9%)	87 (23.5%)	
T2	49 (13.2%)	46 (12.4%)	
T3	36 (9.7%)	44 (11.9%)	
T4	3 (0.8%)	10 (2.7%)	
N stage, n (%)			0.622
N0	126 (48.8%)	128 (49.6%)	
N1	1 (0.4%)	3 (1.2%)	
M stage, n (%)			0.623
M0	132 (48.5%)	136 (50%)	
M1	1 (0.4%)	3 (1.1%)	
Pathologic stage, n (%)			0.293
Stage I	93 (26.6%)	80 (22.9%)	
Stage II	47 (13.4%)	40 (11.4%)	
Stage III	36 (10.3%)	49 (14%)	
Stage IV	2 (0.6%)	3 (0.9%)	
Tumor status, n (%)			**0.033**
Tumor free	110 (31%)	92 (25.9%)	
With tumor	65 (18.3%)	88 (24.8%)	
Gender, n (%)			**0.047**
Female	51 (13.6%)	70 (18.7%)	
Male	136 (36.4%)	117 (31.3%)	
Race, n (%)			0.940
Asian	79 (21.8%)	81 (22.4%)	
Black or African American	8 (2.2%)	9 (2.5%)	
White	88 (24.3%)	97 (26.8%)	
Age, n (%)			0.502
<=60	85 (22.8%)	92 (24.7%)	
>60	102 (27.3%)	94 (25.2%)	
Weight, n (%)			**0.012**
<=70	82 (23.7%)	102 (29.5%)	
>70	95 (27.5%)	67 (19.4%)	
Height, n (%)			0.228
< 170	96 (28.2%)	105 (30.8%)	
>=170	77 (22.6%)	63 (18.5%)	
BMI, n (%)			0.246
<=25	84 (24.9%)	93 (27.6%)	
>25	87 (25.8%)	73 (21.7%)	

(Continued)

10.3389/fimmu.2023.1086342

TABLE 2 Continued

Characteristic	Low expression of *RAB22A*	High expression of *RAB22A*	P value
Residual tumor, n (%)			0.217
R0	170 (49.3%)	157 (45.5%)	
R1	6 (1.7%)	11 (3.2%)	
R2	1 (0.3%)	0 (0%)	
Histologic grade, n (%)			**0.031**
G1	35 (9.5%)	20 (5.4%)	
G2	92 (24.9%)	86 (23.3%)	
G3	55 (14.9%)	69 (18.7%)	
G4	3 (0.8%)	9 (2.4%)	
AFP(ng/ml), n (%)			0.126
<=400	118 (42.1%)	97 (34.6%)	
>400	28 (10%)	37 (13.2%)	

The bold values indicates that the correlation analysis between RAB22A and clinicopathological parameters are statistically significant.

3.4 GSEA and GO/KEGG enrichment analyses

GO and KEGG pathway co-expression analyses of RAB22A-related genes in liver cancer mRNA sequencing data with 371 patients from the TCGA were performed using the functional module of Linkedomics. The top 50 marker genes and their connections with *RAB22A* expression are displayed on the heat map (Figures 4A, B; Supplementary Table 8). These findings revealed a widespread effect of *RAB22A* on the transcriptome.

Next, we conducted an enrichment analysis using the GO and KEGG databases to support the concept that *RAB22A*-related DEGs play a biological role in HCC (Figures 4C, D). The results of GO analysis showed that these DEGs were related to biological processes (BP), cellular components (CC), and molecular functions (MF). In the GO analysis, DEGs were enriched in diverse biological pathways, including proteasomal protein catabolic process, ncRNA processing, ribosome ribosomal subunit, protein serine/threonine kinase activity, and protein serine kinase activity. In the KEGG analysis, DEGs were highly concentrated in endocytosis and non-alcoholic fatty liver disease. GSEA was used to analyze the biological functions related to *RAB22A* expression.

The later criteria were enrichment score | NSE | > 1 ($p < 0.05$), according to which the five most positively relevant signal pathways were selected. GO analysis revealed that *RAB22A* expression was strongly positively correlated with the processes of homophilic cell adhesion *via* plasma membrane adhesion, immunoglobulin, T cell receptor, and plasma membrane signaling receptor complex (Figure 4E). The expression of *RAB22A* was inversely linked to that of co-translational proteins that bind to the membrane, the cytosolic ribosome, the structural components of the ribosome, the ribosomal subunit, and the nonsense-mediated decay of nuclear-transcribed mRNA catabolic processes (Figure 4H). KEGG analysis revealed that *RAB22A* expression was most strongly negatively connected with axon guidance, extracellular matrix receptor interaction, focal adhesion, FCγR-mediated phagocytosis, and the interaction with neuroactive ligand receptors (Figure 4F). The ribosome, Parkinson's disease, retinol metabolism, oxidative phosphorylation, and complement and coagulation cascades were the top five most negatively correlated pathways (Figure 4I). REACTOME pathway analysis determined that phospholipids play a role in phagocytosis, Fc gamma receptor FCGR-dependent phagocytosis, CD22-mediated B cell receptor (BCR) regulation, and FCGR activation, and that second messenger is activated by BCR antigen that all positively correlated with RAB22A expression (Figure 4G). Eukaryotic translation initiation, eukaryotic translation elongation, translocation, the response of eukaryotic initiation factor 2 alpha subunit kappa B cyclin N2 to amino acid deprivation, co-translational protein of SRP-dependent targeting to the membrane, and nonsense-mediated decay were all negatively correlated with *RAB22A* expression (Figure 4J).

3.5 PPI network analysis

The PPI network of co-expressed genes conforming to the STRING conditions was assembled and visualized using Cytoscape, and analysis of the interactions among 108 DEGs in the HCC group was conducted. A total of 51 proteins and 534 edges were screened (Figure 5A; Supplementary Table 9).

After screening 12 nodes and 212 edges, a primary gene cluster with a total score ≥ 14,000 was discovered (Figure 5B). Finally, we screened the top 10 central genes, namely *RAB22A*, *RABGEF1, VPS45, VPS18, VPS11, MON1A, VPS39, VPS16, ZFYV20*, and *VPS8* (Figure 5C).

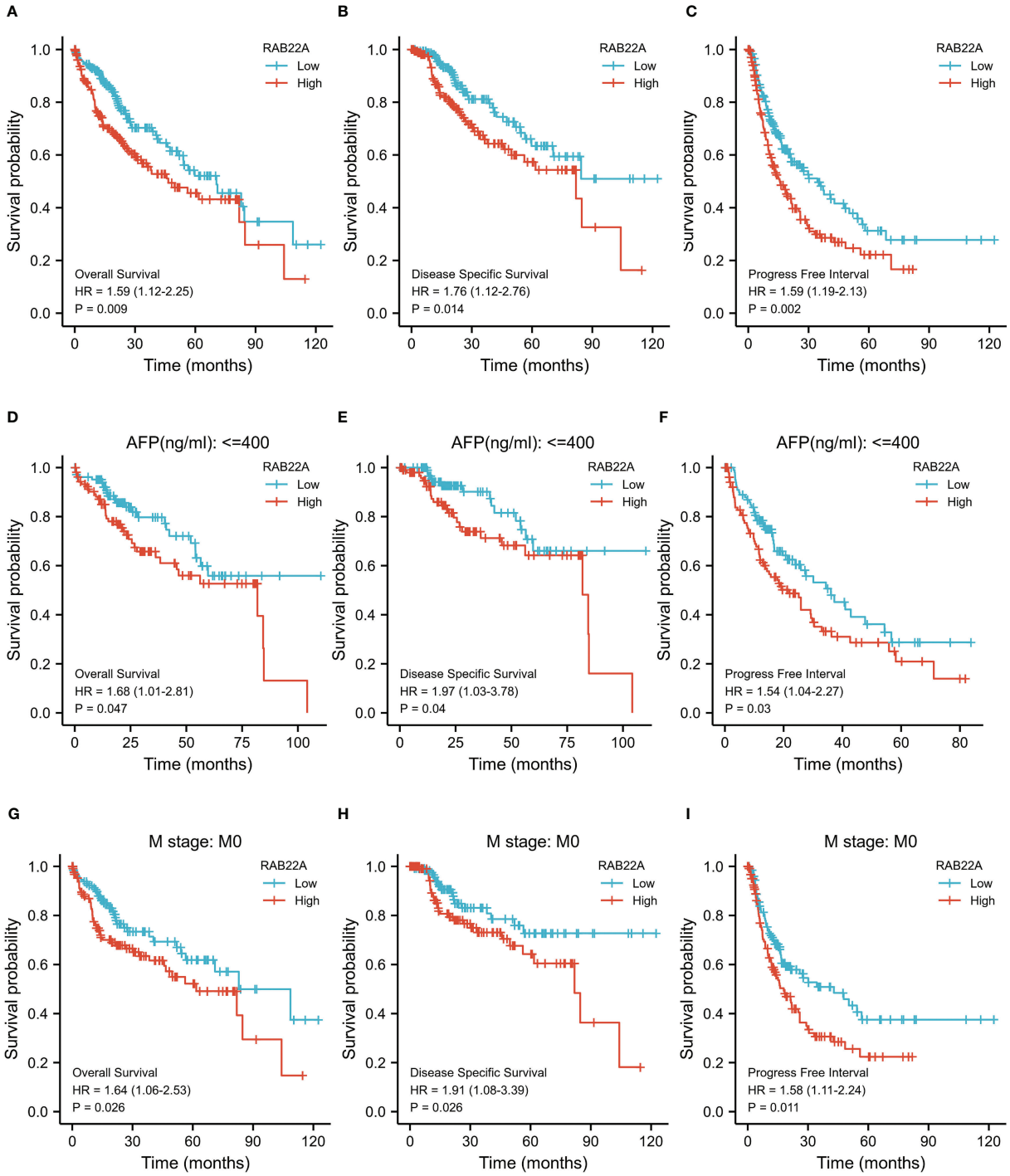


FIGURE 3
Kaplan–Meier survival plots comparing the relationship between *RAB22A* and prognosis in HCC. **(A–C)** Survival curves of OS, DSS, and PFI between RAB22A-high and -low patients with HCC. **(D–F)** OS, DSS, and PFI survival curves of patients with HCC with high and low *RAB22A* expression of AFP (ng/mL) ≤ 400. **(G–I)** Survival curves comparing OS, DSS, and PFI in patients with HCC at the M0 stage with high and low expression of *RAB22A*.

3.6 Role of RAB22A and m6A methylation regulators in HCC

M6A methylation affects the development of HCC (24–27). The expression of *RAB22A* was compared with that of the 23 M6A methylation genes reported in the literature to verify this conclusion (Figure 6A). *RAB22A* expression was closely connected with that of the 23 m6A-related genes in HCC (Figures 6B–X). Moreover, groups were formed according to *RAB22A* median expression. By analyzing the differences in the 23 m6A methylation genes in *RAB22A* between the high- and low-expression groups of patients with HCC, we observed that the expression levels of all genes in the *RAB22A* high-expression group were upregulated (Figure 6Y). Overall, we observed an obvious relationship between m6A methylation and RAB22A expression levels in HCC.

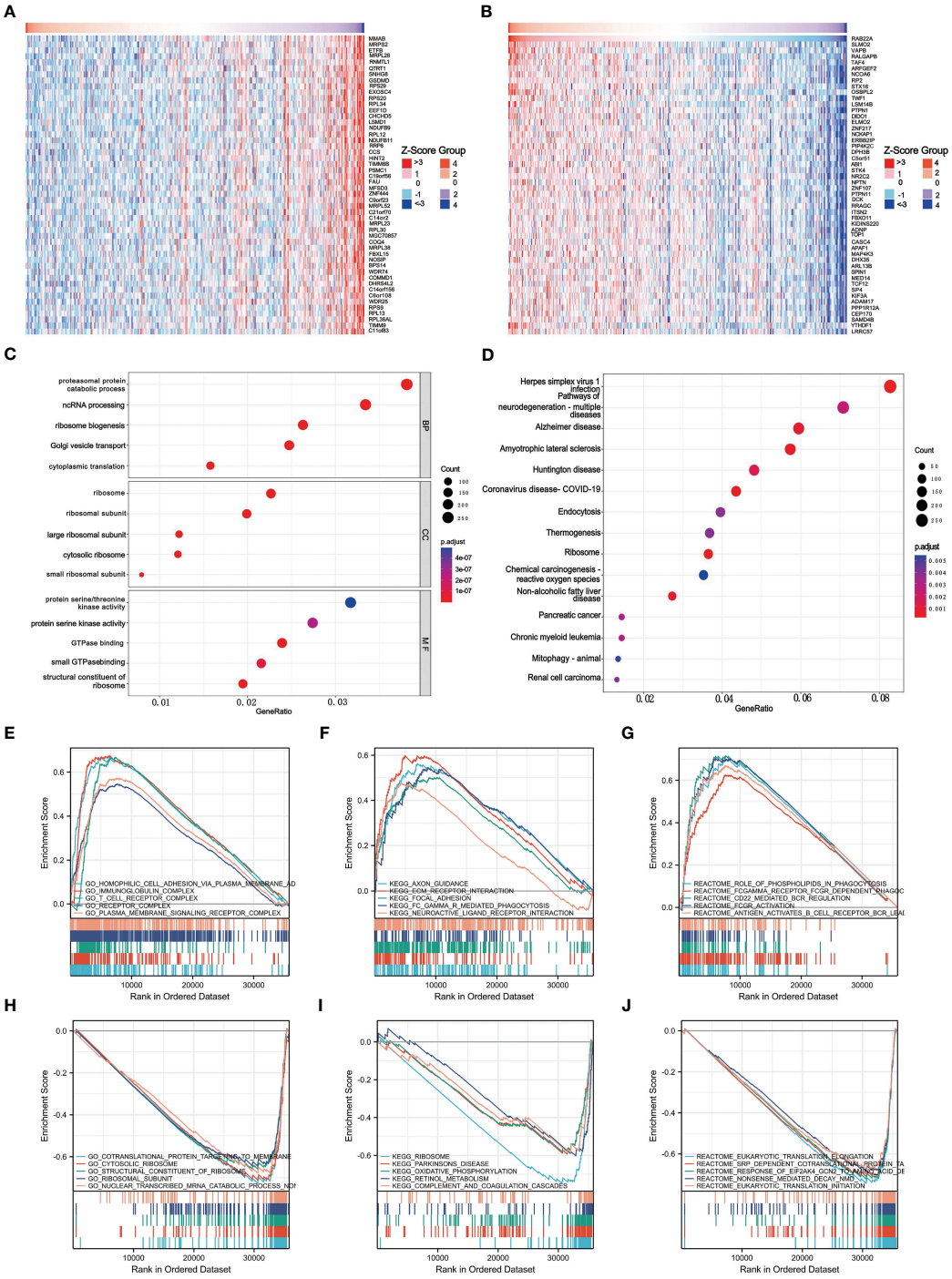


FIGURE 4
Enrichment of biofunction and associated gene analysis of *RAB22A* in HCC. **(A, B)** Heat map showing genes positively and negatively associated with *RAB22A* in liver cancer (top 50). Positively associated genes are indicated in red, while negatively associated genes are in green. **(C)** The enriched terms in GO categories in HCC. **(D)** KEGG pathway analysis based on *RAB22A*-associated DEGs. **(E)** The five most positively correlated pathways were revealed by GO term analysis. **(F)** KEGG pathway analysis revealed the five most positively correlated pathways. **(G)** The five most positively correlated pathways were identified *via* REACTOME pathway analysis. **(H)** The five most negatively correlated pathways were identified *via* GO term analysis. **(I)** KEGG pathway analysis identified the five most negatively correlated pathways. **(J)** The five most negatively correlated pathways were identified *via* REACTOME pathway analysis.

10.3389/fimmu.2023.1086342

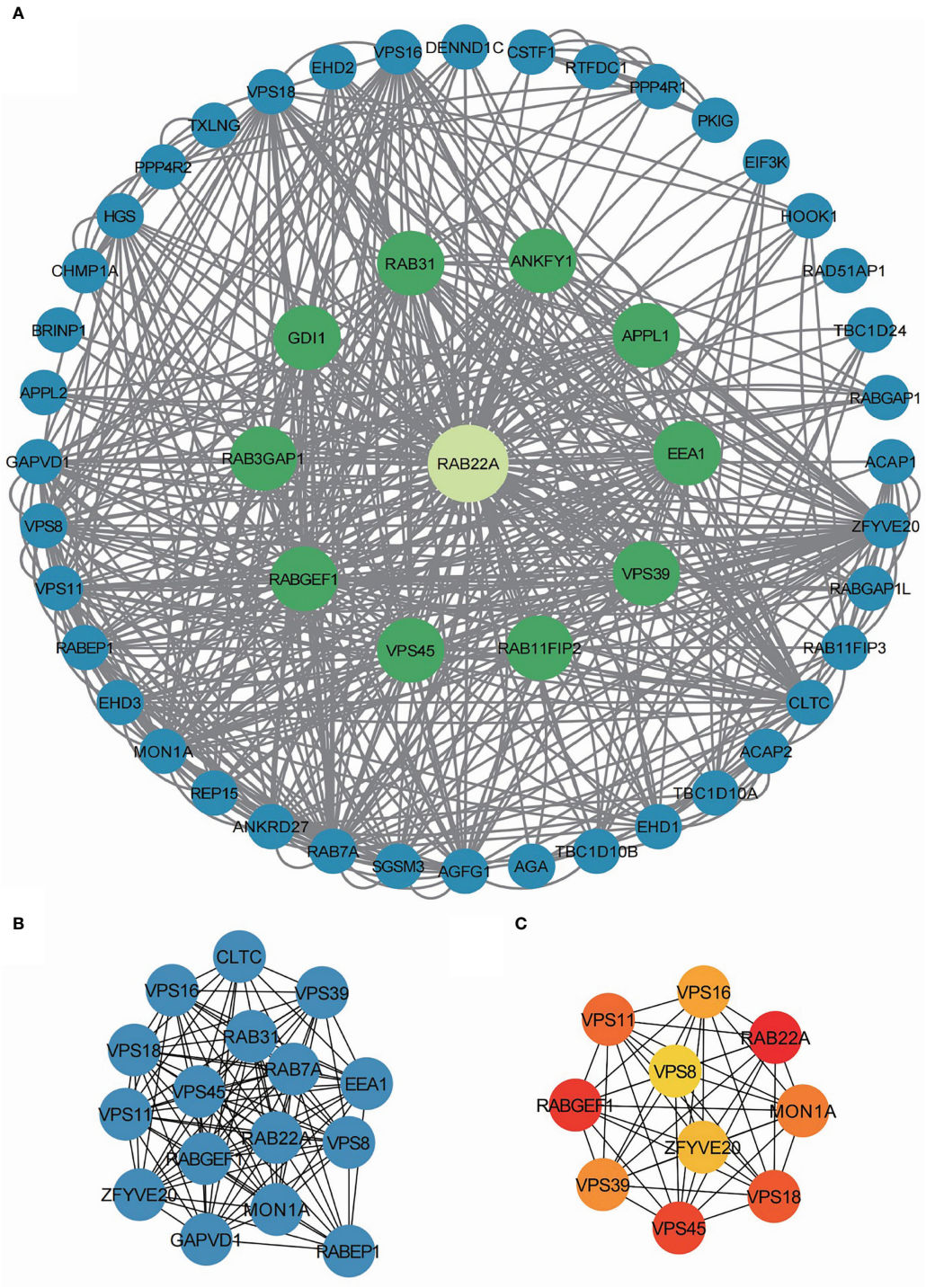


FIGURE 5
PPI network enrichment analysis. **(A)** The PPI network was built based on PPI pairs identified by the STRING dataset. **(B)** Hub gene clusters were selected from the PPI network (criteria of total scores ≥ 14,000). **(C)** Top 10 hub genes in the PPI network.

3.7 Construction of a triple regulatory network for *RAB22A*-associated ceRNA

Increasing evidence has demonstrated the regulatory effect on the lncRNA-miRNA-mRNA ceRNA network in HCC. The Venn diagram showed 41 overlapping miRNAs in the Targerscan, starBase, and MiRDB databases (Figure 7A). Five human-derived miRNAs (miR-328-3p, miR-3163, miR-2114-5p, miR-664b-3p, and miR-204-5p) were verified to negatively correlate with *RAB22A* expression (Figure 7B). The expression of *RAB22A* and target microRNAs is displayed as a scatter plot (Figures 7C–G). We consulted the Rnalnter and starBase databases to predict lncRNAs that can have a mutual effect on target miRNAs (miR-204-5p and miR-328-3p) (Figures 7H–I). The expression levels of lncRNAs and

10.3389/fimmu.2023.1086342

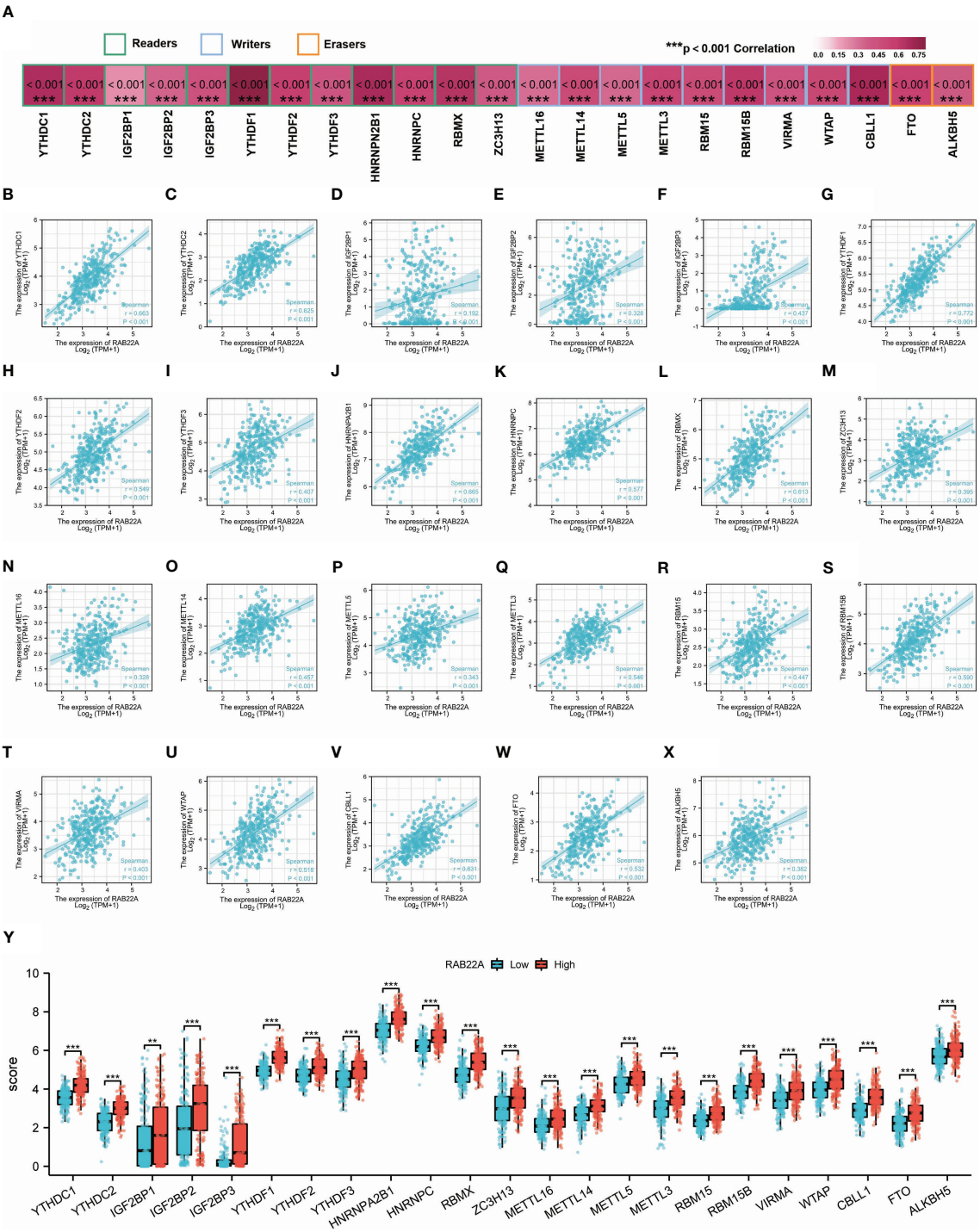


FIGURE 6
Correlation analysis of *RAB22A* expression levels with m6A-related gene expression in HCC tissues. **(A)** Correlation of *RAB22A* expression levels with m6A gene expression in HCC. **(B–X)** Scatter plot showing the relationship between *RAB22A* and the m6A gene. Differences in 23 M6A-related genes between the *RAB22A* high-expression group and *RAB22A* low-expression group in liver cancer patients **(Y)**.**$p < 0.01$, ***$p < 0.001$, NS, no significance.

miRNAs were inversely correlated, which accounted for the mutual influence between the two. We used the starBase database to filter and identify lncRNAs that were adversely associated with the two target miRNAs in HCC. Nine HCC-related ceRNA regulatory networks were constructed (Figure 7J).

3.8 Association of RAB22A expression with immune cell infiltration

Using the ssGSEA method, we verified the strong connection between *RAB22A* and immune cells (Figure 8A). The expression of

10.3389/fimmu.2023.1086342

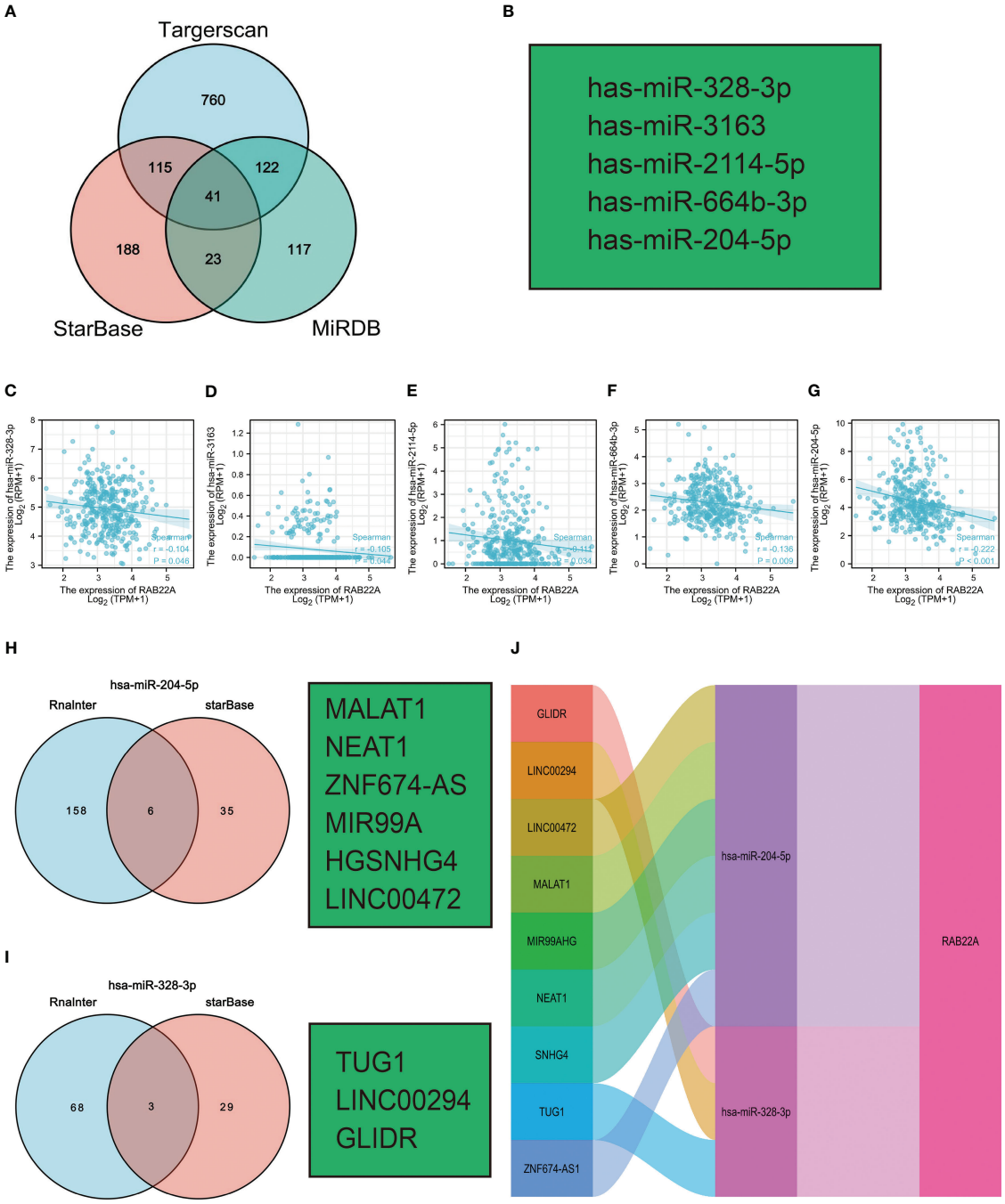


FIGURE 7
Prediction of ceRNA networks in HCC. **(A)** Venn diagram results showing 41 overlapping miRNAs in Targerscan, starBase, and MiRDB databases. **(B)** Five miRNAs screened for negative correlation with *RAB22A* expression. **(C–G)** Scatter plots showed that miRNAs were significantly correlated with mRNAs. **(H, I)** Prediction of lncRNAs bound to target miRNAs using miRNet and starBase online databases and displayed as a Venn diagram, including hsa-miR-204-5p and hsa-miR-328-3p. **(J)** Sankey diagram showing the *RAB22A*-related ceRNA regulatory network.

RAB22A was positively connected with T helper cells, Tcm cells, and Th2 cells ($p < 0.001$) but negatively with cytotoxic cells, DCs, and pDCs ($p < 0.001$) (Figures 8B–G). RAB22A may be heavily involved in the T-cell immune response to HCC. Moreover, *RAB22A* expression in HCC correlated with various immune cell markers (Table 3). In the M2 macrophages in HCC, we found that *RAB22A* expression was substantially relevant to the expression of the immunological markers CD163, VSIG4, and MS4A4A. These results indicate that *RAB22A* caused the macrophages in HCC to adopt an M2 phenotype. The expression of *RAB22A* was substantially linked to 66 immunological markers, including CD8A, CD3D, and T-bet, in an analysis of functional T-cell immunity indicators. Furthermore, *RAB22A* expression was linked to immunological markers for B cells, T cells, TAMs, and neutrophils (Table 3). The TIMER database was utilized to determine whether *RAB22A* expression in HCC was connected

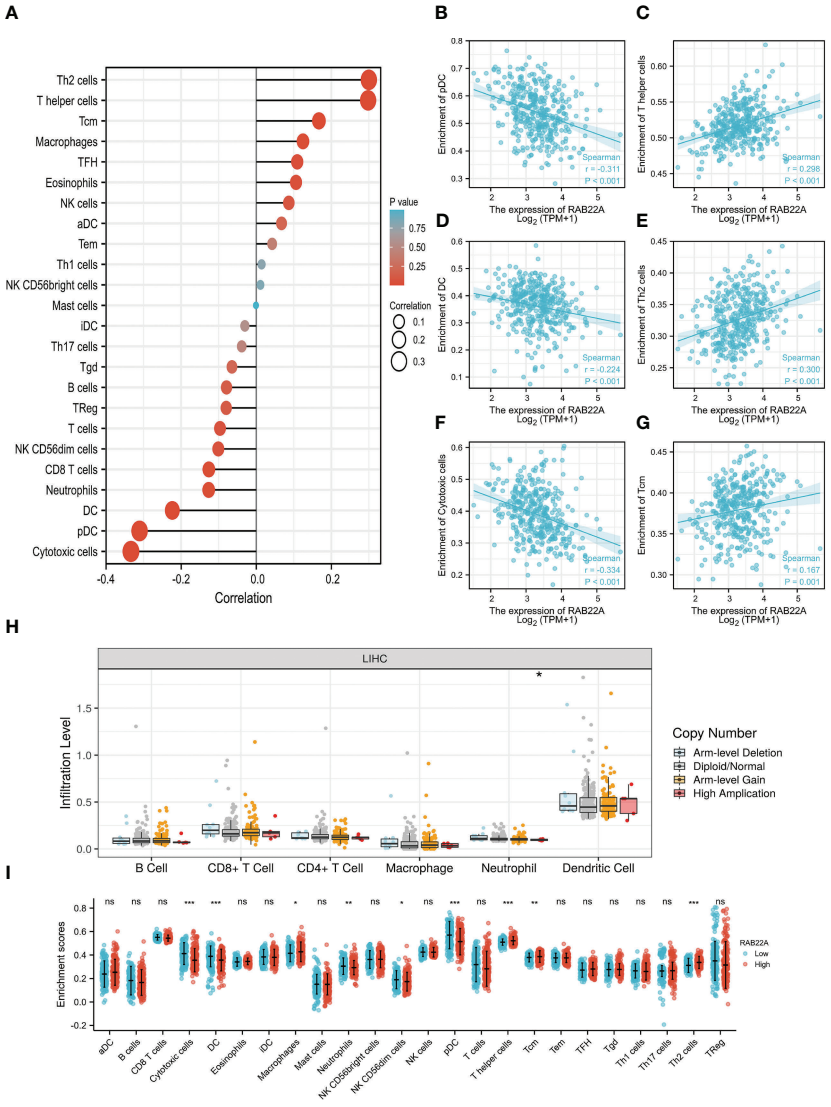


FIGURE 8
Relationship between the expression of RAB22A and microenvironment of immune infiltrating cells in HCC. **(A)** Forest plot depicting the relationship between *RAB22A* expression levels and the relative abundance of the 24 immune cells. **(B–G)** Scatter plots showing the degree of differentiation of pDCs, T helper cells, DCs, Th2 cells, cytotoxic cells, and Tcm cells between the high and low *RAB22A* expression groups. **(H)** SCNA showed that the expression of *RAB22A* correlated with the degree of immune cell infiltration. **(I)** Scatter plot showing the correlation of 24 immune cells with *RAB22A* expression levels. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, NS, no significance.

with immune cell invasion levels. The results indicated that the CNV of *RAB22A* was related to the level of neutrophil infiltration. (Figure 8H). Subsequently, the infiltration of macrophages, T helper cells, Tcm, and Th2 cells increased ($p < 0.001$) in the *RAB22A* high-expression group; however, cytotoxic cells, DCs, and pDCs decreased ($p < 0.001$) (Figure 8I). These results verified that the increased expression of *RAB22A* in HCC is tightly linked with the infiltration of immune cells.

4 Discussion

RAB22A is a member of the RAS oncogene family that controls membrane properties and vesicle budding, delamination, movement, and fusion and is central to ensuring that cargo is transported to its correct destination. *RAB22A* is referred to in the early formation of endosomes and regulates vesicle transport (28, 29).

Furthermore, *RAB22A* is a critical oncogene that has a crucial impact on the course of many different forms of cancer (12, 30). *RAB22A* promotes the epithelial–mesenchymal transition of papillary thyroid cancer cells, thereby promoting their proliferation, migration, and invasion (31). CD147 is recycled by *RAB22A* to control lung carcinoma cell motility and invasion (13). In metastatic breast cancer, hypoxia facilitates MV production and HIF-dependent RAB22A gene expression (14). In addition, *RAB22A* is involved in a miRNA downregulation mechanism in which the overexpression of small GTPases promotes tumor growth and carcinogenesis. Several tumor models, including kidney, colorectal, glioma, and bile duct cancer, have utilized *RAB22A* as

10.3389/fimmu.2023.1086342

TABLE 3 Correlation analysis of *RAB22A* expression with immune cell biomarkers.

Description	Gene markers	LIHC	
		Cor	*P* -value
CD8+ T cell	**CD8A**	-0.463	**<0.001**
	CD8B	-0.418	**<0.001**
T cell (general)	**CD3D**	-0.446	**<0.001**
	CD3E	-0.561	**<0.001**
	CD2	-0.517	**<0.001**
B cell	**CD19**	-0.338	**<0.001**
	CD79A	-0.487	**<0.001**
Monocyte	**CD86**	-0.515	**<0.001**
	CD115 (CSF1R)	-0.530	**<0.001**
TAM	**CCL2**	-0.525	**<0.001**
	CD68	-0.440	**<0.001**
	IL10	-0.472	**<0.001**
M1 Macrophage	INOS (NOS2)	-0.089	0.099
	IRF5	0.003	0.962
	COX2 (PTGS2)	-0.501	**<0.001**
M2 Macrophage	**CD163**	-0.480	**<0.001**
	VSIG4	-0.488	**<0.001**
	MS4A4A	-0.512	**<0.001**
Neutrophils	**CD66b (CEACAM8)**	-0.106	**0.049**
	CD11b (ITGAM)	-0.330	**<0.001**
	CCR7	-0.552	**<0.001**
Natural killer cell	KIR2DL1	-0.043	0.422
	KIR2DL3	-0.184	**<0.001**
	KIR2DL4	-0.186	**<0.001**
	KIR3DL1	-0.105	0.050
	KIR3DL2	-0.221	**<0.001**
	KIR3DL3	-0.050	0.357
	KIR2DS4	-0.036	0.510
	HLA-DPB1	-0.490	**<0.001**
	HLA-DQB1	-0.454	**<0.001**
	HLA-DRA	-0.480	**<0.001**
	HLA-DPA1	-0.485	**<0.001**
	BDCA-1 (CD1C)	-0.426	**<0.001**
Dendritic cell	**BDCA-4 (NRP1)**	-0.195	**<0.001**
	CD11c (ITGAX)	-0.330	**<0.001**
Th1	**T-bet (TBX21)**	-0.436	**<0.001**
	STAT4	-0.259	**<0.001**
	STAT1	-0.192	**<0.001**

(Continued)

10.3389/fimmu.2023.1086342

TABLE 3 Continued

Description	Gene markers	LIHC	
		Cor	*P* -value
	IFN-g (IFNG)	-0.296	**<0.001**
	TNF-a (TNF)	-0.431	**<0.001**
Th2	**GATA3**	-0.499	**<0.001**
	STAT6	-0.003	0.957
	STAT5A	-0.250	**<0.001**
	IL13	-0.013	0.813
Tfh	BCL6	-0.009	0.866
	IL21	-0.160	**0.003**
	STAT3	-0.233	**<0.001**
	IL17A	-0.040	0.457
Th17	**FOXP3**	-0.226	**<0.001**
	CCR8	-0.320	**<0.001**
	STAT5B	0.162	**0.003**
	TGFb (TGFB1)	-0.410	**<0.001**
T cell exhaustion	**PD-1 (PDCD1)**	-0.429	**<0.001**
	CTLA4	-0.413	**<0.001**
	LAG3	-0.234	**<0.001**
	TIM-3 (HAVCR2)	-0.512	**<0.001**
	GZMB	-0.345	**<0.001**
Treg	**FOXP3**	-0.226	**<0.001**

The bold values indicates that the correlation analysis between RAB22A and biomarker of immune cell is statistically significant.

a target gene for miRNAs (32–34). Changes in *RAB22A* in HCC may be significant as hepatocytes always maintain high metabolic levels and active vesicular transport; nevertheless, the potential effect on *RAB22A* in HCC is unclear.

In the present study, we first found that *RAB22A* was upregulated in HCC and various malignant tumors by analyzing multiple databases. Subsequently, we verified the elevation of *RAB22A* expression in HCC cell lines and HCC samples using western blotting, qRT-PCR, and IHC *in vitro*. Overexpression of *RAB22A* in HCC tissues was closely associated with clinicopathologic features. The ROC curve analysis suggested RAB22A as a promising diagnostic biomarker for differentiating HCC from normal tissues. Moreover, the overexpression of *RAB22A* was interrelated with a poor prognosis of HCC, as indicated by OS, DSS, and PFI.

To elucidate the potential biological functions and regulatory pathways of *RAB22A*, we investigated genes encoding *RAB22A*-related proteins and co-expression genes in HCC tissues. mRNA sequencing data with HCC were evaluated in the TCGA database, while the DEGs associated with *RAB22A* in HCC were shown in a heat map. Insights gained from pathway enrichment analyses using GO and KEGG indicated that *RAB22A* has far-reaching effects on the transcriptome. Through enrichment pathway analysis, we verified that these DEGs were involved in proteasomal protein catabolic process, ncRNA processing, ribosomes, and ribosomal subunits, protein serine/threonine kinase activity, GTPase combining, herpes simplex virus type 1 infection, multiple neurodegenerative illnesses, and Alzheimer's disease pathways. Next, we analyzed 30 signaling pathways positively and negatively correlated with *RAB22A* expression using GSEA. Overexpression of *RAB22A* was linked to processes such as cell adhesion *via* the plasma membrane (35), nonsense-mediated decay of nuclear-transcribed mRNA (36), axon guidance (37), oxidative phosphorylation (38), FCGR activation (39), and eukaryotic translation elongation (40) in a GSEA of HCC. Overall, we suggest that *RAB22A* may participate in various cellular immune functions and intracellular transport and may facilitate the advance of HCC by adjusting these signaling pathways.

Subsequently, we built a PPI network using Cytoscape. One central gene cluster (total score ≥ 14,000) and the top 10 central genes were filtered, namely *RAB22A*, *RABGEF1*, *VPS45*, *VPS18*, *VPS11*, *MON1A*, *VPS39*, *VPS16*, *ZFYV20*, and *VPS*. These findings provide important insights for subsequent study designs and experimental validations.

M6A methylation has been examined to elucidate the mechanisms of HCC since it has been proven to affect cancer *via* numerous mechanisms (41). m6A is a critical player in HCC (42, 43).

Methyltransferases (the "Writers"), demethylases (the "Erasers"), and methylated reading proteins have access to the same m6A methylation (Readers). Methylation transferases, such as METTL3/14, WTAP, and KIAA1429, are primarily responsible for catalyzing the m6A alteration of adenosine on mRNA. Demethylases, such as FTO and ALKHB5, facilitate the demethylation of m6A. Methylation reading proteins, such as YTHDF 1-3 and YTHDC 1-3, recognize RNA methylation and play a role in regulatory processes, such as RNA translation, degradation, and miRNA processing (44). Further analysis of the connection between RAB22A expression and m6A methylation proteins revealed a positive and significant association between *RAB22A* expression and the expression of methylation transferases, demethylases, and methylated reading proteins. Patients with HCC have a poor prognosis because m6A-modified proteins are highly elevated in the disease, and their overexpression increases the disease progression. Several reports have verified that IGF2BP1, YTHDF1, and RBM15 are all highly elevated in HCC and contribute to its development and progression. These findings indicate that m6A may alter the *RAB22A* gene to enhance the consistency of its mRNA, hence boosting the occurrence and development of HCC. Evidence for lncRNA-miRNA-mRNA ceRNA networks' regulatory role in cancers is mounting (45). Based on these predictions, we constructed a ceRNA regulatory network that predicted that *RAB22A* might affect several critical pathways of HCC regulatory mechanisms. We intend to conduct further experiments to validate this network.

Cancer cells that invade Immune cells, known as tumor-infiltrating immune cells (TIICs), play a key regulatory role in tumorigenesis and development (46). The HCC prognosis may be affected by the presence of TIICs, which are essential for HCC development (28–30). TIICs facilitate a tangled web of cellular interactions that boost the immunosuppressive milieu, facilitate immune escape, and ultimately aid in tumor progression. Changes in the immune environment of the liver can cause liver lesions, such as chronic inflammation and fibrosis/cirrhosis (22, 47). *RAB22A* is a regulator of immune functions. Independent studies have also shown that Th2 cells contribute to cancer development and progression (48, 49). Effector T helper cell subgroups are essential for coordinating immune responses to diverse infections and participate in the nosogenesis of numerous inflammatory disorders, including autoimmunity and allergies (50). pDCs are a sentinel cell type that can test pathogen-derived nucleic acids and reactions *via* the rapid and significant production of type I interferons, primarily in autoimmune diseases, immune deficiencies, and cancer (51). Cytotoxic T cells and DCs are also essential effectors of antitumor immunity (40, 52). These findings suggest that RAB22A plays an indispensable role in regulating immune cell infiltration in HCC.

We also demonstrated that *RAB22A* expression was significantly correlated with 66 immune markers. These results indicate that the upregulation of *RAB22A* expression in HCC is linked to immune cell infiltration. Understanding the function of RAB22A in immune activation will help to facilitate future research using various immune cell types and animal models.

Although our study identified the molecular mechanism of *RAB22A* in HCC through bioinformatics analysis, there remain limitations. Firstly, to elucidate the effect of *RAB22A* on HCC, several subjective factors, such as the treatment details received by patients and follow-up, should be considered simultaneously. However, some experiments were conducted in different centers, thereby limiting the information or causing inconsistency in the public database, which led to some errors. Secondly, the number of patients with cancer in the experimental control group was different from that in the current study; hence accessional studies are needed to eliminate the error caused by sample offset.

Thirdly, multicenter investigations based on communal databases seek to compensate for the paucity of single-center studies. However, retrospective studies have drawbacks, including inconsistent interventions and a lack of data. Since this study is retrospective, prospective investigations should be undertaken to eliminate analytical bias. Based on previous verifications, the results are robust; advancements in single-cell and spatial transcriptomics technologies allowed for the increased use of single-cell multi-omics technologies to gain insights into complex cellular ecosystems and biological processes. Currently, there is a gap in the rapidly growing single-cell multi-omics data, while effective methods for comprehensive analysis of these inherently sparse and heterogeneous data are limited. Therefore, new algorithms, such as SMGR (53) and spaCI (54), have been derived to address this gap. Single-cell multi-omics gene co-regulation algorithms provide multiple regulatory stages to study the control of cellular heterogeneity and complex biological mechanisms, which provide great clinical value for identifying mechanisms, targets, and predictors to enhance translational therapy. The spaCI algorithm can detect upstream transcription factors (TFS) mediating the L-R signaling axis, which provides insights into the underlying molecular mechanisms of the intercellular crosstalk. These emerging algorithms can be used to verify the biological mechanism of RAB22A in HCC.

In summary, we demonstrated, to the best of our knowledge, for the first time that *RAB22A* promotes carcinogenesis *via* m6A methylation and ceRNA network processes and is strongly linked with HCC development, poor survival, and immune infiltration.

Data availability statement

The article/Supplementary material contains the original contributions presented in the study. Any additional questions can be forwarded to the corresponding authors. The datasets presented in this study can be found in online repositories. The names of the repository can be found below: https://portal.gdc.cancer.gov.

Ethics statement

The studies involving human participants were reviewed and approved by First Affiliated Hospital of Harbin Medical University's ethics committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

FW, FM, XL, QL, and LJ contributed equally to this work. ZL and YC designed this research. FW and FM drafted this manuscript. XL , XW and QL performed the data collection and analysis. JL, RZ, YunZ, YuZ and SJ participated in the data interpretation and study design. All authors approved the final manuscript.

Funding

This work was supported by The National Natural Scientific Foundation of China (Grant Nos. 81972230), The Heilongjiang Postdoctoral Science Foundation (Grant No. LBH-Z20178), The Scientific Foundation of the First Affiliated Hospital of Harbin Medical University (Grant No. 2021B03), the Excellent Youth Science Fund of the First Affiliated Hospital of Harbin Medical University (Grant No. 2021Y01), and the Chen Xiaoping Foundation for the Development of Science and Technology of Hubei Province (CXPJJH122002-092).

Acknowledgments

We thank the patients who provided the experimental samples and the teachers for providing our funding.

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/fimmu.2023.1086342/full#supplementary-material

References

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

2. Konyn P, Ahmed A, Kim D. Current epidemiology in hepatocellular carcinoma. *Expert Rev Gastroenterol Hepatol* (2021) 15:1295–307. doi: 10.1080/17474124.2021.1991792

3. Cui Y, Sun D, Liu L. Upregulation of cystatin SN promotes hepatocellular carcinoma progression and predicts a poor prognosis. *J Cell Physiol* (2019) 234:22623–34. doi: 10.1002/jcp.28828

4. Stenmark H. Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol* (2009) 10:513–25. doi: 10.1038/nrm2728

5. Antonyak MA, Wilson KF, Cerione RAR. R(h)oads to microvesicles. *Small GTPases* (2012) 3:219–24. doi: 10.4161/sgtp.20755

6. Mesa R, Magadán J, Barbieri A, López C, Stahl PD, Mayorga LS. Overexpression of Rab22a hampers the transport between endosomes and the golgi apparatus. *Exp Cell Res* (2005) 304:339–53. doi: 10.1016/j.yexcr.2004.11.017

7. Weigert R, Yeung AC, Li J, Donaldson JG. Rab22a regulates the recycling of membrane proteins internalized independently of clathrin. *Mol Biol Cell* (2004) 15:3758–70. doi: 10.1091/mbc.e04-04-0342

8. Kauppi M, Simonsen A, Bremnes B, Vieira A, Callaghan J, Stenmark H, et al. The small GTPase Rab22 interacts with EEA1 and controls endosomal membrane trafficking. *J Cell Sci* (2002) 115:899–911. doi: 10.1242/jcs.115.5.899

9. Sun L, He M, Xu N, Xu D-H, Ben-David Y, Yang Z-Y, et al. Regulation of RAB22A by mir-193b inhibits breast cancer growth and metastasis mediated by exosomes. *Int J Oncol* (2018) 53:2705–14. doi: 10.3892/ijo.2018.4571

10. Yin Y, Zhang B, Wang W, Fei B, Quan C, Zhang J, et al. miR-204-5p inhibits proliferation and invasion and enhances chemotherapeutic sensitivity of colorectal cancer cells by downregulating RAB22A. *Clin Cancer Res* (2014) 20:6187–99. doi: 10.1158/1078-0432.CCR-14-1030

11. Zheng S, Jiang F, Ge D, Tang J, Chen H, Yang J, et al. LncRNA SNHG3/miRNA-151a-3p/RAB22A axis regulates invasion and migration of osteosarcoma. *BioMed Pharmacother* (2019) 112:108695. doi: 10.1016/j.biopha.2019.108695

12. Zhou Y, Wu B, Li J-H, Nan G, Jiang J-L, Chen Z-N. RAB22A enhances CD147 recycling and is required for lung cancer cell migration and invasion. *Exp Cell Res* (2017) 357:9–16. doi: 10.1016/j.yexcr.2017.04.020

13. Wang T, Gilkes DM, Takano N, Xiang L, Luo W, Bishop CJ, et al. Hypoxia-inducible factors and RAB22A mediate formation of microvesicles that stimulate breast cancer invasion and metastasis. *Proc Natl Acad Sci USA* (2014) 111:E3234–42. doi: 10.1073/pnas.1410041111

14. Mayorga LS, Cebrian I. RAB22A: A novel regulator of immune functions. *Mol Immunol* (2019) 113:87–92. doi: 10.1016/j.molimm.2018.03.028

15. Johnson DC, Baines JD. Herpesviruses remodel host membranes for virus egress. *Nat Rev Microbiol* (2011) 9:382–94. doi: 10.1038/nrmicro2559

16. Isberg RR, O'Connor TJ, Heidtman M. The legionella pneumophila replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* (2009) 7:13–24. doi: 10.1038/nrmicro1967

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

18. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* (2018) 173:400–16.e11. doi: 10.1016/j.cell.2018.02.052

19. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* (2012) 41:D991–5. doi: 10.1093/nar/gks1193

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

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

22. Li B, Severson E, Pignon J-C, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* (2016) 17:174. doi: 10.1186/s13059-016-1028-7

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

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

25. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

26. von Mering C. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* (2003) 31:258–61. doi: 10.1093/nar/gkg034

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

28. He H, Dai F, Yu L, She X, Zhao Y, Jiang J, et al. Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues. *Gene Expr* (2002) 10:231–42. doi: 10.3727/000000002783992406

29. Magadán JG, Barbieri MA, Mesa R, Stahl PD, Mayorga LS. RAB22A regulates the sorting of transferrin to recycling endosomes. *Mol Cell Biol* (2006) 26:2595–614. doi: 10.1128/MCB.26.7.2595-2614.2006

30. Wang J, Luo X, Lu J, Wang X, Miao Y, Li Q, et al. RAB22A promotes the proliferation, migration, and invasion of lung adenocarcinoma *via* up-regulating PI3K/Akt/mTOR signaling pathway. *Exp Cell Res* (2022) 416:113179. doi: 10.1016/j.yexcr.2022.113179

31. Luo X, Wang J, Lu J, Wang X, Miao Y, Li Q, et al. RAB22A promotes epithelial-mesenchymal transition in papillary thyroid carcinoma by activating PI3K/AKT/mTOR signaling pathway. *BioMed Res Int* (2022) 2022:1–17. doi: 10.1155/2022/1874550

32. Liu Q, Lei C. *LINC01232* serves as a novel biomarker and promotes tumour progression by sponging miR-204-5p and upregulating RAB22A in clear cell renal cell carcinoma. *Ann Med* (2021) 53:2153–64. doi: 10.1080/07853890.2021.2001563

33. Kong W, Li H, Xie L, Cui G, Gu W, Zhang H, et al. LncRNA MCF2L-AS1 aggravates the malignant development of colorectal cancer *via* targeting miR-105-5p/RAB22A axis. *BMC Cancer* (2021) 21:1069. doi: 10.1186/s12885-021-08668-w

34. Cui Y, Liang S, Liu L. ABCA8 is regulated by miR-374b-5p and inhibits proliferation and metastasis of hepatocellular carcinoma through the ERK/ZEB1 pathway. *J Exp Clin Cancer Res* (2020) 39:90. doi: 10.1186/s13046-020-01591-1

35. Piao G, Yuchao H, Lu C, Lisha Q, Dongming L, Ziye C, et al. Cytosolic phospholipase A2α modulates cell-matrix adhesion *via* the FAK/paxillin pathway in hepatocellular carcinoma. *Cancer Biol Med* (2019) 16:377. doi: 10.20892/j.issn.2095-3941.2018.0386

36. Zhang H, You Y, Zhu Z. The human RNA surveillance factor up-frameshift 1 inhibits hepatic cancer progression by targeting MRP2/ABCC2. *BioMed Pharmacother* (2017) 92:365–72. doi: 10.1016/j.biopha.2017.05.090

37. Liang JQ, Teoh N, Xu L, Pok S, Li X, Chu ESH, et al. Dietary cholesterol promotes steatohepatitis related hepatocellular carcinoma through dysregulated metabolism and calcium signaling. *Nat Commun* (2018) 9:4490. doi: 10.1038/s4146-018-06931-6

38. Yu H, Wang C, Ke S, Bai M, Xu Y, Lu S, et al. Identification of CFHR4 as a potential prognosis biomarker associated with lmmune infiltrates in hepatocellular carcinoma. *Front Immunol* (2022) 13:892750. doi: 10.3389/fimmu.2022.892750

39. Kudo Y, Sugimoto M, Arias E, Kasashima H, Cordes T, Linares JF, et al. PKCλ/ι loss induces autophagy, oxidative phosphorylation, and NRF2 to promote liver cancer progression. *Cancer Cell* (2020) 38:247–62.e11. doi: 10.1016/j.ccell.2020.05.018

40. Oh DY, Fong L. Cytotoxic CD4+ T cells in cancer: Expanding the immune effector toolbox. *Immunity* (2021) 54:2701–11. doi: 10.1016/j.immuni.2021.11.015

41. Sun T, Wu R, Ming L. The role of m6A RNA methylation in cancer. *BioMed Pharmacother* (2019) 112:108613. doi: 10.1016/j.biopha.2019.108613

42. Wang P, Wang X, Zheng L, Zhuang C. Gene signatures and prognostic values of m6A regulators in hepatocellular carcinoma. *Front Genet* (2020) 11:540186. doi: 10.3389/fgene.2020.540186

43. Zhou T, Li S, Xiang D, Liu J, Sun W, Cui X, et al. m6A RNA methylation-mediated HNF3γ reduction renders hepatocellular carcinoma dedifferentiation and sorafenib resistance. *Sig Transduct Target Ther* (2020) 5:296. doi: 10.1038/s41392-020-00299-0

44. Chen X-Y, Zhang J, Zhu J-S. The role of m6A RNA methylation in human cancer. *Mol Cancer* (2019) 18:103. doi: 10.1186/s12943-019-1033-z

45. Qi X, Zhang D-H, Wu N, Xiao J-H, Wang X, Ma W. ceRNA in cancer: possible functions and clinical implications. *J Med Genet* (2015) 52:710–8. doi: 10.1136/jmedgenet-2015-103334

46. Garnelo M, Tan A, Her Z, Yeong J, Lim CJ, Chen J, et al. Interaction between tumour-infiltrating b cells and T cells controls the progression of hepatocellular carcinoma. *Gut* (2017) 66:342–51. doi: 10.1136/gutjnl-2015-310814

47. Ruf B, Heinrich B, Greten TF. Immunobiology and immunotherapy of HCC: spotlight on innate and innate-like immune cells. *Cell Mol Immunol* (2021) 18:112–27. doi: 10.1038/s41423-020-00572-w

48. Orecchioni M, Bedognetti D, Newman L, Fuoco C, Spada F, Hendrickx W, et al. Single-cell mass cytometry and transcriptome profiling reveal the impact of graphene on human immune cells. *Nat Commun* (2017) 8:1109. doi: 10.1038/s41467-017-01015-3

49. Mantovani A, Romero P, Palucka AK, Marincola FM. Tumour immunity: effector response to tumour and role of the microenvironment. *Lancet* (2008) 371:771–83. doi: 10.1016/S0140-6736(08)60241-X

50. Zhu J. T Helper cell differentiation, heterogeneity, and plasticity. *Cold Spring Harb Perspect Biol* (2018) 10:a030338. doi: 10.1101/cshperspect.a030338

51. Reizis B. Plasmacytoid dendritic cells: Development, regulation, and function. *Immunity* (2019) 50:37–50. doi: 10.1016/j.immuni.2018.12.027

52. Martinek J, Wu T-C, Cadena D, Banchereau J, Palucka K. Interplay between dendritic cells and cancer cells. *Int Rev Cell Mol Biol* (2019) 348:179–215. doi: 10.1016/bs.ircmb.2019.07.008

53. Song Q, Zhu X, Jin L, Chen M, Zhang W, Su J. SMGR: a joint statistical method for integrative analysis of single-cell multi-omics data. *NAR Genom Bioinform* (2022) 4:lqac056. doi: 10.1093/nargab/lqac056

54. Tang Z, Zhang T, Yang B, Su J, Song Q. spaCI: deciphering spatial cellular communications through adaptive graph model. *Brief Bioinform* (2023) 24:bbac563. doi: 10.1093/bib/bbac563

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fimmu.2023.1109503

OPEN ACCESS

EDITED BY
Hao Zhang,
Chongqing Medical University, China

REVIEWED BY
Xiaoting Huang,
Guangzhou Medical University Cancer Hospital, China
Juan Lu,
Southern Medical University, China
Yulai Zhou,
The University of Texas Health Science Center at San Antonio, United States

*CORRESPONDENCE
Sufang Qiu
sufangqiu@fjmu.edu.cn
Desheng Wang
wangds@fjmu.edu.cn

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

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

RECEIVED 27 November 2022
ACCEPTED 21 March 2023
PUBLISHED 31 March 2023

CITATION
Chen X, Ding Q, Lin T, Sun Y, Huang Z, Li Y, Hong W, Chen X, Wang D and Qiu S (2023) An immune-related prognostic model predicts neoplasm-immunity interactions for metastatic nasopharyngeal carcinoma. *Front. Immunol.* 14:1109503. doi: 10.3389/fimmu.2023.1109503

An immune-related prognostic model predicts neoplasm-immunity interactions for metastatic nasopharyngeal carcinoma

Xiaochuan Chen[1,2†], Qin Ding[1,2†], Ting Lin[1,2†], Yingming Sun[3], Zongwei Huang[1,2], Ying Li[1,2], Wenquan Hong[1,2], Xin Chen[1,2], Desheng Wang[4*] and Sufang Qiu[1,2*]

[1]Department of Radiation Oncology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China, [2]Fujian Provincial Key Laboratory of Translational Cancer Medicine, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China, [3]Department of Radiation and Medical Oncology, Affiliated Sanming First Hospital of Fujian Medical University, Sanming, China, [4]Department of Otolaryngology, Fujian Medical University Union Hospital, Fuzhou, China

Background: The prognosis of nasopharyngeal carcinoma (NPC) has been recognized to improve immensely owing to radiotherapy combined with chemotherapy. However, patients with metastatic NPC have a poor prognosis. Immunotherapy has dramatically prolonged the survival of patients with NPC. Hence, further research on immune-related biomarkers is imperative to establish the prognosis of metastatic NPC.

Methods: 10 NPC RNA expression profiles were generated from patients with or without distant metastasis after chemoradiotherapy from the Fujian Cancer Hospital. The differential immune-related genes were identified and validated by immunohistochemistry analysis. The method of least absolute shrinkage and selection operator (LASSO)was used to further establish the immune-related prognostic model in an external GEO database (GSE102349, n=88). The immune microenvironment and signal pathways were evaluated in multiple dimensions at the transcriptome and single-cell levels.

Results: 1328 differential genes were identified, out of which 520 were upregulated and 808 were downregulated. Notably, most of the immune genes and pathways were down-regulated in the metastasis group. A prognostic immune model involving nine hub genes. Patients in low-risk group were characterized by survival advantage, hot immune phenotype and benefit from immunotherapy. Compared with immune cells, malignant cell exhibited the most active levels of risk score by ssGSEA. Accordingly, intercellular communications including LT, CD70, CD40 and SPP1, and the like, between high-risk and low-risk were explored by the R package "Cellchat".

Conclusion: We have constructed a model based on immunity of metastatic NPC and determined its prognostic value. The model identified the level of immune cell infiltration, cell-cell communication, along with potential immunotherapy for metastatic NPC.

KEYWORDS

immunotherapy, bioinformatics, metastatic nasopharyngeal carcinoma, immune microenvironment, mRNA transcriptome sequencing and single cell sequencing

Introduction

Nasopharyngeal carcinoma (NPC), an Epstein-Barr virus (EBV)-associated cancer that is prevalent in Southern China (1), has been recognized to have a favorable prognosis owing to radiotherapy combined with chemotherapy during the past decades. Apart from EBV infection, human papillomavirus (HPV) infection, alcohol and tobacco consumption, smoking, and the consumption of salt-preserved foods have recently been identified as high-risk factors (2). Although most patients reach complete clinical remission, it has been suggested that patients with recurrence or metastasis have a poor prognosis. The application of intensity-modulated radiotherapy has improved the treatment outcome of NPC, especially the local control rate, but the impact on distant metastasis is minimal. The 5-year survival rate of patients with early-stage NPC can reach more than 90% with a relatively low rate of 60% for patients with advanced stage (3). Hence, currently, the focus should be on the cure of metastatic NPC. Exploring new therapeutic targets and developing new molecularly targeted drugs are definitely the direction of future research. In addition, exploring the molecular mechanism of distant metastasis of NPC and screening high-risk groups will also facilitate individualized response in the initial treatment.

Immune checkpoint blockade (ICB)-based immunotherapy, such as programmed cell death ligand 1 (PD-L1) and interferon (IFN)-γ, has dramatically changed the treatments of cancer to prolong the patients' survival (4). Particularly, the clinical research of immunotherapy has contributed majorly to the individual treatments of malignant tumors (5). However, the most well-known research recommends Pembrolizumab as the first-line treatment for PD-L1-positive recurrent or metastatic head and neck squamous cell carcinomas (6). The response to immunotherapy for the treatment of metastatic NPC is inconclusive.

In recent years, the assessment of immunotherapy efficacy has become a major challenge for clinicians to individualize treatment. Although no accepted immune-related risk model for predicting prognosis exists, the reported models have shown decent predictive validity in certain cancers (7–12). The focus of immune-related prediction models is not only restricted to the genomic level but also extended to the transcriptome level, single-cell level, and so on (13–18). However, there are not many studies on immune-related prognosis models integrating single-cell RNA and mRNA levels in metastatic NPC. Therefore, it is of great significance to explore novel immune-related diagnostics and therapeutics for patients with metastatic NPC.

This study aimed to (i) identify the immune-related genes, (ii) reveal the underlying pathway associated with metastatic NPC, (iii) establish the prognostic immune model and evaluate its prognostic value, and (iv) validate the predictive validity of the model from various aspects.

Materials and methods

Patients' samples

10 NPC tumor tissue samples were obtained from the patients who were diagnosed and treated at the Fujian Cancer Hospital between May 9, 2013, and August 2, 2016. All 10 patients met the following eligibility criteria: newly diagnosed NPC, received standardized radiotherapy and chemotherapy, ≥18 years old, adequate hematological, renal and hepatic functions, and no other malignant diseases. All the patients provided written informed consent. The study was approved by the Ethics Committee of Fujian Cancer Hospital and Fujian Medical University Cancer Hospital (approval number SQ2019-035-01). The tissue samples were stored in liquid nitrogen for subsequent RNA extraction. During the 5 years of follow-up, 5 samples were from patients with disease progression after radiotherapy and chemotherapy, containing 3 liver metastases, 1 bone metastasis, and 1 lung metastasis. While the other 5 samples were evaluated as having a complete or partial response after the treatment.

As an external validation cohort, RNA-seq data of NPC from the GEO database (https://www.ncbi.nlm.nih.gov/geo/, GSE102349) were selected to verify the reliability and applicability of the data of this study (19, 20). The single-cell dataset GSE150430 was designed to validate the accuracy of the model at the individual cell level and to probe the communication of cells and ligand receptors in the immune microenvironment of NPC. Also, the tumor tissue biopsies of 74 NPC patients treated in our hospital in 2021 and 2022 were used for the immunohistochemistry to validate CD8 T cell infiltration and immune checkpoints expression, including 11 cases in the metastatic group and 63 cases in the

non-metastatic group (Supplementary Table 1). The 8th edition of the American Joint Committee on Cancer (AJCC) Staging Manual was used to restage all the patients.

Immunohistochemistry analysis

NPC biopsies were fixed with 10% formalin overnight and processed into 5-μm-thick paraffin sections. The slides were then analyzed by immunohistochemistry with anti-human CD8 (Cat # ab237709; Abcam), anti-human PD1 (Cat # ab52587; Abcam), and anti-human PD-L1 (Cat # ab213524; Abcam) followed by HRP secondary antibody (Cat #ab205718; Abcam) and DAB staining. Images were obtained using a microscope (BX43; Olympus, Japan). Histochemistry score (H-score) was used to evaluate the expression. H-score = (percentage of cells of weak intensity × 1) + (percentage of cells of moderate intensity × 2) + (percentage of cells of strong intensity × 3).

Construction and validation of immune-related prognostic model

The R package "ggplot2" was employed to visualize DEGs from sequencing data of NPC samples in Fujian Cancer Hospital (21). The cut-off values met the following two conditions: fold-change of >2 and the p-value of<0.05. Gene ontology (GO) (22, 23) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses (24) were applied to further explore the pathways of DEGs enrichment. A false-discovery rate of<0.05 was set as the cut-off value. The immune gene data was downloaded through the ImmPort data portal (www.immport.org/immport-open/public/home/home), and 2,498 immune-related genes were obtained. Then the intersection of the DEGs and the immune-related genes was selected as differentially expressed immune-related genes. Progression-free survival (PFS) was subjected to minimum absolute shrinkage and selection operator (LASSO) Cox regression with 10-fold cross-validation to screen for DEGs with prognostic value on the basis of the univariate Cox analysis. The R package "glmnet" was employed to determine the gene signatures containing the biomarkers most helpful for prognosis (25). The prognosis risk score was established by linearly combining the following formula:

$$\text{risk score} = \sum_{1}^{n}(\text{exp} \times \text{coef})$$

where exp denotes the gene expression value, while coef refers to the coefficient of a gene in LASSO analysis.

To assess the predictive power of our prognostic risk model, receiver operating characteristic (ROC) for 1- and 3-year survival were performed in the validation cohort GSE102349 using the R package "timeROC". Next, the samples were divided into high-risk and low-risk groups according to the best cut-off value of the risk score from the R package "survival" for survival analysis. The survival curves were compared using the Kaplan-Meier method and the log-rank test. The univariate and multivariate Cox regression models were applied to determine whether the risk score was an independent prognostic factor.

Immune- and carcinogenesis-related estimation in multiple dimensions

To evaluate the infiltration of immune cells from several aspects, we adopted multiple immune scoring approaches, like TIMER and ssGSEA algorithms (26, 27). The immune scores and tumor purity were estimated by the R package "ESTIMATE" (28). From an earlier study, we retrieved a group of six inhibitory immune checkpoints that displayed immune therapeutic efficacy (29). Gene sets that displayed T cell-inflamed gene expression profile (GEP) and tertiary lymphatic structure (TLS) were acquired (30, 31). Furthermore, we assessed the enrichment of 10 oncogenic pathways using the ssGSEA method (32). The score of activation minus the score of repression represented the final score of each pathway. We used a validated set of 31 genes related to cell cycle progression (CCP) to estimate the rate of cell proliferation (33). The cluster score was calculated as the average expression level of CCP-related pathways by subtracting the mean level.

Prediction of the immunotherapy response

To assess the predictive efficacy of the model for immunotherapy efficacy, we collected several immunotherapy cohorts from the GEO database and the TIGER website (http://tiger.canceromics.org/#/), including nasopharyngeal carcinoma-GSE102349, melanoma-GSE91061, melanoma-PRJEB23709, NSCLC-GSE126044. We visually compared the proportion of patients with and without response to immunotherapy in high- and low-risk groups.

Single-cell RNA-seq analysis

This study performed quality control, downscaling, and clustering of scRNA-seq data as well using Seurat (v.4.0.4) (34). To ensure data quality, genes detected in less than 3 cells and cells with less than 250 genes detected were excluded, and the percentage of mitochondria was limited to less than 35% (35). Data were processed by the logNormalize method for normalization. The nonlinear dimensionality reduction method Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) was utilized for unsupervised classification and unbiased visualization of cell populations on two-dimensional maps (36). TISCH (http://tisch.comp-genomics.org/) provides detailed cell type annotations at the single-cell level (35). After that, the "FindAllMarkers" function was configured to identify marker genes in each cluster using a filter value of absolute log2 fold change (FC) ≥ 0.3 and a minimum cell cluster fraction of 0.25.

Risk score calculation in single-cell samples

A risk score of each single cell sample from GSE150430 was calculated by single sample Gene Set Enrichment Analysis (ssGSEA) method and was completed using the "GSVA" and "GSEABase" packages in R. We used single-cell data as a reference, apply a newly developed deconvolution algorithms (CIBERSORTx) to the bulk transcriptome data to quantitatively estimate cell-type proportions for each tumor in GEO database (37).

Cell–cell chat analysis

CellChat v1.1.3 software inferred cell-cell communication based on ligand-receptor interactions (38). Cell groups with less than 10 cells were filtered out of cell-cell communication. Pairwise tests were performed on communication probability values to assess their statistical significance.

Statistical analysis

Statistical analysis was done using R software (V.3.6.1) and SPSS software (ver. 25.0). Wilcoxon rank sum test and chi-square test were conducted for continuous and categorical variables, respectively. For all analysis, two-by-two pairs indicate statistically significant differences. *, **, *** and **** indicate, respectively <0.05, <0.01, <0.001, and <0.0001.

Results

Identification of differential immune-related expressions in NPC

The schematic diagram presents the workflow of our study (Figure S1). The RNA-seq profiles were generated for the NPC samples of 10 patients treated at the Fujian Cancer Hospital, 5 of whom were assigned to the non-metastasis group, while the other 5 were in the metastasis group owing to distant metastasis after chemoradiotherapy. The baseline characteristics of patients in the metastatic and non-metastatic groups could be seen in Table 1 (n=10). In general, PCA indicated distinct transcriptional profiles between the metastatic group and the non-metastatic group (Figure S2A). Then, 1328 DEGs were conspicuously illustrated in the volcano plot, with 520 upregulated genes and 808 downregulated genes (Figure 1A). The KEGG and GO analyses are the universally applicable statistical methods of enrichment analysis. The DEGs were enriched in the immune-related pathways of the bubble chart containing signal transduction, adaptive immune system, innate immune system, and hemostasis (Figure 1B). Simultaneously, they were also centralized in the cell periphery, plasma membrane, and immune system processes (Figure 1C). The expression levels of the top 154 immune-related genes selected from the DEGs can be significantly distinguished between the two groups in the heat-map (Figure 1D). Overall, immune gene expression and immune signaling pathway were down-regulated in the metastasis group, indicating a potential "immune-cold" tumor phenotype in the metastasis group. For the validation, Therefore, we performed immunohistochemistry staining of CD8 T cell, PD1, and PD-L1. Our immunohistochemistry analysis showed that PD1 and PD-L1 expressions were down-regulated, and the infiltration of CD8 T cells was decreased in the metastasis NPC group (n=11) compared to the non-metastasis group (n=63, Figures 1E, F).

Establishment and validation of the risk model

The LASSO logistic regression model was applied to establish the prognostic immune biomarkers, which involved 9 hub genes (A2M, APLNR, CD8B, RAC3, PRDX2, ULBP1, TMSB15B, KIR3DL2, and SEMA4F; Figure 2A). The standard for high and low risk scores was evaluated based on cut points associated with the median risk score. Cut-off value of 1.31 for the risk model was identified, which served to divide the patients into high-risk group (with levels of risk score ≥ 1.31) and a low-risk group (with levels of risk score< 1.31). The risk scores were significantly distinguished between the clinical stages I–III and stage IV in GSE102349 (Figure S2B), which indicated that the clinical stage of the tumor could be one of the critical factors in assessing the effect of the treatment. The risk scores were also apparently different between the metastasis and non-metastasis groups in our hospital cohort (Figure S2C). Patients in the high-risk group had worse tumor metastatic presentation, which is indicative of a worse prognosis (Figure 2B). This finding was further validated in a cohort of patients from Fujian Cancer Hospital (Figure 2C).

It was found that APLNR, KIR3DL2, CD8B, and A2M were upregulated in the low-risk group, while PRDX2, ULBP1, TMSB15B, SEMA4F and RAC3 were upregulated in the high-risk group (Figure 2D). The assumption could be proposed that the former 4 genes were protective biomarkers, while the latter 5 genes were risk biomarkers. The area under the ROC curve (AUC) was 0.79 at 1-year, and 0.81 at 3-years, respectively, indicating a high predictive value (Figure 2E). Combining the results of univariate (Figure 2F) and multivariate (Figure 2G) Cox analysis, it appeared that risk scores could be an independent prognostic factor compared to other clinical traits.

Expression profiles and prognostic potency of nine hub genes

In the mRNA sequencing data of NPC from Fujian Cancer Hospital, the expressions of the nine immune-related hub genes were apparently different in the metastasis and non-metastasis groups (Figure 3A). Of the 9 genes, CD8B, APLNR, A2M and KIR3DL2 were upregulated in the non-metastasis group where the patients would

TABLE 1 The baseline characteristics of patients in the metastatic and non-metastatic groups (n=10).

Variables	metastatic group (n=5)	non-metastatic group (n=5)	*P* value[a]
Gender			*1.000*
Male	*4*	*4*	
Female	*1*	*1*	
Age			*0.167*
≤*50*	*2*	*5*	
>*50*	*3*	*0*	
T stage			*0.524*
T1-*2*	*1*	*3*	
T3-*4*	*4*	*2*	
N stage			*1.000*
N0-*1*	*1*	*2*	
N2-*3*	*4*	*3*	
M stage			*0.008*
M*0*	*0*	*5*	
M*1*	*5*	*0*	
Clinical stage			*0.008*
II-III	*0*	*5*	
IV	*5*	*0*	
Survival			*0.048*
Alive	*1*	*5*	
Dead	*4*	*0*	
Pathological type[b]			*1.000*
WHO I	*0*	*1*	
WHO II	*1*	*1*	
WHO III	*4*	*3*	

[a]*P* values were two-sided using Fisher's exact test, [b]Pathological type includes WHO type I: keratinizing squamous cell carcinoma, WHO type II: non-keratinizing differentiated carcinoma and WHO type III: non-keratinizing undifferentiated carcinoma.

have lower risk and gain better outcomes. In contrast, SEMA4F, PRDX2, RAC3, ULBP1 and TMSB15B were upregulated in the metastasis group where the patients would have higher risk and suffer worse outcomes (Figure 3A). To verify the predictive validity of nine hub genes for prognostic outcome, survival analysis demonstrated promising prognostic differentiation (Figures 3B–J).

Enrichment pathways of hub genes and correlation with oncogenic pathways, proliferative activity

The pathway in which the gene is enriched tends to indicate that the gene plays a role in that physiological process. Using the GSEA method, the high-risk group was mainly distributed into the E2F, G2M checkpoint, and MYC targets, which were closely related to interactions on angiogenesis, extracellular matrix remodeling, and tumor cell-endothelial cell interactions (Figure 4A). Correspondingly, the low-risk group was mainly distributed in the INF-γ, INF-α, and inflammatory responses, which were closely related to antitumor effect in anti-tumor immune response (Figure 4B).

Moreover, patient samples from high- and low-risk groups showed significant differences in scores across the ten carcinogenic pathways (Figure 4C). Patients in the high-risk group had higher oncogenic pathogenic activity, predicting that a higher risk of cancer progression was involved. And the CCP scores corroborated this finding, with patients in the high-risk group having high CCP scores, which suggested that the tumors had stronger proliferative activity (Figure 4D).

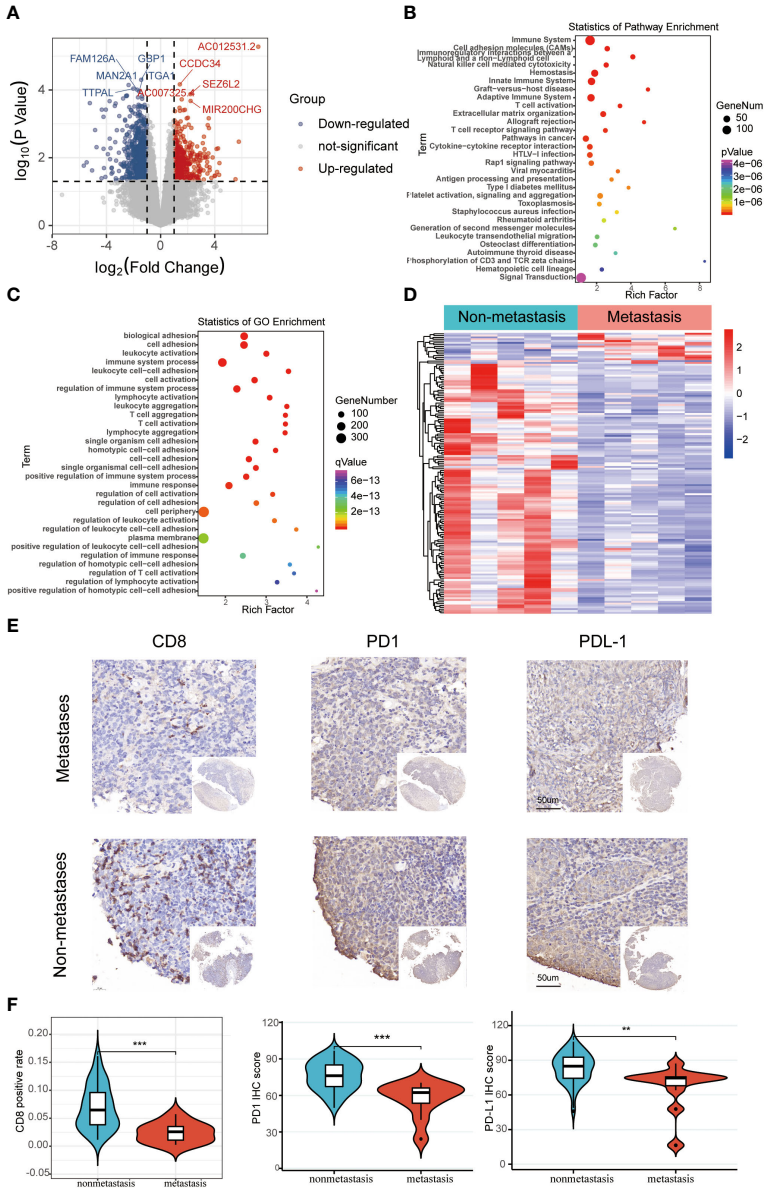


FIGURE 1
The differentially expressed immune-related genomic biomarkers in nasopharyngeal carcinoma (NPC). **(A)** All 1328 differential genes assessed from the tumor tissues are shown in the volcano plot; red dots for upregulated genes (520 genes), while blue dots represent downregulated genes (808 genes); **(B, C)** Statistics of enrichment analysis using KEGG and GO were concentrated on the immune-related cellular components, biological processes and pathways in the bubble charts; **(D)** The top 154 immune-related genes were significantly differentiated between the metastasis and non-metastasis group in the heatmap; **(E, F)** Immunohistochemical staining results of CD8, PD1, and PD-L1 in metastatic and non-metastatic NPC samples from Fujian Cancer Hospital. **P < 0.01, ***P < 0.001.

Assessment of the tumor immune microenvironment and immune checkpoints

Here, we estimated how the immune microenvironment differed between patients in high- and low-risk groups in terms of immune scores and levels of immune cell infiltration. The patients from the low-risk group had higher immune scores but lower tumor purity (Figures 5A, B). Additionally, the compositions of the 29 immune-cell types were significantly different in the high- and low-risk groups (Figure 5C). In the low-risk group, almost all levels of immune cell infiltration were higher than in the high one, including B cells, CD8 T cells, dendritic cells (DC), macrophages (Figure 5D). Moreover, there were significant statistical differences in the immune checkpoint inhibitors (CTLA-4, HAVCR2, SIGLEC15, TIGIT, PD1 and LAG3) between the high- and low-risk groups (Figure 5E).

Predictive power for immunotherapy efficacy

We were the first to evaluate GEP and TLS score, and showed that there were higher levels of immune cell receptors in low-risk

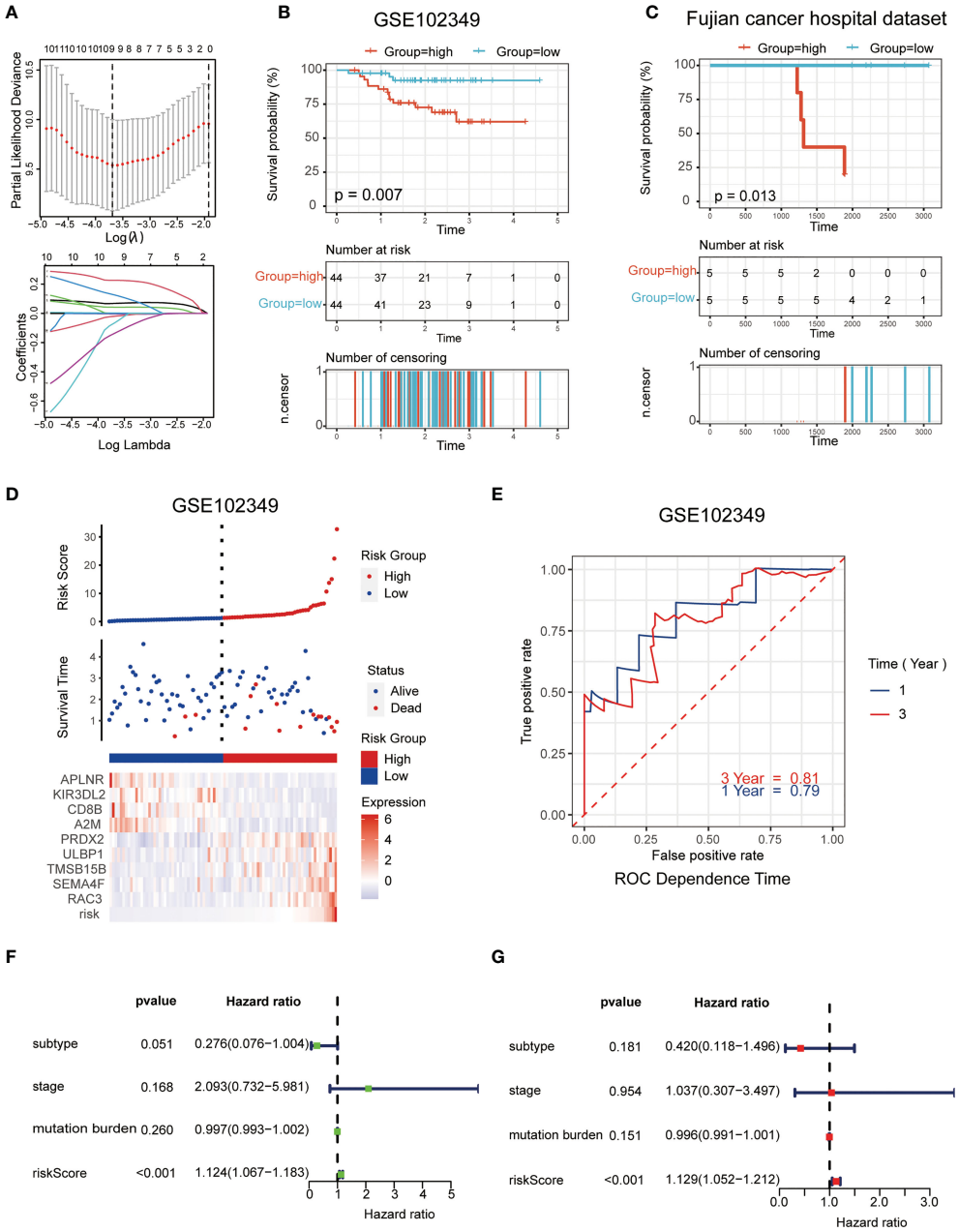


FIGURE 2
Establishment and validation of the immune-related risk model. **(A)** The LASSO logistic regression model was applied to establish prognostic immune biomarkers which involved 9 signatures (A2M, APLNR, CD8B, RAC3, PRDX2, ULBP1, TMSB15B, KIR3DL2 and SEMA4F) identified by the GEO dataset (GSE102349); **(B, C)** The Kaplan-Meier plot of the immune-related genes in GSE102349 **(B)** and Fujian Cancer Hospital corhort **(C)** revealed the statistical significance between the high- and low-risk groups; **(D)** Patient survival status and expression of 9 hub genes in high and low risk groups; **(E)** Receiver operating characteristic (ROC) curves of 1-year and 3-year survival in GSE102349; **(F, G)** Univariate **(F)** and multivariate **(G)** Cox regression analyses for the immune-related risk score model as an independent prognostic factor.

patients (Figures 6A, B). Subsequently, the same results were observed in numerous immune-related indicators (Figure 6C). These results suggested that tumors stimulate more immune cell activation and strong ligand-receptor activation in patients in the low-risk group, laying the biological foundation for a positive response to this immunotherapy. As Figure 6D–G showed, patients in the low-risk group had a higher immune response in a cohort of patients with whether nasopharyngeal carcinoma or melanoma, or non-small cell lung cancer. It was evident that the patients of the high-risk group had less chance of benefiting from immunotherapy, which represented a worse prognosis when compared with the patients of the low-risk group (Figure 6H).

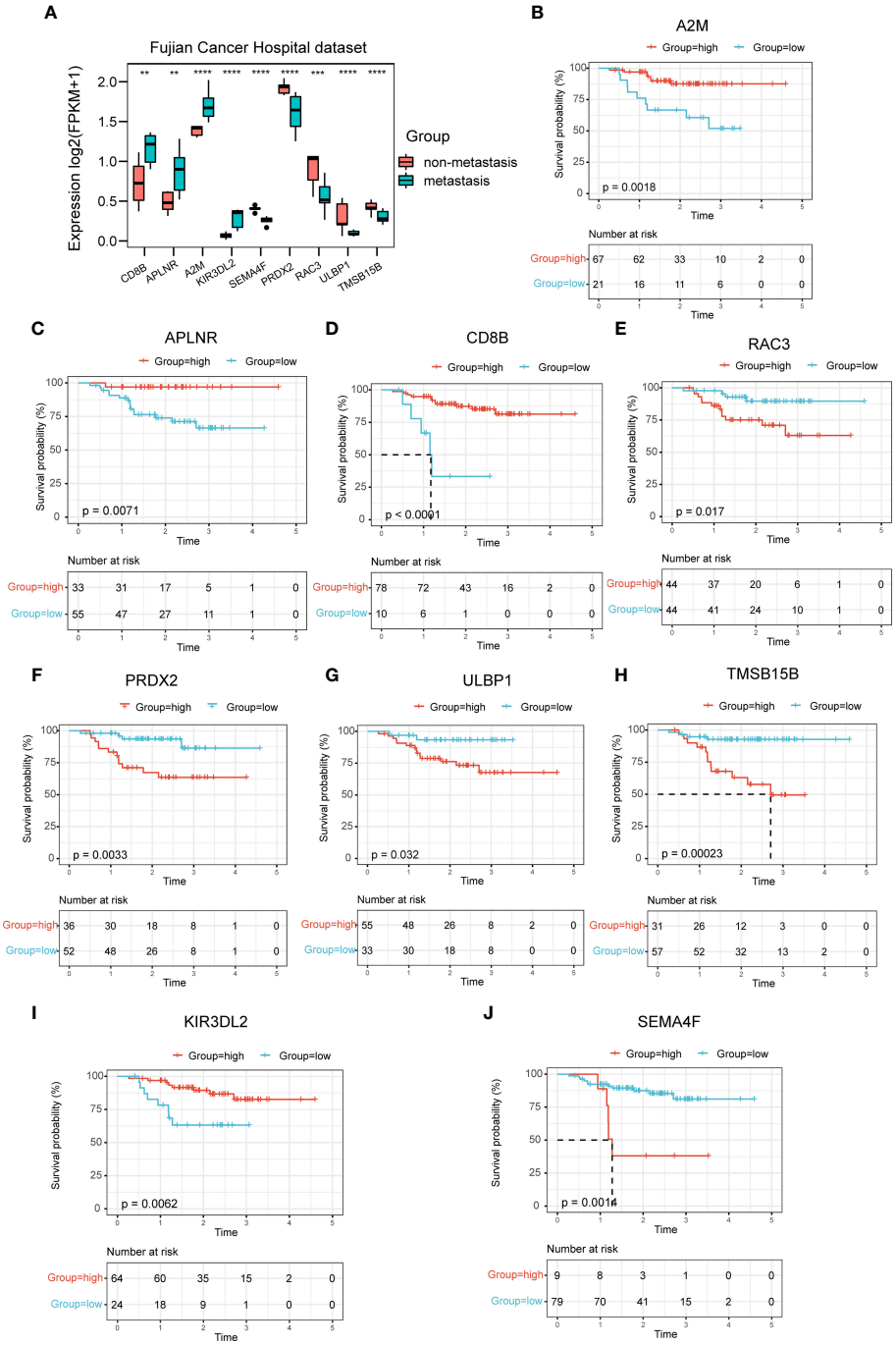


FIGURE 3
Expression profiles and prognostic potency of nine hub genes. **(A)** The 9 immune-related signatures were significantly different between the non-metastasis group and metastasis group of this hospital cohort. **(B–J)** A2M, APLNR, CD8B, RAC3, PRDX2, ULBP1, TMSB15B, KIR3DL2 and SEMA4F had extraordinary differences of survival probability between the high-risk and the low-risk groups in GSE102349. **P < 0.01, ***P < 0.001, ****P < 0.0001.

Immune landscapes and cellular communication at the single-cell level

A cluster of 29 distinct cell types in GSE150430 cohort was defined by two-dimensional spatial visualization of UMAP analysis (Figure 7A). Cell lineages were distributed to each cluster by gene expression with reference to the human primary cell atlas data in TISCH. As a result, cells were annotated (Figure 7B). We targeted the most significantly differentially expressed genes in each cluster to better understand the species of cell fascicles (Figure S3A). In the identified cell subsets, the GSVA and ssGSEA algorithm was employed to calculate the performance of the nine hub genes at the single-cell level. Significantly higher risk scores were observed in malignant cells than in B cells and CD8 T cells (Figures 7C, S3B). The same conclusion can be drawn in the cellular localization map (Figures S3C, D). Moreover, the percentage of B cells and CD8 T

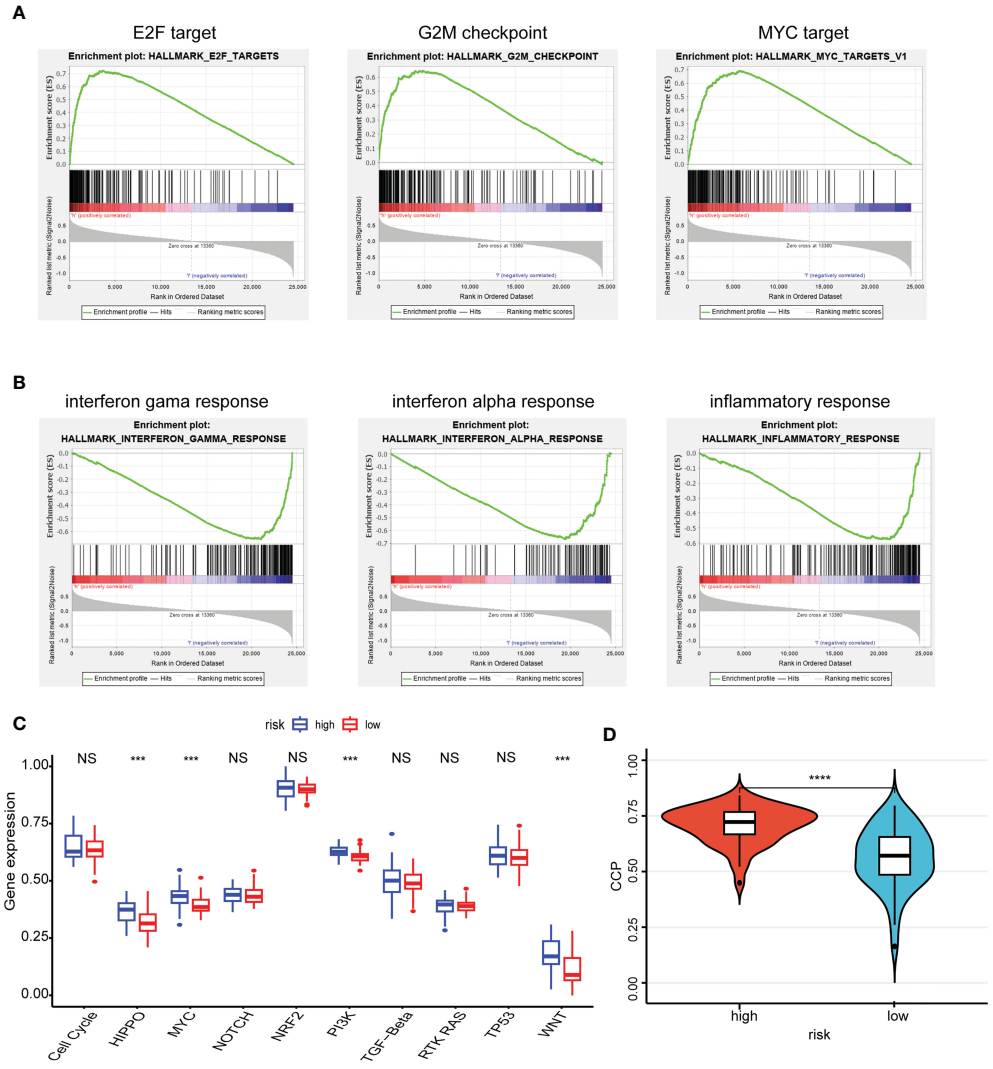


FIGURE 4
Enrichment pathways of hub genes and correlation with oncogenic pathways, proliferative activity. **(A, B)** The high-risk group **(A)** was mainly distributed in the E2F target, G2M checkpoint and MYC target using the GSEA method and the low-risk group **(B)** was mainly distributed in the INF-γ, INF-α and inflammatory responses using the GSEA method; **(C)** Patient samples from high- and low-risk groups showed significant differences in scores across the ten carcinogenic pathways; **(D)** Patients in the high-risk group having high CCP scores. ***P < 0.001, **** P < 0.0001, ^{ns}P > 0.05.

cells in the low-risk samples was notablely higher than that of the high ones; however, the percentage of malignant cells in the high-risk samples was significantly higher than that of the low ones (Figures 7D, E). This was consistent with previous findings indicating that the high-risk scores predicted worse biological behavior.

Next, we carried out functional exploration. The major pathways enriched for differential genes between high- and low-risk groups were those related to intercellular adhesion and immune cell activation, suggesting that the response to distant metastasis and immune resistance differed between high and low-risk groups (Figure 7F). Also, active pathways were observed to vary in the high- and low-risk groups, like LT, TGFb, SEMA3, KIT, FGF and CD70 pathways being active in the high group while CALCR, CD40, and SPP1 pathways being vibrant in the low group (Figures 7G, S3E). In Figures 7H, I, the distinction of CD70 and SPP1 signaling pathways in high- and low-risk groups was more intuitive. Finally, the intracellular expression of nine hub genes is exhibited (Figure 7J). It can be seen that the expression of PRDX2, TMSB15B, ULBP1, and RAC3 was specifically increased in malignant cells, and the high expression of these genes coincides with a worse survival prognosis (Figures 3E–H).

Discussion

In this study, we screened nine hub genes to construct an immune-related risk model from differently expressed genes of metastatic and non-metastatic NPC patients in Fujian Cancer Hospital. The model accurately predicted overall survival and was strongly associated with immune infiltration at both the transcriptome level and the single-cell level.

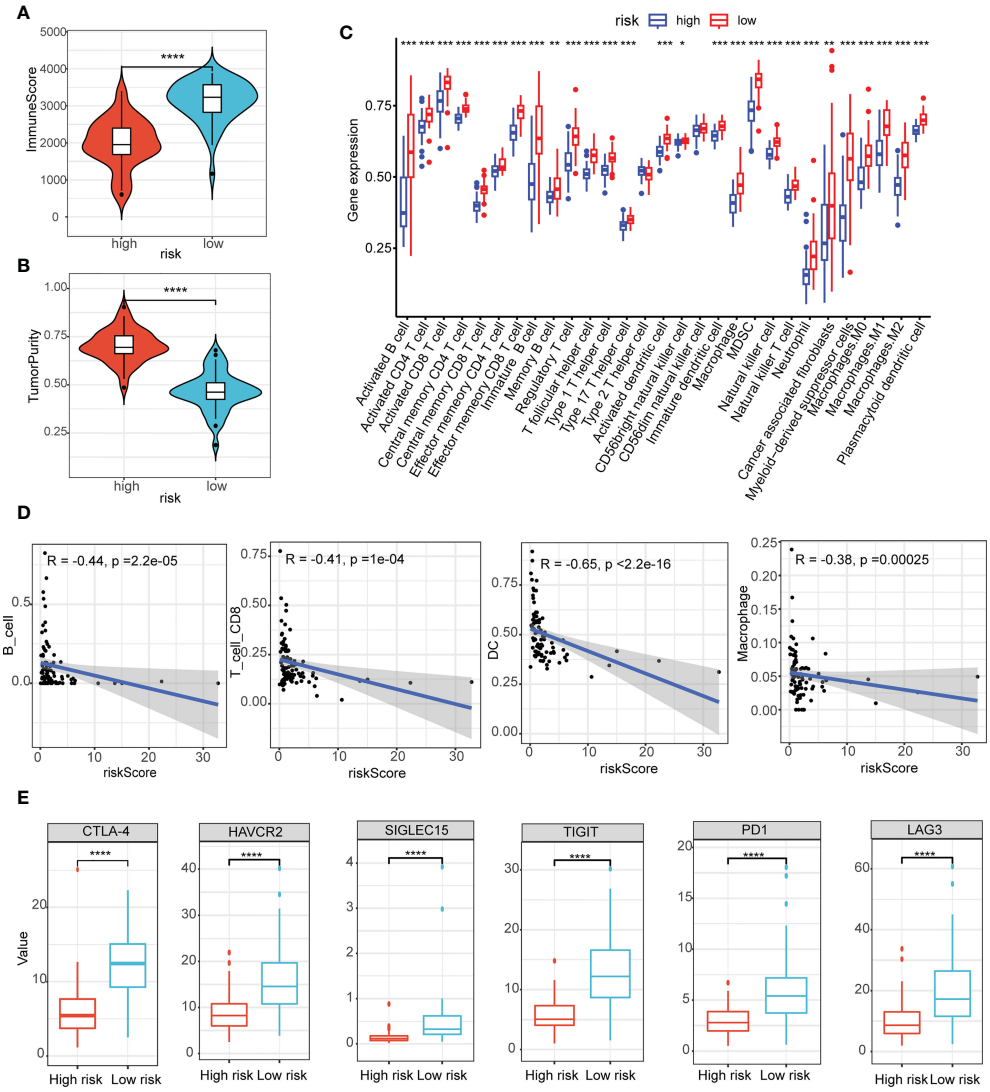


FIGURE 5
Assessment of the tumor immune microenvironment and immune checkpoints. **(A, B)** The immune scores **(A)** and scores of tumor purity **(B)** between the high- and low-risk group had notable statistical differences in the violin plot; **(C)** The compositions of the 29 immune-cell types were significantly different in the high- and low-risk groups; **(D)** B cells, CD8 T cells, dendritic cells (DC), macrophages infiltration were negatively related to the risk scores; **(E)** Immune checkpoint inhibitors (CTLA-4, HAVCR2, SIGLEC15, TIGIT, PD1 and LAG3) between the high- and low-risk groups had notable statistical differences in the box plots. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

In NPC, polygenic models for predicting prognosis based on gene expression levels have been rarely reported. More attention has focused on predicting prognosis at the miRNA level, single gene level. Prediction models are constructed by integrating various different factors, such as clinicopathological features, imaging features, genomic features, etc. A study identified a prognostic predictive risk model for patients with nasopharyngeal carcinoma based on three miRNA signatures (ebv-miR-BART19-3p, hsa-miR-135b, hsa-miR-141), which can be used to predict the overall survival of patients with nasopharyngeal carcinoma. (3-year ROC = 0.76) (39). In a CT-based and PET-based signatures for individual induction chemotherapy (IC) in advanced NPC, the researchers proposed a radiomics nomogram with a C-index of 0.754 [95% confidence interval (95% CI), 0.709-0.800] in the training set and 0.722 (95% CI, 0.652-0.792) in the test set (40). Another study investigated the prognostic significance of tumor-infiltrating immune cells and microenvironment-relevant genes in NPC (NPC) and their correlations. A risk score model composed of DARC, IL33, IGHG1, and SLC6A8 was established with a good performance for PFS prediction (AUC = 0.738) (41). In our study, one of the novelties is the construction of a predictive model for metastatic NPC, and the good predictive accuracy achieved. The area under the ROC curve (AUC) of our model was 0.79 at 1-year, and 0.81 at 3-years, respectively, indicating a high predictive value. We filled the research gap of genetic prognostic prediction model for metastatic NPC. The results of the study are expected to provide

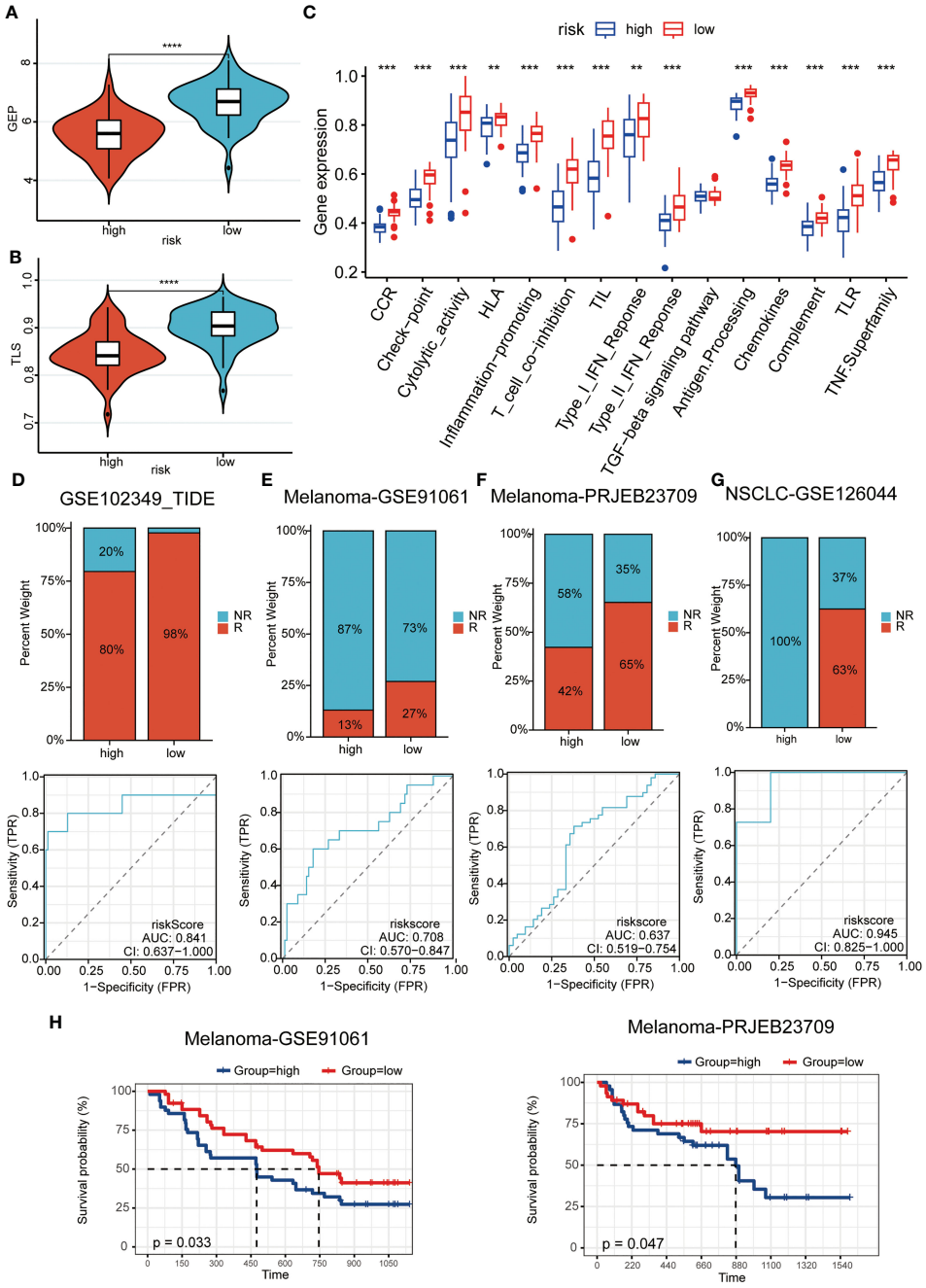


FIGURE 6
Predictive power for immunotherapy efficacy. **(A, B)** GEP **(A)** and TLS **(B)** score were higher in the low-risk group; **(C)** numerous immune-related indicators were over-expressed in low-risk patients; **(D–G)** Patients in the low-risk group had a higher immune response in a cohort of patients with whether nasopharyngeal carcinoma **(D)** or melanoma **(E-F)**, or non-small cell lung cancer **(G)**; **(H)** The high-risk patients had a worse prognosis when compared with the patients of the low-risk group in melanoma cohorts. **P < 0.01, ***P < 0.001, ****P < 0.0001.

a theoretical basis for accurate prognostic assessment of metastatic NPC.

Although improving the responsiveness of immunotherapy is very promising for the treatment of metastatic tumors, the effectiveness of strategies to improve the immune response to cancer varies from patient to patient, due to the heterogeneity of cancer cells and immune cells in TME, the crosstalk of biological signaling pathways, and the varying composition of specific immune cells (42). Our study proposes a robust risk prediction model based on metastatic NPC cases in Fujian Cancer Hospital, which can accurately predict the prognosis and immunotherapy efficacy of metastatic NPC patients.

Tumor-infiltrating lymphocytes determine the progression and aggressiveness of tumors and are a source of important

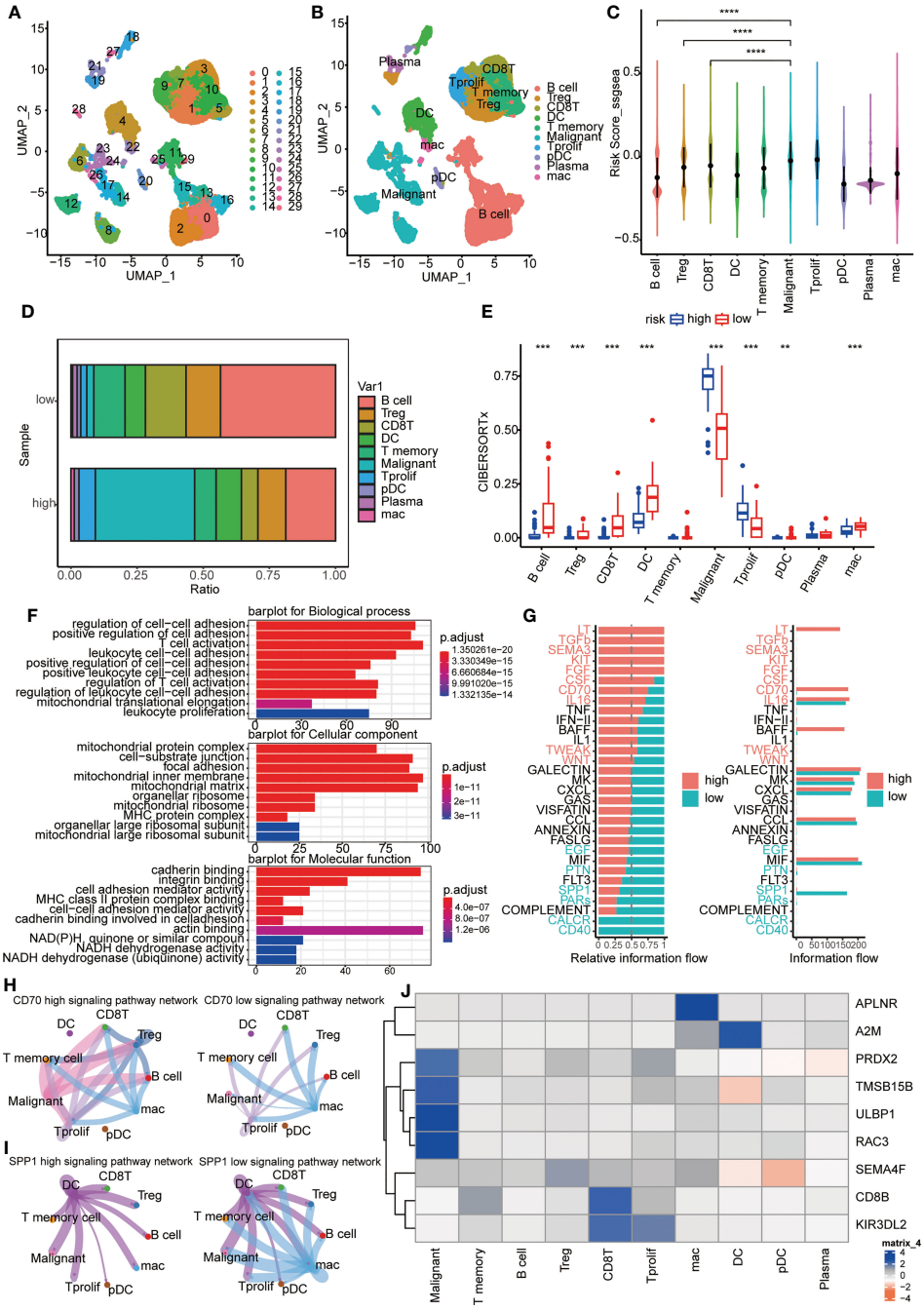


FIGURE 7
Immune landscapes and cellular communication at the single-cell level. **(A, B)** A cluster of 29 distinct cell types in GSE150430 cohort was defined by two-dimensional spatial visualization of UMAP analysis; **(C)** Risk scores for samples in different cell subsets; **(D)** The proportion of cell composition in high- and low-risk groups; **(E)** Immune cell infiltration in high- and low-risk groups using CIBERSORTx; **(F)** The major pathways enriched for differential genes between high- and low-risk groups; **(G)** Active pathways were observed to vary in the high- and low-risk groups; **(H, I)** CD70 and SPP1 signaling pathways in high- and low-risk groups; **(J)** The intracellular expression of nine hub genes. **P < 0.01, ***P < 0.001, ****P < 0.0001.

prognostic information for patients (43, 44). In this study, samples from the low-risk group had higher immune scores, lower scores of tumor purity, and higher value of immune checkpoint inhibitors simultaneously. It can be reasonably speculated that the patients from the low-risk group will benefit from immunotherapy as compared with patients from the high-risk group. The well-established prognostic model could make an obvious distinction of the patients with metastatic NPC to predict the risk of poor prognosis. For the advanced patients assigned to the low-risk group, the combination of chemoradiotherapy and immunotherapy would be an appropriate choice to attempt a better outcome.

The pro-oncogenic pathways, including E2F, G2M checkpoint and MYC targets pathway, favor tumor cells to promote growth, migration, invasion, and angiogenesis. In our analysis, GESA identified the enrichment of E2F, G2M checkpoint, and MYC targets pathway in the high-risk group, which may contribute to the dismal prognosis. On the contrary, inflammatory response contributes to cancer cell death by inducing an anti-tumor immune response and therefore accounts for a favor prognosis of low-risk group.

Recently, the prediction and evaluation of the efficacy and outcome after immunotherapy for a specific tumor or the patient with a specific tumor is a hot spot in the development of contemporary medical treatment. Tumors and their microenvironments constantly interact with each other (45). According to the type and number of infiltrated immune cells, it can be divided into hot tumors and cold tumors. Hot tumors refer to tumors that have triggered the body's immune responses with a certain number of immune cell infiltration, which tend to respond well to immune checkpoint inhibitors. While cold tumors are considered as those with few immune cells where it is difficult to stimulate the autoimmune responses and where immune checkpoint inhibitors could not play an effective role when compared with hot tumors. In this study, the risk model we constructed can predict immune cell infiltration in patients and even infer specific immune cell content levels in both transcriptome level and single-cell level. In addition, patients in the high-risk group had a large proportion of malignant cells in their cellular composition, whereas immune cells in the low-risk group had a large proportion. There was also a dramatic difference in the ligand receptors for cellular communication between the high- and low-risk patients. The low-risk group or the non-metastasis group had high immune scores and abundant immune cell infiltration, which means that they have a hot tumor component and superior immune response in their bodies, indicating a higher likelihood of benefiting from immunotherapy and a better prognosis. Therefore, accurate prediction of our model holds great value for individualized treatment and efficacy detection in clinical settings for advanced NPC patients.

Immunotherapy drugs targeting PD-L1 and CTLA-4 are playing an increasingly critical role in the treatment of malignant tumors (46). The expression levels of PD-L1 or other immune checkpoints will directly affect the therapeutic effect of immune checkpoint inhibitors, by which the application of immune checkpoint inhibitors can be guided. TLS is an ectopic lymphoid-like structure that is mostly formed in tissues where inflammation occurs (47). In recent years, many studies have revealed that tumor-infiltrating B lymphocytes (48) and tumor-associated TLS have a non-negligible correlation with the response to immune-checkpoint blockade treatment, which provides a new biological marker for the clinical decision-making of immunotherapy. In this study, there are a higher number of B memory lymphocytes and increased immune checkpoint expression in the low-risk group, which insinuates more opportunity to benefit from immunotherapy. The accuracy of the risk model predictions was likewise validated in multiple immunotherapy cohorts.

To the best of our knowledge, our study presented the first immunopredictive risk model for metastatic NPC based on realistic cases. However, our study had some limitations. A major limitation was the lack of a prospective NPC cohort to validate the prognostic role and stratification performance of the model. In addition, the role of predicting immunotherapy efficacy in real-world settings needs further investigation.

Conclusions

We have constructed a model based on immunity of metastatic NPC and determined its prognostic value. In addition, the model identified cell-cell communication between tumor and immunity, along with potential therapeutic approaches to target metastatic NPC.

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

All included patients gave their written informed consent. The study was approved by the Ethics Committee of Fujian Cancer Hospital and Fujian Medical University Cancer Hospital (approval number SQ2019-035-01). The patients/participants provided their written informed consent to participate in this study.

Author contributions

XCC and QD designed the article. QD, TL and YS organized the public data and wrote the manuscript. XC, TL, WH and XC took charge for data visualization. YS, ZH and YL obtained the clinical information. DW and SQ contributed to the concept and revised the article. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Fujian Provincial Clinical Research Center for Cancer Radiotherapy and Immunotherapy (2020Y2012); Supported by the National Clinical Key Specialty Construction Program (2021); Fujian Clinical Research Center for Radiation and Therapy of Digestive, Respiratory and Genitourinary Malignancies; United Fujian Provincial Health and Education Project for Tackling the Key Research, China (2019-WJ-03); National Natural Science Foundation of China (11974077);

National Natural Science Foundation of China (82072986); Major Research Projects for Young and Middle-aged Researchers of Fujian Provincial Health Commission (2021ZQNZD010); Science and Technology Pilot Program of Fujian Province, China (2021Y0053); Innovative Medicine Subject of Fujian Provincial Health Commission, China (2021CXA029); Wu Jieping Medical Foundation (320.6750.2021-01-27); Joint Funds for the Innovation of Science and Technology, Fujian province (Grant number: 2021Y9196); and High-level Talent Training Program of Fujian Cancer Hospital.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1109503/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The schematic diagram of this study.

SUPPLEMENTARY FIGURE 2
(A) PCA cluster analysis of the metastatic group (group M) and the non-metastatic group (group N); **(B, C)** The risk score was significantly distinguished between **(B)** clinical stage I–III and stage IV in GSE102349 (n=73) as well as **(C)** metastatic and the non-metastatic patients in Fujian Cancer Hospital dataset (n=10).

SUPPLEMENTARY FIGURE 3
(A) The cell annotation and the specially expressed genes in each cluster; **(B)** Risk scores for samples in different cell subsets; **(C, D)** Comparison of cellular composition of high and low risk groups; **(E)** The signaling pathway of high and low risk groups in comparison.

References

1. Chen Y-P, Chan ATC, Le Q-T, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet* (2019) 394(10192):64–80. doi: 10.1016/S0140-6736(19)30956-0

2. Chua MLK, Wee JTS, Hui EP, Chan ATC. Nasopharyngeal carcinoma. *Lancet* (2016) 387(10022):1012–24. doi: 10.1016/S0140-6736(15)00055-0

3. Wong KCW, Hui EP, Lo KW, Lam WKJ, Johnson D, Li L, et al. Nasopharyngeal carcinoma: An evolving paradigm. *Nat Rev Clin Oncol* (2021) 18(11):679–95. doi: 10.1038/s41571-021-00524-x

4. Lin W, Chen L, Zhang H, Qiu X, Huang Q, Wan F, et al. Tumor-intrinsic YTHDF1 drives immune evasion and resistance to immune checkpoint inhibitors *via* promoting MHC-I degradation. *Nat Commun* (2023) 14(1):265. doi: 10.1038/s41467-022-35710-7

5. Morad G, Helmink BA, Sharma P, Wargo JA. Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* (2021) 184(21):5309–37. doi: 10.1016/j.cell.2021.09.020

6. Burtness B, Harrington KJ, Greil R, Soulieres D, Tahara M, de Castro GJr., et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet* (2019) 394(10212):1915–28. doi: 10.1016/S0140-6736(19)32591-7

7. Wu J, Li L, Zhang H, Zhao Y, Zhang H, Wu S, et al. A risk model developed based on tumor microenvironment predicts overall survival and associates with tumor immunity of patients with lung adenocarcinoma. *Oncogene* (2021) 40(26):4413–24. doi: 10.1038/s41388-021-01853-y

8. Yuan H, Liu J, Zhao L, Wu P, Chen G, Chen Q, et al. Prognostic risk model and tumor immune environment modulation of m5C-related LncRNAs in pancreatic ductal adenocarcinoma. *Front Immunol* (2021) 12:800268. doi: 10.3389/fimmu.2021.800268

9. Gu W, Mo S, Wang Y, Kawabata-Iwakawa R, Zhang W, Yang Z, et al. Robust validation and comprehensive analysis of a novel signature derived from crucial metabolic pathways of pancreatic ductal adenocarcinoma. *Cancers (Basel)* (2022) 14 (7), 1825. doi: 10.3390/cancers14071825

10. Liu Y, Wu J, Huang W, Weng S, Wang B, Chen Y, et al. Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer. *J Transl Med* (2020) 18(1):201. doi: 10.1186/s12967-020-02366-0

11. Luo C, Lei M, Zhang Y, Zhang Q, Li L, Lian J, et al. Systematic construction and validation of an immune prognostic model for lung adenocarcinoma. *J Cell Mol Med* (2020) 24(2):1233–44. doi: 10.1111/jcmm.14719

12. Wang D, Wei G, Ma J, Cheng S, Jia L, Song X, et al. Identification of the prognostic value of ferroptosis-related gene signature in breast cancer patients. *BMC Cancer* (2021) 21(1):645. doi: 10.1186/s12885-021-08341-2

13. Song J, Xu Q, Zhang H, Yin X, Zhu C, Zhao K, et al. Five key lncRNAs considered as prognostic targets for predicting pancreatic ductal adenocarcinoma. *J Cell Biochem* (2018) 119(6):4559–69. doi: 10.1002/jcb.26598

14. Zhao Z, He B, Cai Q, Zhang P, Peng X, Zhang Y, et al. Combination of tumor mutation burden and immune infiltrates for the prognosis of lung adenocarcinoma. *Int Immunopharmacol* (2021) 98:107807. doi: 10.1016/j.intimp.2021.107807

15. Ding Q, Chen X, Hong W, Wang L, Liu W, Cai S, et al. The prognostic role of cuproptosis in head and neck squamous cell carcinoma patients: A comprehensive analysis. *Dis Markers* (2022) 2022:9996946. doi: 10.1155/2022/9996946

16. Yang M, Zheng H, Xu K, Yuan Q, Aihaiti Y , Cai Y, et al. A novel signature to guide osteosarcoma prognosis and immune microenvironment: Cuproptosis-related lncRNA. *Front Immunol* (2022) 13:919231. doi: 10.3389/fimmu.2022.919231

17. Lu J, Chen Y, Zhang X, Guo J, Xu K, Li L. A novel prognostic model based on single-cell RNA sequencing data for hepatocellular carcinoma. *Cancer Cell Int* (2022) 22(1):38. doi: 10.1186/s12935-022-02469-2

18. Liu X, Li J, Wang Q, Bai L, Xing J , Hu X, et al. Analysis on heterogeneity of hepatocellular carcinoma immune cells and a molecular risk model by integration of scRNA-seq and bulk RNA-seq. *Front Immunol* (2022) 13:1012303. doi: 10.3389/fimmu.2022.1012303

19. Huang D, Liu Q, Zhang W, Huang C, Zheng R, Xie G, et al. Identified IGSF9 association with prognosis and hypoxia in nasopharyngeal carcinoma by bioinformatics analysis. *Cancer Cell Int* (2020) 20:498. doi: 10.1186/s12935-020-01587-z

20. Bao YN, Cao X, Luo DH, Sun R, Peng LX, Wang L, et al. Urokinase-type plasminogen activator receptor signaling is critical in nasopharyngeal carcinoma cell growth and metastasis. *Cell Cycle* (2014) 13(12):1958–69. doi: 10.4161/cc.28921

21. Singer J, Irmisch A, Ruscheweyh HJ, Singer F, Toussaint NC, Levesque MP, et al. Bioinformatics for precision oncology. *Brief Bioinform* (2019) 20(3):778–88. doi: 10.1093/bib/bbx143

22. Consortium TGO. Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res* (2017) 45(D1):D331–8. doi: 10.1093/nar/gkw1108

23. The Gene Ontology C. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* (2019) 47(D1):D330–8. doi: 10.1093/nar/gky1055

24. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* (2017) 45 (D1):D353–61. doi: 10.1093/nar/gkw1092

25. Vidyasagar M. Identifying predictive features in drug response using machine learning: Opportunities and challenges. *Annu Rev Pharmacol Toxicol* (2015) 55:15–34. doi: 10.1146/annurev-pharmtox-010814-124502

26. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* (2020) 48(W1):W509–14. doi: 10.1093/nar/gkaa407

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

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

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

30. Ott PA, Bang Y-J, Piha-Paul SA, Razak ARA, Bennouna J, Soria J-C, et al. T-Cell-Inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J Clin Oncol* (2019) 37(4):318–27. doi: 10.1200/JCO.2018.78.2276

31. Bergomas F, Grizzi F, Doni A, Pesce S, Laghi L, Allavena P, et al. Tertiary intratumor lymphoid tissue in colo-rectal cancer. *Cancers (Basel)* (2011) 4(1):1–10. doi: 10.3390/cancers4010001

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

33. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17(1):218. doi: 10.1186/s13059-016-1070-5

34. Lin W, Xu L, Chen Y, Go K, Sun C, Zhu Y. Single-cell transcriptomics-based study of transcriptional regulatory features in the mouse brain vasculature. *BioMed Res Int* (2021) 2021:7643209. doi: 10.1155/2021/7643209

35. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* (2019) 20(2):163–72. doi: 10.1038/s41590-018-0276-y

36. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7), 1888–1902.e21. doi: 10.1016/j.cell.2019.05.031

37. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* (2019) 37(7):773–82. doi: 10.1038/s41587-019-0114-2

38. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* (2021) 12(1):1088. doi: 10.1038/s41467-021-21246-9

39. Zhou J, Zhang B, Zhang X, Wang C , Xu Y. Identification of a 3-miRNA signature associated with the prediction of prognosis in nasopharyngeal carcinoma. *Front Oncol* (2021) 11:823603. doi: 10.3389/fonc.2021.823603

40. Peng H, Dong D, Fang MJ, Li L, Tang LL, Chen L, et al. Prognostic value of deep learning PET/CT-based radiomics: Potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res* (2019) 25 (14):4271–9. doi: 10.1158/1078-0432.CCR-18-3065

41. Zou Z, Ha Y, Liu S, Huang B. Identification of tumor-infiltrating immune cells and microenvironment-relevant genes in nasopharyngeal carcinoma based on gene expression profiling. *Life Sci* (2020) 263:118620. doi: 10.1016/j.lfs.2020.118620

42. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Sci (New York NY)* (2018) 359(6377):801–6. doi: 10.1126/science.aan5951

43. Sidaway P. TILs show early efficacy. *Nat Rev Clin Oncol* (2021) 18(10):603. doi: 10.1038/s41571-021-00555-4

44. Kristensen VN. The antigenicity of the tumor cell - context matters. *N Engl J Med* (2017) 376(5):491–3. doi: 10.1056/NEJMcibr1613793

45. Goliwas KF, Deshane JS, Elmets CA, Athar M. Moving immune therapy forward targeting TME. *Physiol Rev* (2021) 101(2):417–25. doi: 10.1152/physrev.00008.2020

46. Englinger B, Pirker C, Heffeter P, Terenzi A, Kowol CR, Keppler BK, et al. Metal drugs and the anticancer immune response. *Chem Rev* (2019) 119(2):1519–624. doi: 10.1021/acs.chemrev.8b00396

47. Sautes-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat Rev Cancer* (2019) 19(6):307–25. doi: 10.1038/s41568-019-0144-6

48. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* (2020) 577 (7791):549–55. doi: 10.1038/s41586-019-1922-8

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 11 April 2023
DOI 10.3389/fimmu.2023.1014378

OPEN ACCESS

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

REVIEWED BY
Kamariah Ibrahim,
University of Malaya, Malaysia
Yanbing Yu,
China-Japan Friendship Hospital, China
Qianqian Song,
Wake Forest University, United States

*CORRESPONDENCE
Zhiwei Xia
xiazhiwei2011@gmail.com
Li Peng
mj.pengli@csu.edu.cn

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

RECEIVED 08 August 2022
ACCEPTED 13 March 2023
PUBLISHED 11 April 2023

CITATION
Lei Q, Yuan B, Liu K, Peng L and Xia Z (2023) A novel prognostic related lncRNA signature associated with amino acid metabolism in glioma. *Front. Immunol.* 14:1014378. doi: 10.3389/fimmu.2023.1014378

A novel prognostic related lncRNA signature associated with amino acid metabolism in glioma

Qiang Lei[1], Bo Yuan[2], Kun Liu[2], Li Peng [3,4]* and Zhiwei Xia [5]*

[1]Department of Neurology, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]Department of Cerebrovascular Surgery, The Second People's Hospital of Hunan Province, The Hospital of Hunan University of Chinese Medicine, Changsha, Hunan, China, [3]Department of Ophthalmology, Central South University Xiangya School of Medicine Affiliated Haikou Hospital, Haikou, Hainan, China, [4]Department of Ophthalmology, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China, [5]Department of Neurology, Hunan Aerospace Hospital, Changsha, Hunan, China

Background: Glioma is one of the deadliest malignant brain tumors in adults, which is highly invasive and has a poor prognosis, and long non-coding RNAs (lncRNAs) have key roles in the progression of glioma. Amino acid metabolism reprogramming is an emerging hallmark in cancer. However, the diverse amino acid metabolism programs and prognostic value remain unclear during glioma progression. Thus, we aim to find potential amino-related prognostic glioma hub genes, elaborate and verify their functions, and explore further their impact on glioma.

Methods: Glioblastoma (GBM) and low-grade glioma (LGG) patients' data were downloaded from TCGA and CCGA datasets. LncRNAs associated with amino acid metabolism were discriminated against *via* correlation analysis. LASSO analysis and Cox regression analysis were conducted to identify lncRNAs related to prognosis. GSVA and GSEA were performed to predict the potential biological functions of lncRNA. Somatic mutation data and CNV data were further built to demonstrate genomic alterations and the correlation between risk scores. Human glioma cell lines U251 and U87-MG were used for further validation *in vitro* experiments.

Results: There were eight amino-related lncRNAs in total with a high prognostic value that were identified *via* Cox regression and LASSO regression analyses. The high risk-score group presented a significantly poorer prognosis compared with the low risk-score group, with more clinicopathological features and characteristic genomic aberrations. Our results provided new insights into biological functions in the above signature lncRNAs, which participate in the amino acid metabolism of glioma. LINC01561 is one of the eight identified lncRNAs, which was adopted for further verification. In *in vitro* experiments, siRNA-mediated LINC01561 silencing suppresses glioma cells' viability, migration, and proliferation.

Conclusion: Novel amino-related lncRNAs associated with the survival of glioma patients were identified, and a lncRNA signature can predict glioma prognosis and therapy response, which possibly has vital roles in glioma. Meanwhile, it emphasized the importance of amino acid metabolism in glioma, particularly in providing deeper research at the molecular level.

KEYWORDS

glioma, lncRNA, amino acid, prognostic, target

1 Introduction

Gliomas are among the most severe and common primary human brain malignancies with a dismal prognosis and a very low 5-year survival rate, which are characterized by an immunosuppressive microenvironment (1–3). Currently, the clinical treatment includes surgery, chemotherapy, radiotherapy, targeted therapy, and immunotherapy in gliomas (4, 5). However, patients with gliomas remain to have a poor prognosis, especially in high-grade gliomas (6). Hence, further exploring the mechanism of glioma is of great significance for finding new targets for the treatment of glioma.

The microenvironment is recognized as playing a vital role in the development and progression of tumors (7, 8). In recent years, it has been recognized as an emerging hallmark of cancer (9, 10), which plays a major role in cancer occurrence and development. Metabolism reprogramming has been recognized as having a critical role in both cancer progression and effective immune responses in the tumor microenvironment (11). Metabolism must be altered to meet the rapid biosynthetic demands for growing tumors in metabolism proliferating cancer (12). As with sugar metabolism, amino acid metabolism is also an ordered process, which is important for the maintenance of cellular homeostasis (13). Amino acid transport is a vital aspect of amino acid metabolism. Amino acids and their metabolites play a critical role in metabolism and are essential to life (14). Cancer cells need large amounts of energy and compounds to meet the requirements of metabolic reprogramming. Amino acid metabolism affects the prognosis of tumor (15). The uptake and metabolism of amino acids are increased in cancer cells *via* upregulation of specific amino acid transporters (16). Amino acids promote the proliferation and survival of cancer cells in the context of genotoxic, oxidative damage, nutritional change, and stress (11, 16). Currently, a new focus on the amino acid metabolism of glioma is considered to play a major role in glioma. Therefore, better knowledge of amino acid metabolism in terms of cancer is strongly warranted.

Long non-coding RNA (lncRNA) has been reported to be involved in multiple physiological and pathological processes and has shown an essential role among them (17). Glioblastoma cells increase under conditions of hypoxic stress (18, 19). Amino acid metabolism-related risk signatures can be used to predict the prognosis for glioma (19). In addition, studies have demonstrated that lncRNA has shown a crucial role in the proliferation, progression, invasion, and prognosis of gliomas (20–22). Therefore, research about new lncRNAs as novel biomarkers will provide deeper insights into the progression and prognosis of glioma with the purpose of strengthening the management of the disease.

Previous studies have indicated that amino-related lncRNAs are involved in the prognosis of malignant tumor (23, 24). However, the role of amino acid-related gene sets remains unclear in glioma. Herein, we collected data from The Cancer Genome Atlas (TCGA) and Chinese Glioma Genome Atlas (CGGA) databases by using multiple algorithms to investigate the prognostic value of amino-related lncRNAs and correlation in the microenvironment of glioma.

2 Methods

2.1 Data resources

Human gene expression profiles and corresponding clinical information on glioma were downloaded from TCGA dataset (http://cancergenome.nih.gov) and the CGGA dataset (http://www.cgga.org.cn/). TCGA dataset was regarded as the training set, the CGGA dataset as the validation set. Log2(tpm+0.001) was used to normalize the data of gene expression. We collected 672 samples from TCGA dataset and 322 samples from the CGGA dataset. Somatic mutation and copy number variation (CNV) data were obtained from the dataset of TCGA.

2.2 Screening for prognostic lncRNAs associated with amino acids

To extract the amino-related lncRNA, gene set variation analysis (GSVA) was implemented using the GSVA R package. The amino-related gene sets were extracted from the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/i). A correlation analysis was conducted using the limma package of R statistical software. Amino-related lncRNAs were characterized *via* correlation analysis on the basis of Gene Ontology (GO) information of the target amino-related gene sets. Then, a univariate Cox regression analysis was performed to analyze all the lncRNAs related to the glioma patients' overall survival, and a multivariable Cox regression analysis was used to further select and identify lncRNAs that exhibited independent prognostic value. LASSO regression analysis was used in the construction of the lncRNAs with prognostic gene signatures. Then, we obtained a set of prognostic lncRNAs and regression coefficients (β) (Figure 1).

2.3 LncRNAs used for consensus clustering

Consensus clustering was performed to identify glioma subgroups (GBM and LGG) and cluster amino-related genes from TCGA dataset using the ConsensusClusterPlus R package. Permuting clustering runs by changing the category number k = 2 to k = 10. Based on a relatively high consistency within the clusters, cumulative distribution function (CDF) and a relatively small incremental change in area under the CDF curve were used to determine the optimal number of clusters k. Survival analyses were conducted with R statistical software and packages.

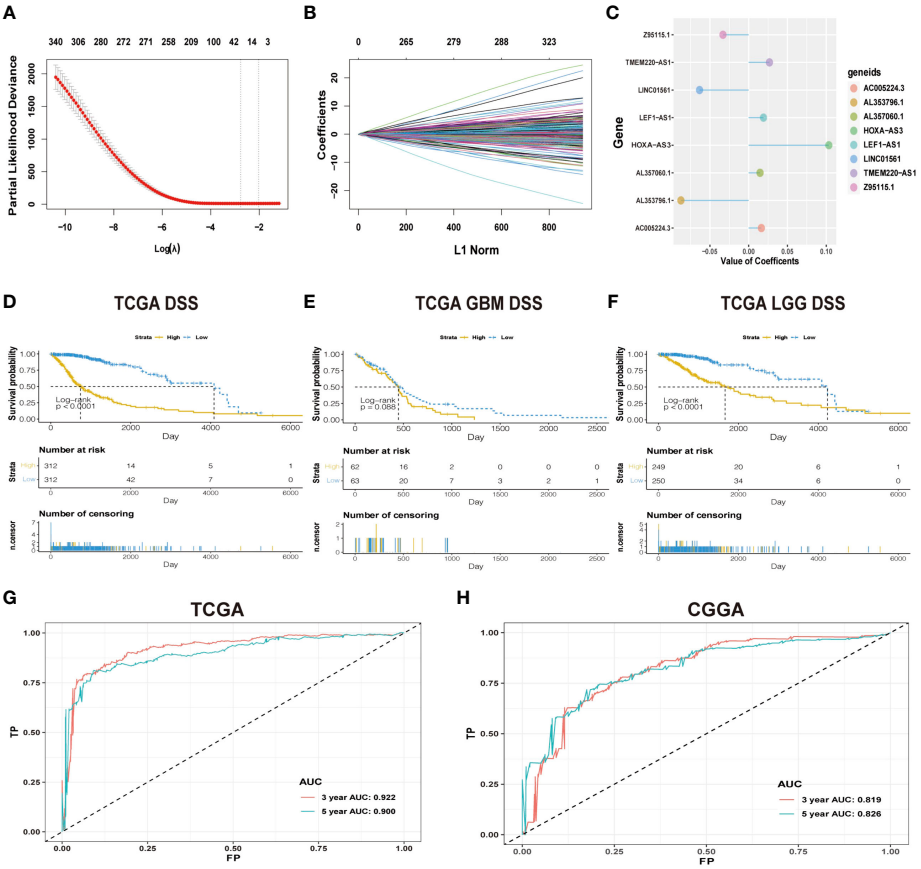


FIGURE 1
(A) Cross-validation in LASSO regression; dashed lines indicate the best-fit log (λ) value. **(B)** LASSO coefficients of the lncRNAs that have independent prognostic value. **(C)** Coefficient profie of the eight signature lncRNAs. **(D-F)** Difference in K-M curves of the DSS between high- and low-risk groups in the dataset of TCGA. **(G)** The ROC for predicting the survival at 3 and 5 years in TCGA dataset. The AUC of the risk score model showed good accuracy in the dataset of TCGA. **(H)** ROC for predicting the survival at 3 and 5 years in the dataset of CGGA.

2.4 Genomic alteration cluster

Somatic mutation analysis and copy number variation (CNV) were utilized to determine whether the risk score levels were related to specific genomic characteristics in gliomas *via* TCGA dataset. The somatic mutation data were analyzed with the "maftools" R package, and the genes with the most frequent somatic mutations were screened and presented. The enrichment of genomic events was determined by GSITIC analysis through an online analysis platform (https://www.genepattern.org).

2.5 Constructing a prognostic risk score model based on the clinical features and risk score

Univariate survival analyses were carried out by using the Cox proportional hazards regression package for the risk score and clinical features (age, gender, risk, IDH status, grade, 1p/19q) with a cutoff of p-value less than 0.05. Then, the multivariate Cox regression model was built based on the selected features, and visualization was achieved through a nomogram chart with the Regplot package. The area under the curve (AUC) value of the receiver operating characteristic (ROC) curve was established as an effective risk model to evaluate the prognosis of high-risk and low-risk patients. Both the calibration curve and AUC were carried out to evaluate the risk model.

2.6 Gene set variation analysis

Gene set variation analysis (GSVA) was performed using the ClusterProfiler package of R to calculate the enrichment analysis within TCGA and CGGA samples. The correlation between the risk score and GO terms was performed based on the significant GO terms of biological processes identified ($p < 0.05$), and a high correlation coefficient was selected.

2.7 Predicting the response to immunotherapy

The TIDE algorithm (http://tide.dfci.harvard.edu/) and the submap algorithm on the GenePattern website (https://cloud.genepattern.org/gp) are used to predict the likelihood of

response to immune checkpoint blockade for individual samples and subtypes.

2.8 Cell culture, treatments, and siRNA transfection

Human glioma cells (U251 and U87) were purchased from Procell Life Science & Technology Company (Hubei, China). The logarithmic growth phase cells were grouped into the following groups: control group, siRNA-negative control (NC) group, and LINC01561-siRNA group. Cells were treated with 5 μl siRNA, and Lipofectamine 2000 was diluted in Opti-MEM medium for 5 min. They were then mixed and incubated at room temperature for 20 min. The composite was then added to the cell culture plate. Following transfection for 48 h, the cells were collected for the subsequent experiments.

2.9 RT-qPCR assay

RT-qPCR assay was performed. The primers of β-actin (F: ACCCTGAAGTACCCCATCGAG, R: AGCACAGCCTGGATAGCAAC) and LINC01561 (F:CCAGGAGGAGCAGAGAAAGC, R: CCCAGCTGCTGTCTGGTTTA) were designed using Primer Premier 5.0. The total RNAs were extracted and then reversely transcribed into cDNA by HiScript Q RT SuperMix for RT-qPCR. The expression levels of β-actin and LINC01561 were quantified and calculated with the method of 2-ΔΔCT. The reaction conditions were as follows: 95°C for 10 min, then 95°C for 15 s, and 60°C for 30 s, for a total of 40 cycles.

2.10 Validation *in vitro* cell experiments

Cell viability was carried out by using a cell counting kit-8 (CCK-8) assay. Cell migration was measured by the Transwell migration assay.

2.11 Colony formation assay

Colony forming ability was assessed by a colony formation assay. Cell proliferation was evaluated and measured by using the 5-ethynyl-2′-deoxyuridine (EdU) assay kit (RiboBio, China). Each group has three biological replicates. Detailed vitro experimental protocol can be found in the supplemental material (Table S1).

2.12 Statistical analysis

Statistical calculations were performed using R statistical analysis package (version 3.5.3). One-way ANOVA followed by the Tukey posttest was used to identify the differences among groups, respectively. Correlations between categorical variables were assessed using chi-square tests. Overall survival analysis was evaluated using the Kaplan–Meier method, followed by Cox regression analysis. The ClusterProfiler package was performed to measure the enrichment analysis in TCGA and CGGA samples. Both the somatic mutations and CNA data were obtained through TCGA database. A comparison was accepted to be statistically significant when a p value was <0.05.

3 Results

3.1 Identification of prognostic lncRNAs related to amino acids

The workflow diagram of this study is illustrated in Figure 2. In total, in TCGA and CCGA datasets, 13,895 lncRNAs were extracted *via* intersecting the lncRNAs. Univariate and multivariate Cox regression analyses were carried out to find information on the lncRNA expression level of patients between the lncRNA and overall survival (OS) *via* the survival R package. There were 417 and 396 lncRNAs with significant prognostic potential screened out by univariate and multivariate Cox regression analysis ($p < 0.05$), respectively. Meanwhile, LASSO regression is an effective method for high-dimensional data and predictors. Then, in both TCGA and CCGA datasets, there were eight considered amino-related lncRNAs regarded with independent prognostic values identified *via* LASSO regression (Figures 1A, B), and the coefficient profile of the eight signature lncRNAs is illustrated in Figure 1C. They were AL357060.1, HOXA-AS3, LINC01561, Z95115.1, AL353796.1, LEF1-AS1, AC005224.3, and TMEM220-AS1. Then, a prognostic amino-related lncRNA signature was built (Figures 1D–F). The sensitivity and specificity of the eight-lncRNA prognostic model were determined by measuring the area under the receiver operating characteristic (AUC) and ROC between 3-year survival and 5-year survival. The AUC was 0.922 at 3-year survival and 0.900 at 5-year survival in TCGA dataset. The AUC was 0.819 at 3-year survival and 0.826 at 5-year survival in the CGGA dataset (Figures 1G, H). It can be seen that the AUC of the ROC and risk scoring models predicting the survival rates at 3 and 5 years show good accuracy in the two datasets.

3.2 Consensus cluster analyses of the 8 lncRNA expression

Based on the analysis of TCGA data, genes related to prognosis were identified. Consensus clustering analysis was applied for grouping the samples of glioma patients. With the expression similarity based on the dataset of TCGA, clustering stability rises from k = 2 to k = 10. The incremental change is small in the area under the CDF curve when k = 2 (Figures 3A, B). Hence, k = 2 was considered an optimal choice. Consensus clustering of gliomas was clustered into two clusters as follows: cluster 1 and cluster 2. Moreover, cluster 1 was associated with a worse poor prognosis. Meanwhile, cluster 1 was worse than cluster 2 in overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) (Figures 3D–L). Based on the eight prognostic-related

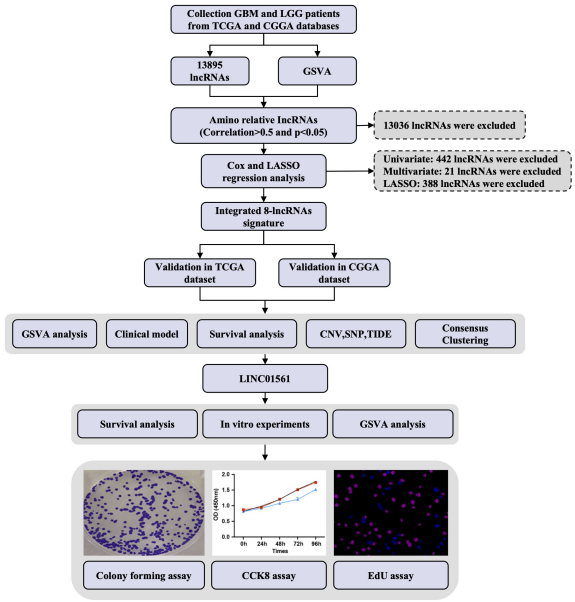


FIGURE 2
Study design flowchart.

lncRNAs, the results of the principal component analysis of the two clusters showed that there was an obvious separation between the survival probability distribution and PCI distribution in TCGA and CGGA datasets (Figures 3C, S3C).

3.3 Establish and validate the risk score model

According to the median risk score, patients were assigned into high-risk and low-risk groups. The results showed that high-risk patients had a significantly worse prognosis than the low-risk group in both TCGA dataset and CGGA dataset ($p < 0.05$, Figures 1, S1). To further evaluate the clinical endpoints derived from the two risk groups, DSS, OS, and PFI were used to compare the survival results of the patients in TCGA dataset. OS was used to compare the survival results of the patients in the CGGA dataset. Univariate analysis shows that when we use DSS, OS, and PFI as clinical endpoints, the high-risk group showed a worse prognosis than the low-risk group in TCGA dataset (Figures 1D–F, S1A–F) and OS in the CGGA dataset (Figures S1G–I).

3.4 Risk models of different subgroups of gliomas

To explore the survival status of different subgroups of gliomas, single-factor Cox regression analysis was performed on the 1p/19q status, gender, age, pathological grade, IDH type, MGMT methylation status, cancer type, and other factors in TCGA database. In addition to the gender of patients, other subtypes are associated with prognosis. Patients with MGMT methylation had lower risk scores (Figures S2A–I). The patient's risk is related to age. Patients older than 45 years of age have a higher risk than patients younger than 45 years. 1p/19q no deletion and IDH wild type have higher risk scores with a poor prognosis. As the pathological grade of glioma increases, the risk score also increases and has a poor prognosis. These results demonstrate that clinical features were significantly associated with the risk score of the model, and a significant positive correlation between high-risk scores and dangerous clinical features was observed (Table 1).

3.5 Correlation between genomic alterations and the glioma risk model

Analysis of somatic mutation and CNV showed that there were significant differences between the low-risk and high-risk groups. Compared with the overall population, high-risk groups of patients had more gene mutations on chromosomes 7 and 10 and fewer gene mutations on chromosome 1. 9p is prone to copy number deletion gene mutation, and 1q, 4q, 7p, and 12q are prone to copy number amplification gene mutation in the high-risk group (Figures 4C, D). IDH1, TP53, ATRX, and CIC were the main variant genes in the low-risk group, whereas TP53, IDH1, ATRX, and EGFR were the main CNV genes in the high-risk group. Patients with low-risk scores had significantly higher mutation frequencies of IDH1 and TP53 than high-risk score patients (IDH1, 90% vs. 33%; TP53, 47% vs. 38%) (Figures 4A, B). The risk of IDH1 missense mutations is significantly reduced within the high-risk group. IDH1 missense mutations are related to low-grade gliomas and may bring patients better prognoses. The variation of CIC (27%) mainly occurred in the low-risk group. PTEN (18%) mutations mainly occur in high-risk groups. Studies have shown that PTEN deficiency can cause PAX7 to be up-regulated, which in turn promotes the carcinogenic transformation of NSC, which may be associated with a poor prognosis.

10.3389/fimmu.2023.1014378

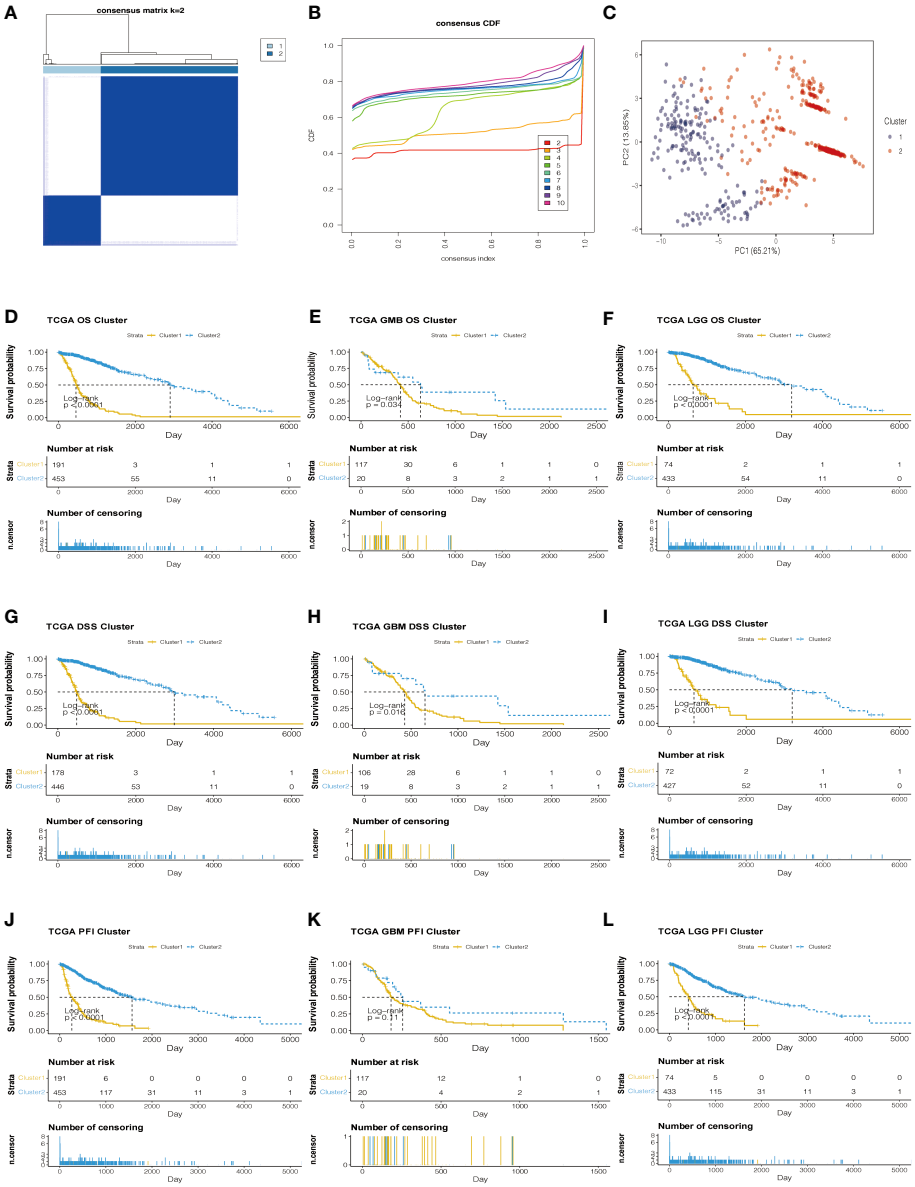


FIGURE 3
Consensus Cluster analyses of the 8 lncRNA expressions in the TCGA dataset. **(A)** Clustering stability for k=2 to 10. **(B)** The incremental change is small under the CDF curve area when k = 2. **(C)** The consensus clustering of gliomas was clustered into two clusters in principal component analysis (PCA). **(D–L)** The K-M plots of the difference in OS, DSS, and PFI between cluster 1 and cluster 2 in TCGA datasets.

3.6 Functional assays of the selected prognostic lncRNAs

To further verify the potential function and target gene of the eight prognostic lncRNAs, GSVA was used to perform with TCGA and CGGA data. It can be seen that the representative GO terminology is significantly related to the prognostic lncRNA in TCGA dataset. Samples from the high-risk group were enriched with more biological processes related to amino acid metabolism compared with the low-risk group. In both datasets of TCGA and CGGA, the 20 most significant biological processes of amino metabolism correlated with risk score were identified by correlation analysis of both TCGA and CGGA datasets, respectively (Figure 5). The most enriched functions of lncRNAs included regulation of cellular amino acid metabolism processes, aromatic amino acid family metabolism, cell-modified amino acid biosynthesis, aromatic amino acid family metabolic process, and amino sugar metabolism. Among them, cell-modified amino acid biosynthesis, aromatic amino acid family metabolic processes, and amino sugar metabolic processes are strongly and positively correlated. The lncRNA risk model is involved in the regulation of the tumor microenvironment, directly or indirectly promoting tumor cell growth, tumor blood vessel formation, migration and invasion, and inhibiting cell apoptosis.

TABLE 1 Clinical baseline characteristics between high- and low-risk groups in TCGA dataset.

Characteristic	N	High (N = 322)	Low (N = 323)	p-value
Age	645			<0.001
		96 (30%)	222 (69%)	
		226 (70%)	101 (31%)	
Gender	645			>0.9
Female		136 (42%)	136 (42%)	
Male		186 (58%)	187 (58%)	
Cancer	645			<0.001
GBM		136 (42%)	1 (0.3%)	
LGG		186 (58%)	322 (99.7%)	
IDH	638			<0.001
Mutant		113 (36%)	308 (95.7%)	
WT		203 (64%)	14 (4.3%)	
1p/19q	641			<0.001
Codel		30 (9.4%)	138 (43%)	
Non-codel		289 (90.6%)	184 (57%)	
MGMT	614			<0.001
Methylated		171 (59%)	292 (90.4%)	
Unmethylated		120 (41%)	31 (9.6%)	

3.7 Prognostic nomogram assessment of overall survival prediction

Based on the association between clinicopathological characteristics and risk score, a nomogram was conducted based on age, glioma risk, 1p/19q, IDH mutation, and glioma grades to predict the survival rate of patients (Figure 6A). The predicted 3- and 5-year survival was agreed strongly with the actual rates in TCGA and CGGA datasets (Figures 6B, D). The AUC and ROC were performed to compare the survival rate at 3 and 5 years and the risk score model showed a good consistency (Figures 6C–E).

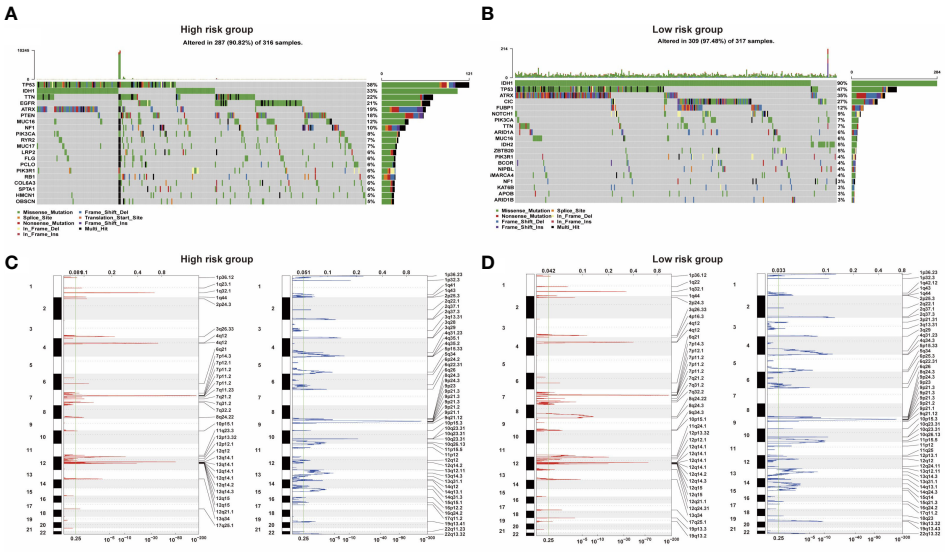


FIGURE 4
(A, B) Two groups of somatic mutation profiles. **(C, D)** Deleted and amplified chromosomal regions were presented in blue and red in high- and low-risk groups.

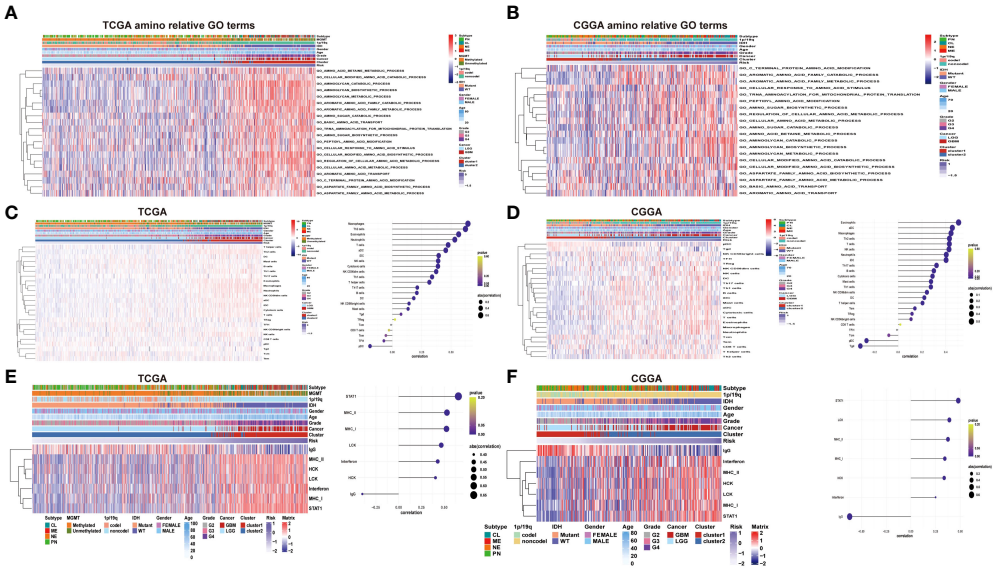


FIGURE 5
(A–B) Heatmaps of the identified amino acid metabolism functions, clinicopathological characteristics, and risk score in the TCGA and CGGA datasets. **(C–D)** Heatmap of immune infiltrating cells in different levels of the risk score in TCGA and CGGA datasets. **(E–F)** Heatmap of inflammatory signature genes under different risk score levels in both the TCGA and CGGA datasets.

3.8 LINC01561 is a prognosis-related biomarker of gliomas correlated with immune infiltration

LINC01561 has been reported in other cancers, but its effect on glioma has not been reported. Hence, LINC01561 was chosen for further analysis. We compared GBM and LGG samples in TCGA dataset with normal samples in the GTEx dataset in the GEPIA online database (http://gepia.cancer-pku.cn). The expression of LINC01561 in the tumor samples is higher than that in the normal samples, but significant differences were found only in the LGG samples (Figure 7A).

In addition to the lncRNA signature, the prognostic value in gliomas was also further confirmed. GSVA results also indicated that LINC01561 may be involved in regulating the prognosis of gliomas. The tumor immune microenvironment is considered closely associated with cancer initiation, prognosis, and response to immunotherapy (25). Submap analysis manifested that the high risk-score group is more sensitive to anti–PD-1 therapy and CTLA4 therapy (nominal p value <0.05, Figure 7B), suggesting that it may

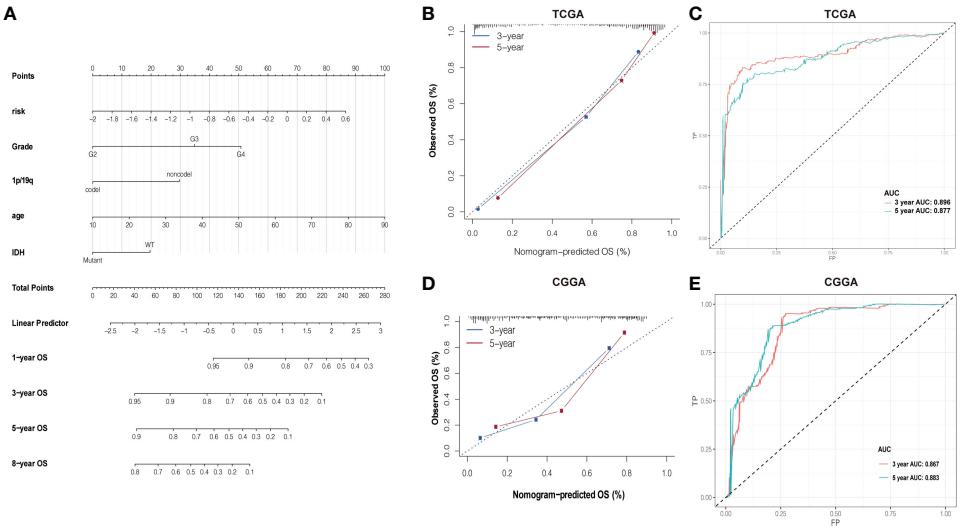


FIGURE 6
(A) Nomogram of summing the points for each variable. **(B, D)** The difference in K-M plots of the nomogram predicted OS between high- and lowrisk groups in both TCGA and CGGA datasets. **(C, E)** Nomogram of ROC curves and AUC values in both TCGA and CGGA datasets.

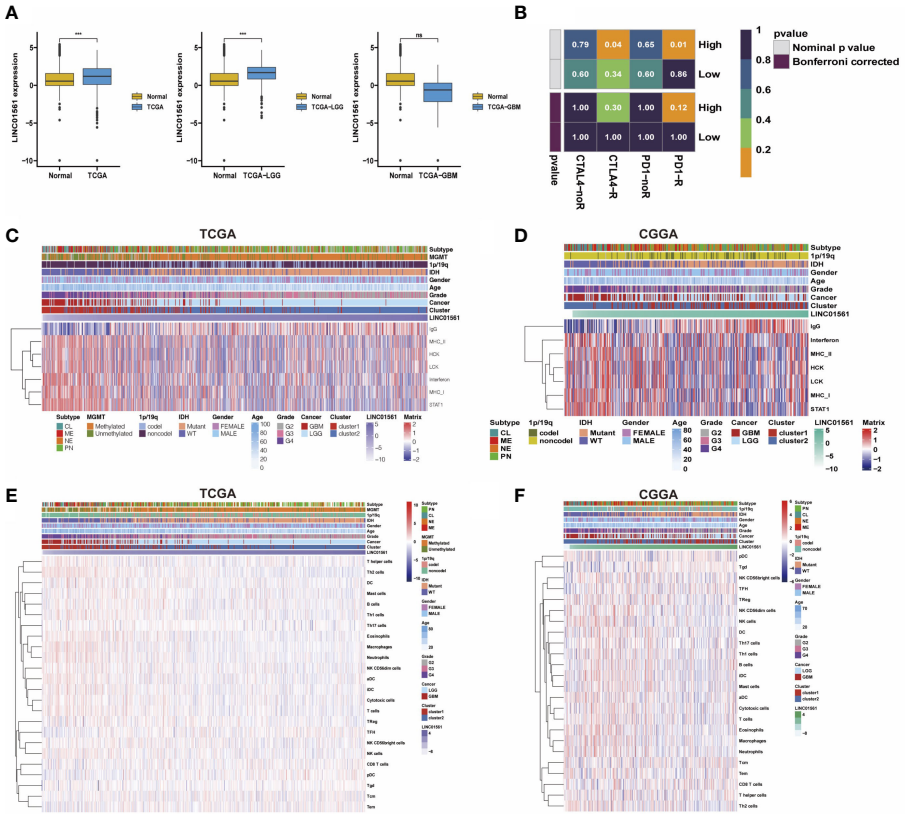


FIGURE 7
(A) Comparison of LINC01561 expression in different groups. **(B)** The differences of the risk score were compared using submap analysis in the anti-PD-1 response and anti-cytotoxic T lymphocyte-associated antigen-4 (CTLA-4) response. **(C-F)** Correlograms between immune cell infiltration, inflammation activities, and LINC01561 expression in TCGA dataset and the CGGA dataset. ***p < 0.001.

benefit from immunotherapy. Moreover, the correlation between LINC01561 and immune activity was explored. The correlation analysis showed that the infiltration of pDC, TFH, Tem, CD8T cells, and Tgd were positively associated with the expression of LINC01561, whereas TH17 cells, NK CD56bright cells, B cells, Treg, mast cells, DC, Th1 cells, NK cells, T helper cells, cytotoxic cells, NK CD56dim cells, iDC, eosinophils, aDC, T cells, neutrophages, macrophages, and Th2 cells were negatively associated. On the level of the inflammatory response, the expression of LINC01561 was positively correlated with IgG, MHC-II, STAT1, LCK, MHC-I, HCK, and interferon. These results hinted that LINC01561 was involved in regulating the immune microenvironment of gliomas (Figures 7C–F).

3.9 LINC01561 affected glioma cell viability, migration, and proliferation

To know the role of LINC01561 on glioma, we further conducted the effects of the increased level of LINC01561 in glioma cell lines (U251 and U87-MG) in *in vitro* experiments. The SiRNA technique was conducted to interfere with the mRNA expression of LINC01561 successfully, and it was confirmed by RT-PCR. CCK-8 assay results showed inhibited cell viability after silencing LINC01561 expression. To further know the effect of LINC01561, a colony-forming assay was also conducted to suggest that the viability of glioma cells was significantly inhibited after LINC01561 was knocked down by siRNA. To further detect whether LINC01561 affects the metastasis of glioma, Transwell experiments were performed to assess the migration and invasion of glioma cell. These results demonstrated that siRNA-mediated LINC01561 silencing suppresses the invasion of glioma. Furthermore, the EdU proliferation assay indicated that inhibition of LINC01561 expression can lead to a significantly decreased EdU-positive rate of glioma cells (Figures 8, S4).

4 Discussion

Glioma accounts for 80% of all malignant brain tumors (25) and is one of the most frequently occurring malignant central nervous system tumors in adults with a highly invasive and a poor prognosis (26). Currently, the prognosis of glioma patients was judged according to the criteria of the WHO classification in clinic (27). The development of molecular typing in gliomas has progressed rapidly based on the development of sequencing technology and bioinformatic technologies in recent years. Surgery is often unable to remove or completely resect the tumor because it is characterized by a highly invasive potential ability due to the aggressive growth of glioma and peripheral brain tissue, and residual tumor cells are difficult to be treated by radiation and chemotherapy thoroughly (28, 29). New targets for

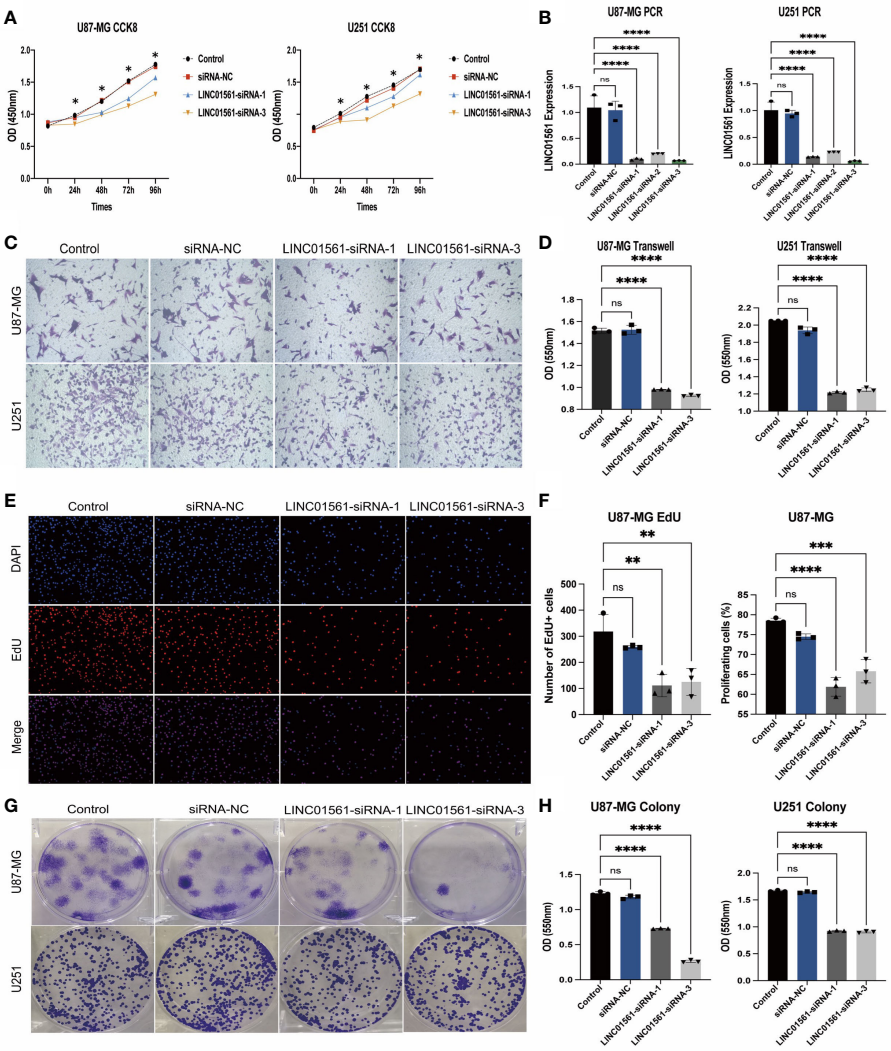


FIGURE 8
(A) Expression results of the CCK-8 assay of the silencing LINC01561 at different time points: 24, 48, 72, and 96h. **(B)** Relative expression levels of LINC01561 in control, siRNA-NC, and LINC01561-siRNA groups. **(C, D)** The results of Transwell assay in the control, siRNA-NC, and LINC01561-siRNA groups. **(E, F)** EdU assay showing the proliferation cells (U87-MG); EdU (red) and DAPI (blue) were stained. **(G, H)** Results of the colony-forming assay in LINC01561 expression. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns p > 0.05.

therapeutic approaches are significant for improving the prognosis of patients with glioma. Altered cellular metabolism is considered a hallmark of glioma cell biology (30, 31). Carbohydrates, lipids, and amino acids utilized by glioma cells and their initiating cells in the hypoxic lesions have important roles in the tumor microenvironment (30). Previous accumulating research mainly focused on lipid metabolism and glucose metabolism in glioma (32). Currently, a new focus on the amino acid metabolism of glioma is considered to play a major role in the glioma (33). In comparison with normal glioma tissues, glioma tissues exhibited higher expression levels of amino acid. LncRNAs have been reported to be involved in many cellular processes, which have been found abnormally expressed in many cancers (34). However, detailed analyses of lncRNAs and amino acid metabolism in glioma have not been reported previously; moreover, the role of amino acid metabolism remains unclear in glioma.

In our study, eight amino-related lncRNAs associated with glioma prognosis by bioinformatics analysis were found, namely, AL357060.1, HOXA-AS3, LINC01561, Z95115.1, AL353796.1, LEF1-AS1, AC005224.3, and TMEM220-AS1, which had significant differences in the expression in normal brain and glioma tissue. Therefore, an accurate predict prognostic lncRNA signature of patients was established. Of note, the amino acid metabolism-related risk signature identified in our study is still considered to be an independent prognostic factor after adjusting to the clinical and molecular features. The metabolic conditions of amino acid metabolism hold great potential and accurate predictive power for clinicopathological features. Together, combining the risk signature and other constraints can better predict a prognosis in patients with glioma. Moreover, in terms of further investigating the lncRNA signature, some researchers make an innovative methodological

contribution, like having someone develop a new system assay for investigating associations on patterns of gene expression (35).

Long non-coding RNAs have been reported as new potentially promising therapeutic targets involved in multiple cancers (36). HOXA-AS3 has been reported to be upregulated in human pulmonary artery smooth muscle cells in the presence of hypoxia (5). HOXA-AS3 can regulate chemoresistance in several types of cancer (37). The expression of HOXA-AS3 increased in glioma patient samples and glioma cell lines (38). LINC01561, a newly identified tumor-related lncRNA, has critical regulatory roles in several tumors including non-small-cell lung carcinoma (39), lung squamous cell carcinoma (40), and breast cancer (41). However, no reports have shown the involvement of LINC01561 in glioma and its underlying mechanisms remain unknown. TMEM220-AS1 was identified as a new prognostic lncRNA biomarker in hepatocellular carcinoma. Downregulated lncRNA TMEM220-AS1 was associated with poorer prognosis in hepatocellular carcinoma (42). LEF1-AS1 is an lncRNA whose expression was significantly upregulated in glioma tissues and cell lines. The knockdown of LEF1-AS1 represses cell proliferation and suppressed tumor growth while activating apoptosis in glioma *via* the downregulated LEF1-AS1/miR-489-3p/HIGD1A axis (43). AL357060.1 has been reported as a new potential prognostic biomarker for hepatocellular carcinoma (44), but there is no report of its role in glioma. Low-risk patients have significantly longer progression-free and overall survival compared with the high-risk group.

Our study provided new insights into biological functions in the above signature lncRNAs, which participate in the amino acid metabolism of glioma. These lncRNAs mainly regulate amino acid biosynthesis, modification, catabolism, and translation. Amino acids (AAs) are essential for the function and survival of the cell. Pathways of glucose, amino acids, and fatty acid metabolism are involved in glioma reprogramming. Recently, great progress has been made in single-cell analysis (45), which is being generally used in terms of tumor study (33). Some scholars have developed a novel approach of the single-cell multi-omics co-regulatory algorithm to broaden our understanding of the underlying mechanisms in complex tumor (46). There is currently a lack of high-quality single-cell gene expression data for glioma. Our team is currently working with single-cell sequencing of glioma and will use this method as a reference for future research.

In the case of metabolic reprogramming of glioma, carbohydrates, lipids, and amino acids were efficiently utilized in glioma cells with hypoxic lesions. This ensures adequate energy to improve glioma rapid growth and migration while altering the role of cell features and its microenvironment (30). The tumor microenvironment holds an indispensable role in tumor metastasis. In recent years, the role of the immune microenvironment has been the subject of intensive research in gliomas (47). Immunotherapy, tumor microenvironment, and a combination of several efficacious methods have piqued ever-increasing interest (48). They include NK and T-cell dysfunction, T-cell and myeloid-derived suppressor cell expansion, immunosuppressive glioma cell surface factors and cytokines, and tumor microenvironment hypoxia. Gliomas have created a profoundly immunosuppressive environment, both topical and systemic within the tumor (49). Immunosuppressive tumor microenvironment reprogramming improved antitumor effectiveness (27). Moreover, tumor microenvironment-targeted therapies for gliomas may be a comprehensive source of the immune landscape and offers insights into possible strategies to overcome tumor with practical significance. MAT2A is an essential amino acid, and a high expression of MAT2A or an inhibitor of MAT2A can reduce the proliferation of glioma cells. Histone methylation is promoted by MAT2A, and cells can be prompted to proliferate in a methionine-restricted environment, which is associated with the progression of glioma (50). Compared with normal cells, rapidly proliferating tumor cells need higher demand for proteinogenic amino acids, and the bioavailability of the use of amino acids is affected largely in human proteomes (51). There were tight associations between serine, glycine, and other non-essential amino acids and tumorigenesis and tumor development. Much research has been devoted to molecular biomarkers that may contribute to the diagnosis or treatment of glioma, which revealed a series of biomarkers associated with the prognosis of glioma. Initiation and development of glioma involve metabolic alterations and genetic and epigenetic alterations on the molecular level, including lncRNA expression. However, the effects of molecular biological mechanisms on amino acid metabolism are poorly understood. Therefore, amino-related lncRNAs found in our study provided new insights to be considered as new genetic biomarkers for the prognosis and treatment of glioma.

As mentioned above, LINC01561 is a novel identified tumor-related lncRNA associated with the progression and poor prognosis of glioma. However, to date, no reports have investigated the involvement of LINC01561 in glioma and its underlying mechanisms remain unknown. We, therefore, have selected LINC01561 for further study of verification analysis.

Our study revealed a negative correlation between LINC01561 expression and the prognosis of glioma patients, indicating that LINC01561 can serve as a reliable predictor of glioma prognosis. It was somewhat that lncRNA exhibited the highest prognostic value. Moreover, LINC01561 was involved in regulating the immune microenvironment of gliomas, which plays a crucial role in tumor initiation and prognosis (25).

Based on our *in vitro* experiments, we found that siRNA-mediated inhibition of LINC01561 suppressed the proliferation of U87-MG and U251 cells, as confirmed by the CCK-8, colony formation assays, and EdU assays. Transwell experiments revealed that silencing LINC01561 inhibits glioma cell invasion, indicating that downregulation of LINC01561 significantly hindered glioma progression. Taken together, these results demonstrate that LINC01561 is a novel identified tumor-related lncRNA playing an important regulatory role in glioma, potentially involved in multiple biological processes.

5 Conclusion

In this study, our bioinformatics analysis identified eight novel amino acid-related lncRNAs associated with glioma patient survival and established an lncRNA signature for prognostic and therapeutic prediction. These findings suggest that the identified lncRNAs may play important roles in glioma, highlighting the importance of amino acid metabolism in glioma and the need for further research. Although our study identified and selected several key lncRNAs, such as LINC01561, more detailed studies are still required to

elucidate their roles in regulating amino acid metabolism and their specific mechanisms of action. Therefore, further studies are warranted to advance our understanding of glioma and develop more effective management strategies for this disease.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conception and design: QL and ZX. Foundation support: QL and ZX. Experiment: LP, and BY. Acquisition and analysis of data: KL, LP, ZX, and QL. Interpretation of data: ZX and QL. Drafting the manuscript and revising for submission quality: QL, BY, KL, LP, and ZX. Study supervision: LP and ZX. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Hainan Province (No. 821QN1005) and the Science Foundation of AMHT Group (NO.2022YK04).

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/fimmu.2023.1014378/full#supplementary-material

SUPPLEMENTARY FIGURE 1
(A-F) The difference in K-M plots of the OS and PFI between high- and low-risk groups in the dataset of TCGA. **(A-F)** The difference in K-M plots of OS between high- and low-risk groups in the dataset of CGGA.

SUPPLEMENTARY FIGURE 2
The survival status of different subgroups of gliomas. **(A)** Age, **(B)** Cancer type, **(C)** Gender, **(D)** Grade, **(E)** IDH type, **(F)** MGMT status, **(G)** 1p/19q status, **(H)** Cancer subtype, **(I)** Combined cancer subtype.

SUPPLEMENTARY FIGURE 3
Consensus Cluster analyses of the 8 lncRNA expressions in the CCCA dataset. **(A)** Clustering stability for k=2 to 10. **(B)** The incremental change under the CDF curve area. **(C)** The consensus clustering of gliomas was clustered into two clusters in principal component analysis (PCA). **(D-F)** The difference in K-M plots of the OS between the cluster 1 and cluster 2 in CGGA dataset.

SUPPLEMENTARY FIGURE 4
(A) The results of Transwell assay in control, siRNA-NC, and LINC01561-siRNA groups in U251 cells. **(B, C)** The EdU assay showing the U251 cells proliferation; EdU (red) and DAPI (blue) were stained.

SUPPLEMENTARY TABLE 1
In Vitro Cell Experiments supplementary materials

References

1. van Hooren L, Vaccaro A, Ramachandran M, Vazaios K, Libard S, van de Walle T, et al. Agonistic CD40 therapy induces tertiary lymphoid structures but impairs responses to checkpoint blockade in glioma. *Nat Commun* (2021) 12(1):4127. doi: 10.1038/s41467-021-24347-7

2. Szopa W, Burley TA, Kramer-Marek G, Kaspera W. Diagnostic and therapeutic biomarkers in glioblastoma: Current status and future perspectives. *BioMed Res Int* (2017) 2017:8013575. doi: 10.1155/2017/8013575

3. Cheng Q, Tang A, Wang Z, Fang N, Zhang Z, Zhang L, et al. CALD1 modulates gliomas progression via facilitating tumor angiogenesis. *Cancers (Basel)* (2021) 13 (11):2705. doi: 10.3390/cancers13112705

4. D'Ammando A, Raspagliesi L, Gionso M, Franzini A, Porto E, Di Meco F, et al. Sonodynamic therapy for the treatment of intracranial gliomas. *J Clin Med* (2021) 10 (5):1101. doi: 10.3390/jcm10051101

5. Anghileri E, Di Ianni N, Paterra R, Langella T, Zhao J, Eoli M, et al. High tumor mutational burden and T-cell activation are associated with long-term response to anti-PD1 therapy in lynch syndrome recurrent glioblastoma patient. *Cancer Immunol Immunother* (2021) 70(3):831–42. doi: 10.1007/s00262-020-02769-4

6. Shi J, Zhang Y, Yao B, Sun P, Hao Y, Piao H, et al. Role of exosomes in the progression, diagnosis, and treatment of gliomas. *Med Sci Monit* (2020) 26:e924023. doi: 10.12659/MSM.924023

7. Zhou F, Chen B. Acute myeloid leukemia carrying ETV6 mutations: Biologic and clinical features. *Hematology* (2018) 23(9):608–12. doi: 10.1080/10245332.2018.1482051

8. Du H, Huang Y, Hou X, Yu X, Lin S, Wei X, et al. DT-13 inhibits cancer cell migration by regulating NMIIA indirectly in the tumor microenvironment. *Oncol Rep* (2016) 36(2):721–8. doi: 10.3892/or.2016.4890

9. Yang E, Wang X, Gong Z, Yu M, Wu H, Zhang D. Exosome-mediated metabolic reprogramming: The emerging role in tumor microenvironment remodeling and its influence on cancer progression. *Signal Transduct Target Ther* (2020) 5(1):242. doi: 10.1038/s41392-020-00359-5

10. Hu B, Lin JZ, Yang XB, Sang XT. Aberrant lipid metabolism in hepatocellular carcinoma cells as well as immune microenvironment: A review. *Cell Prolif* (2020) 53 (3):e12772. doi: 10.1111/cpr.12772

11. Wang W, Zou W. Amino acids and their transporters in T cell immunity and cancer therapy. *Mol Cell* (2020) 80(3):384–95. doi: 10.1016/j.molcel.2020.09.006

12. Avolio R, Matassa DS, Criscuolo D, Landriscina M, Esposito F. Modulation of mitochondrial metabolic reprogramming and oxidative stress to overcome chemoresistance in cancer. *Biomolecules* (2020) 10(1):135. doi: 10.3390/biom10010135

13. Lauinger L, Kaiser P. Sensing and signaling of methionine metabolism. *Metabolites* (2021) 11(2):83. doi: 10.3390/metabo11020083

14. Alfarsi LH, El-Ansari R, Craze ML, Masisi BK, Mohammed OJ, Ellis IO, et al. Co-Expression effect of SLC7A5/SLC3A2 to predict response to endocrine therapy in oestrogen-Receptor-Positive breast cancer. *Int J Mol Sci* (2020) 21(4):1407. doi: 10.3390/ijms21041407

15. Wang L, Zhao X, Fu J, Xu W, Yuan J. The role of tumour metabolism in cisplatin resistance. *Front Mol Biosci* (2021) 8:691795. doi: 10.3389/fmolb.2021.691795

16. Wei Z, Liu X, Cheng C, Yu W, Yi P. Metabolism of amino acids in cancer. *Front Cell Dev Biol* (2020) 8:603837. doi: 10.3389/fcell.2020.603837

17. Xu S, Tang L, Liu Z, Yang K, Cheng Q. Bioinformatic analyses identify a prognostic autophagy-related long non-coding RNA signature associated with immune microenvironment in diffuse gliomas. *Front Cell Dev Biol* (2021) 9:694633. doi: 10.3389/fcell.2021.694633

18. Zhang B, Chen Y, Shi X, Zhou M, Bao L, Hatanpaa KJ, et al. Regulation of branched-chain amino acid metabolism by hypoxia-inducible factor in glioblastoma. *Cell Mol Life Sci CMLS* (2021) 78(1):195–206. doi: 10.1007/s00018-020-03483-1

19. Brożyna AA, Jóźwicki W, Jetten AM, Slominski AT. On the relationship between VDR, RORα and RORγ receptors expression and HIF1-α levels in human melanomas. *Exp Dermatol* (2019) 28(9):1036–43.

20. Gao W, Li H, Liu Y, Zhang Y, Zhao H, Liu F. Long non−coding RNA FLVCR1 −AS1 promotes glioma cell proliferation and invasion by negatively regulating miR −30b−3p. *Mol Med Rep* (2020) 22(2):723–32. doi: 10.3892/mmr.2020.11149

21. Zong Z, Song Y, Xue Y, Ruan X, Liu X, Yang C, et al. Knockdown of LncRNA SCAMP1 suppressed malignant biological behaviours of glioma cells *via* modulating miR-499a-5p/LMX1A/NLRC5 pathway. *J Cell Mol Med* (2019) 23(8):5048–62. doi: 10.1111/jcmm.14362

22. Huang L, Jiang X, Wang Z, Zhong X, Tai S, Cui Y. Small nucleolar RNA host gene 1: A new biomarker and therapeutic target for cancers. *Pathol Res Pract* (2018) 214 (9):1247–52. doi: 10.1016/j.prp.2018.07.033

23. Yin X, Yin Y, Dai L, Shen C, Chen N, Li J, et al. Integrated analysis of long non-coding RNAs and mRNAs associated with malignant transformation of gastrointestinal stromal tumors. *Cell Death Dis* (2021) 12(7):669. doi: 10.1038/s41419-021-03942-y

24. Li Y, Huo J, He J, Ma X. LncRNA MONC suppresses the malignant phenotype of endometrial cancer stem cells and endometrial carcinoma cells by regulating the MiR-636/ GLCE axis. *Cancer Cell Int* (2021) 21(1):331. doi: 10.1186/s12935-020-01730-w

25. Han W, Hu P, Wu F, Wang S, Hu Y, Li S, et al. FHL3 links cell growth and self-renewal by modulating SOX4 in glioma. *Cell Death Differ* (2019) 26(5):796–811. doi: 10.1038/s41418-018-0152-1

26. Hu Q, Huang K, Tao C, Zhu X. Protein disulphide isomerase can predict the clinical prognostic value and contribute to malignant progression in gliomas. *J Cell Mol Med* (2020) 24(10):5888–900. doi: 10.1111/jcmm.15264

27. Yang R, Zhang Z, Fu S, Hou T, Mu W, Liang S, et al. Charge and size dual switchable nanocage for novel triple-interlocked combination therapy pattern. *Advanced Sci (Weinheim Baden-Wurttemberg Germany)* (2020) 7(18):2000906. doi: 10.1002/advs.202000906

28. Sukumari-Ramesh S, Prasad N, Alleyne CH, Vender JR, Dhandapani KM. Overexpression of Nrf2 attenuates carmustine-induced cytotoxicity in U87MG human glioma cells. *BMC cancer* (2015) 15:118. doi: 10.1186/s12885-015-1134-z

29. Zhu Y, Jia J, Zhao G, Huang X, Wang L, Zhang Y, et al. Multi-responsive nanofibers composite gel for local drug delivery to inhibit recurrence of glioma after operation. *J nanobiotechnol* (2021) 19(1):198. doi: 10.1186/s12951-021-00943-z

30. Qiu R, Zhong Y, Li Q, Li Y, Fan H. Metabolic remodeling in glioma immune microenvironment: Intercellular interactions distinct from peripheral tumors. *Front Cell Dev Biol* (2021) 9:693215. doi: 10.3389/fcell.2021.693215

31. Di Ianni N, Musio S, Pellegatta S. Altered metabolism in glioblastoma: Myeloid-derived suppressor cell (MDSC) fitness and tumor-infiltrating lymphocyte (TIL) dysfunction. *Int J Mol Sci* (2021) 22(9):4460. doi: 10.3390/ijms22094460

32. Calzadilla PI, Maiale SJ, Ruiz OA, Escaray FJ. Transcriptome response mediated by cold stress in lotus japonicus. *Front Plant sci* (2016) 7:374. doi: 10.3389/ fpls.2016.00374

33. Way GP, Greene CS. Bayesian Deep learning for single-cell analysis. *Nat Methods* (2018) 15(12):1009–10. doi: 10.1038/s41592-018-0230-9

34. Jiang H, Wu FR, Liu J, Qin XJ, Jiang NN, Li WP. Effect of astragalosides on long non-coding RNA expression profiles in rats with adjuvant-induced arthritis. *Int J Mol Med* (2019) 44(4):1344–56. doi: 10.3892/ijmm.2019.4281

35. Song Q, Wang H, Bao J, Pullikuth AK, Li KC, Miller LD, et al. Systems biology approach to studying proliferation-dependent prognostic subnetworks in breast cancer. *Sci Rep* (2015) 5:12981. doi: 10.1038/srep12981

36. Song J, Xu Q, Zhang H, Yin X, Zhu C, Zhao K, et al. Five key lncRNAs considered as prognostic targets for predicting pancreatic ductal adenocarcinoma. *J Cell Biochem* (2018) 119(6):4559–69. doi: 10.1002/jcb.26598

37. Chen D, Xie S, Wu Y, Cui Y, Cai Y, Lan L, et al. Reduction of bladder cancer chemosensitivity induced by the effect of HOXA-AS3 as a ceRNA for miR-455-5p that upregulates Notch1. *Front Oncol* (2020) 10:572672. doi: 10.3389/fonc.2020.572672

38. Wu F, Zhang C, Cai J, Yang F, Liang T, Yan X, et al. Upregulation of long noncoding RNA HOXA-AS3 promotes tumor progression and predicts poor prognosis in glioma. *Oncotarget* (2017) 8(32):53110–23. doi: 10.18632/oncotarget.18162

39. Gao W, Qi CQ, Feng MG, Yang P, Liu L, Sun SH. SOX2-induced upregulation of lncRNA LINC01561 promotes non-small-cell lung carcinoma progression by sponging miR-760 to modulate SHCBP1 expression. *J Cell Physiol* (2020) 235(10):6684–96. doi: 10.1002/jcp.29564

40. Shi Y, Li Y, Yan C, Su H, Ying K. Identification of key genes and evaluation of clinical outcomes in lung squamous cell carcinoma using integrated bioinformatics analysis. *Oncol Lett* (2019) 18(6):5859–70. doi: 10.3892/ol.2019.10933

41. Jiang R, Zhao C, Gao B, Xu J, Song W, Shi P. Mixomics analysis of breast cancer: Long non-coding RNA linc01561 acts as ceRNA involved in the progression of breast cancer. *Int J Biochem Cell Biol* (2018) 102:1–9. doi: 10.1016/j.biocel.2018.06.003

42. Du W, Chen W, Shu Z, Xiang D, Bi K, Lu Y, et al. Identification of prognostic biomarkers of hepatocellular carcinoma *via* long noncoding RNA expression and copy number alterations. *Epigenomics* (2020) 12(15):1303–15. doi: 10.2217/epi-2019-0385

43. Cheng Z, Wang G, Zhu W, Luo C, Guo Z. LEF1-AS1 accelerates tumorigenesis in glioma by sponging miR-489-3p to enhance HIGD1A. *Cell Death Dis* (2020) 11 (8):690. doi: 10.1038/s41419-020-02823-0

44. Ye J, Zhang J, Lv Y, Wei J, Shen X, Huang J, et al. Integrated analysis of a competing endogenous RNA network reveals key long noncoding RNAs as potential prognostic biomarkers for hepatocellular carcinoma. *J Cell Biochem* (2019) 120 (8):13810–25. doi: 10.1002/jcb.28655

45. Ji Z, Zhou W, Hou W, Ji H. Single-cell ATAC-seq signal extraction and enhancement with SCATE. *Genome Biol* (2020) 21(1):161. doi: 10.1186/s13059-020-02075-3

46. Song Q, Zhu X, Jin L, Chen M, Zhang W, Su J. SMGR: A joint statistical method for integrative analysis of single-cell multi-omics data. *NAR Genomics Bioinf* (2022) 4 (3):lqac056. doi: 10.1093/nargab/lqac056

47. Klemm F, Maas RR, Bowman RL, Kornete M, Soukup K, Nassiri S, et al. Interrogation of the microenvironmental landscape in brain tumors reveals disease-specific alterations of immune cells. *Cell* (2020) 181(7):1643–60.e17. doi: 10.1016/j.cell.2020.05.007

48. Yang K, Wu Z, Zhang H, Zhang N, Wu W, Wang Z, et al. Glioma targeted therapy: Insight into future of molecular approaches. *Mol cancer* (2022) 21(1):39. doi: 10.1186/s12943-022-01513-z

49. Grabowski MM, Sankey EW, Ryan KJ, Chongsathidkiet P, Lorrey SJ, Wilkinson DS, et al. Immune suppression in gliomas. *J neuro-oncol* (2021) 151(1):3–12. doi: 10.1007/s11060-020-03483-y

50. Wang L, Hu B, Pan K, Chang J, Zhao X, Chen L, et al. SYVN1-MTR4-MAT2A signaling axis regulates methionine metabolism in glioma cells. *Front Cell Dev Biol* (2021) 9:633259. doi: 10.3389/fcell.2021.633259

51. Zhang H, Wang Y, Li J, Chen H, He X, Zhang H, et al. Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat Commun* (2018) 9(1):4124. doi: 10.1038/s41467-018-06461-1

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 26 April 2023
DOI 10.3389/fimmu.2023.1142319

OPEN ACCESS

EDITED BY
Ping Zheng,
The University of Melbourne, Australia

REVIEWED BY
Noha Mousaad Elemam,
University of Sharjah, United Arab Emirates
Siriporn Jitkaew,
Chulalongkorn University, Thailand
Kai Wang,
Amgen, United States

*CORRESPONDENCE
Shuyun Zhang
13214501198@163.com

†These authors have contributed equally to this work

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

RECEIVED 11 January 2023
ACCEPTED 29 March 2023
PUBLISHED 26 April 2023

CITATION
Dong J, Zhang R, Xia Y, Jiang X, Zhou K, Li J, Guo M, Cao X and Zhang S (2023) The necroptosis related gene LGALS3 can be used as a biomarker for the adverse progression from chronic HBV infection to HCC.
Front. Immunol. 14:1142319.
doi: 10.3389/fimmu.2023.1142319

The necroptosis related gene LGALS3 can be used as a biomarker for the adverse progression from chronic HBV infection to HCC

Jianming Dong[1†], Rongzheng Zhang[1†], Yan Xia[1†], Xu Jiang[2], Kun Zhou[1,3], Jiaqi Li[1], Mengrui Guo[1], Xinyang Cao[1] and Shuyun Zhang[1*]

[1]Scientific Research Center, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Parasitology, Harbin Medical University, Harbin, China, [3]Beidahuang Industry Group General Hospital Department of Clinical Laboratory, Harbin, China

The number of patients with hepatocellular carcinoma (HCC) caused by hepatitis B virus (HBV) infection remains large, despite the remarkable effectiveness of antiviral drugs and vaccines for HBV in preventing and treating HBV infection. Necroptosis is closely related to the occurrence of inflammation, clearance of viral infection, and tumor progression. Presently, little is known about the changes in necroptosis-related genes in the progression from chronic HBV infection (CHI) to HBV-related hepatic fibrosis (HBV-HF) and HBV-related hepatocellular carcinoma (HBV-HCC). In this study, Cox regression analysis was performed using GSE14520 chip data and a necroptosis-related genes survival prognosis score (NRGPS) was established for HBV-HCC patients. NRGPS was constructed using three model genes (G6PD, PINK1 and LGALS3), and verified by data sequencing in the TCGA database. The HBV-HCC cell model was established by transfection of pAAV/HBV1.2$_{C2}$, constructed by homologous recombination, into HUH7 and HEPG2 cells. The expression levels of G6PD, PINK1, and LGALS3 were detected using RT-qPCR. We further analyzed the expression of the model genes in GSE83148, GSE84044, and GSE14520 and found that LGALS3 was consistently highly expressed in CHI, high fibrosis score and high NRGPS. In addition, immune microenvironment analysis showed that LGALS3 was not only associated with the infiltration of regulatory T cells in the immune microenvironment but also with expression of CCL20 and CCR6. The expression levels of model genes, FOXP3 and CCR6, were analyzed using RT-qPCR in peripheral blood mononuclear cells of 31 hepatitis B surface antibody positive patients, 30 CHI, 21 HBV-HF, and 20 HBV-HCC. In further cell-model experiments, we analyzed the expression of CCL20 by RT-qPCR and the changes in cell proliferation and migration by CCK8 and transwell assays, respectively, in HBV-HCC cell models after LGALS3 knockdown. The findings of this study suggest that LGALS3 could be a biomarker for adverse progression following

chronic HBV infection and may also be involved in the regulation of the immune microenvironment, making it a potential therapeutic target.

KEYWORDS

necroptosis related gene, HBV infection, hepatocellular carcinoma, immune microenvironment, prognosis

1 Introduction

Hepatitis B virus (HBV) infection is a worldwide epidemic, with an estimated 257 million patients with chronic hepatitis B worldwide. Furthermore, about 887,000 people die each year from diseases related to HBV infection, such as HBV-related hepatic fibrosis (HBV-HF) and HBV-related hepatocellular carcinoma (HBV-HCC). In China, 77% of hepatic fibrosis and 84% of hepatocellular carcinoma cases are caused by HBV (1). During chronic HBV infection (CHI) progression, the activity of the necroptotic pathway in diseased liver tissues is often enhanced (2).

Necroptosis is an important means of interacting with pathogenic microorganisms, in addition to the immune system, and it is mainly composed of receptor-interacting protein kinase 1 (RIPK1), receptor-interacting protein kinase 3 (RIPK3), and mixed lineage kinase domain-like (MLKL) protein. The changes in the conformation of the MLKL protein and translocation to the cell membrane, as well as the flow of Na^+, Ca^{2+}, and other cations increases the intracellular osmotic pressure. Furthermore, it can also directly form pore structures on the cell membrane, leading to cell destruction and release of a large number of cell contents. The leaked cell contents also form damage-related molecular patterns that allow for a massive invasion of surrounding immune cells, triggering intense inflammation (3, 4). Infected cells are eliminated through necroptosis and virus transmission is limited by the release of inflammatory mediators that promote the activation of innate and adaptive immunity (5). In this study, we used the Gene Expression Omnibus (GEO) database to develop prognostic markers of HBV-HCC composed of three necroptosis-related genes (NRGs), including LGALS3, G6PD, and PINK1. The relative-expression levels of model genes mRNA in HEPG2, HUH7, LX2, and peripheral blood mononuclear cells (PBMCs) were measured by RT-qPCR. Finally, the cancer genome atlas (TCGA) database was analyzed to verify the necroptosis-related genes survival prognosis score (NRGPS). LGALS3, a member of the galactoagglutinin family, is a vital gene in the regulation of the liver immune microenvironment (6, 7) and is significantly associated with increased risk of liver failure, cirrhosis, chronic active hepatitis B, and shorter survival time for hepatocellular carcinoma (HCC) (8, 9). Analysis of the relationship between LGALS3 and adverse disease progression after persistent HBV infection and its immune microenvironment may help understand the role of necroptosis in chronic HBV infection, which is of great significance for the prevention and treatment of adverse disease progression. Therefore, this study aimed to analyze the relationship between LGALS3 and adverse disease progression after persistent HBV infection and its immune microenvironment.

2 Materials and methods

Figure 1, a flowchart, illustrates this study's design and procedure.

2.1 Acquisition of data from patients with HBV infection

We downloaded the following gene expression profiles from the GEO database: GSE83148, GSE84044, and GSE14520. As the GSE14520 dataset contained two different batches of expression profile chip datasets, we selected the Affymetrix HT Human Genome U133A Array dataset, which contained the majority of patients, to avoid batch processing effects. Sequencing data from 89 HBV-HCC were obtained from TCGA. Supplementary Table 1 shows the specific clinical information.

2.2 Identification of differentially expressed NRGs

Overall, 583 NRGs were obtained from GeneCards (www.genecards.org/) and their expression levels were obtained by intersection with gene expression data from GSE14520, using the "limma" R package, and with | log2 fold change (FC) | > 1 and adjusted $p < 0.05$ as filter conditions. Finally, differentially expressed necroptosis-related genes (DENRGs) were obtained from the tumor and paracancerous tissues. The "pheatmap," "ggpubr," "ggrepel," "dplyr," "ggplot2," and "rcircos" R packages were used to generate heat maps, volcano maps, and chromosome pattern maps of DENRGs. Gene ontology (GO) analysis was performed using the "clusterprofiler," "enrichment plot," and "ggplot2" R packages.

2.3 Construction and validation of the NRGPS

Follow-up time ≥ 30 days was used as inclusion criteria. The "caret" R package was used to divide the 212 HBV-HCC patients

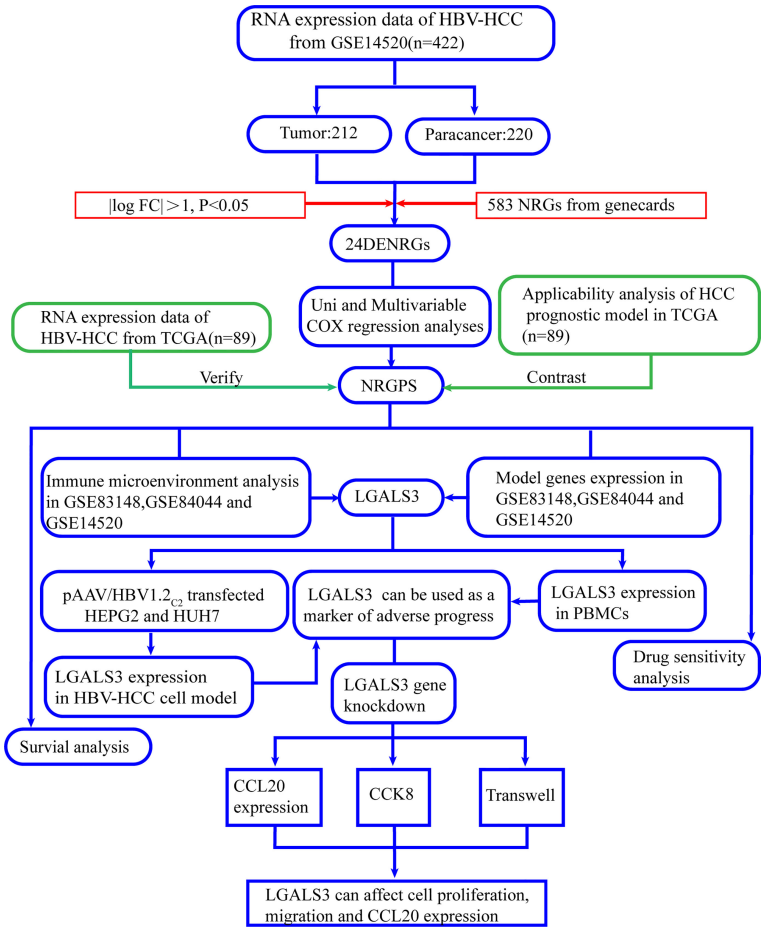


FIGURE 1
Flowchart of this study.

randomly into the training and test sets in a ratio of 6:4. Table 1 presents the specific clinical information for the training and test sets. Univariable Cox regression analysis was performed for 24 DENRGs in the training set using "survival" and "survminer" R packages to screen for DENRGs associated with prognosis in the training set. Subsequently, based on the survival information of the samples, including survival time and survival state, combined with the expression values of prognostic related genes in each sample, through multivariable Cox regression analysis and stepwise regression algorithm, the gene combination with the greatest

TABLE 1 There was no significant difference between the two groups of clinical phenotypes.

Parameters	Training set	Test set	P value
N	129	83	
Age >55 years (%)	29.46	32.53	0.636
Gender, male (%)	88.37	83.13	0.279
TNM stage			0.236
I	53	36	
II	46	30	
III	30	17	
OS time, median (IQR)	4.33 (1.5–4.82)	4.22 (1.48–4.71)	0.482

OS, overall survival; IQR, interquartile range.

impact on survival was found, combined with the corresponding regression coefficient to construct NRGPS, the calculation formula is as follows:

$$\mathrm{NRGPS} = \sum_{i=1}^{N}(Exp(gene_i) \times Coef(gene_i))$$

Kaplan–Meier curve analysis was performed using "survival" and "survminer" R packages to determine the survival rate of high and low NRGPS groups. Time-dependent receiver operating characteristic curve analysis was performed using "survivalROC" R package to evaluate the prognostic value of survival indicators. Moreover, "Rtsne" package principal component analysis was used to reduce the dimensionality of the multivariable data in the NRGPS, and the data was visualized. Univariable and multivariable Cox regression analyses were conducted to evaluate the independent prognostic value of the NRGPS. Additionally, "c2.cp.kegg.v7.5.1. symbols.gmt" was used as the reference file, and Gene Set Enrichment Analysis (GSEA) was performed using "limma," "GSEABase," "GSVA", and "pheatmap". Finally, the potential biological processes and risk paths differing between the high and low NRGPS in the GSE14520 cohort were visualized. A nomogram based on NRGPS was developed using the "rms" R package to predict the prognosis of patients with HBV-HCC. The TCGA dataset was used as an external validation of NRGPS.

2.4 Prognostic analysis of existing prognostic models of necroptosis-related genes in HCC

We collected several studies (10–15) on the construction of prognostic models of HCC patients using NRGs, and conducted a Kaplan–Meier curve analysis of 89 patients with HBV-HCC in the TCGA database using the prognostic models in the literature to determine whether the prognostic models in the literature are suitable for HBV-HCC.

2.5 Drug sensitivity analysis

Differences in drug sensitivity between high and low NRGPS groups were analyzed using "ggplot2" and "pRRophetic" R packages in the TCGA cohort. A Ridge regression model was constructed using the pRRophetic algorithm based on the genomics of drug sensitivity in cancer cell line and TCGA gene expression profiles to calculate the difference in IC50 of chemotherapy drugs between the high and low NRGPS groups. Additionally, we downloaded the drug structures using the PubChem database (https://pubchem.ncbi.nlm.nih.gov/).

2.6 Immune cell infiltration analysis and functional analysis

In GSE83148, GSE84044, GSE14520, and TCGA cohorts, single-sample gene set enrichment analysis (ssGSEA) was conducted using the "GSVA" R package to analyze related immune cells and immune pathways in the microenvironment and predict the degree of immune cell infiltration. Furthermore, "reshape2," "limma," "tidyverse," "corrplot," and "ggplot2" R packages were used to analyze and map the correlation of modeling genes with chemokine receptors and ligands, immune cellsand function.

2.7 Clinical specimen collection

From February 2022 to August 2022, 1 mL of EDTA anticoagulant peripheral blood was collected from 71 outpatients at the Second Affiliated Hospital of Harbin Medical University as an experimental specimen, including 30 cases of CHI, 21 cases of HBV-HF, and 20 cases of HBV-HCC. The EDTA anticoagulant peripheral blood of 31 healthy patients with hepatitis B surface antibody positive (HBsAb+) was used as the control group. The inclusion criteria are shown in Supplementary Table 2. The clinical information is shown in Supplementary Table 3. Human whole blood mononuclear cell isolation solution was used to extract mononuclear cells and, after addition of 1 mL Seven RNAkey™ Reagent (SM129-02, Seven, China), samples were frozen at -80°C and stored until further use.

2.8 Construction of HBV-HCC cell model

Based on GenBank: JQ688404.1 Whole genome of HBV C2 subtype and 1.2 fold gene sequence of pAAV/HBV1.2 plasmid constructed by Huang et al. (16): NT140-NT3182/1-NT1987 using pAAV-MCS as carrier, pAAV/HBV1.2$_{C2}$ recombinants were constructed by gene synthesis and homologous recombination. The process was completed by Beijing Liuhe Huada Genomics Technology Co., LTD. pAAV-MCS carrier was purchased from Harbin Suit Biotechnology Co., Ltd. A high-purity Midiprep Kit (ZP104-1, Zoman, China) was used to extract the plasmid. Moreover, HEPG2 and HUH7 cells were purchased from Wuhan Procell Life Technology Co., Ltd.; LX-2 cells were purchased from Shanghai Gaining Biotechnology Co. Ltd. Additionally, HEPG2, HUH7, and LX-2 were cultured in Dulbecco's modified Eagle medium (DMEM) high-glucose medium (D6429, Sigma, USA) with 10% fetal bovine serum (11011-8611, Every Green, China) and 5% CO_2 at 37°C. The pAAV/HBV1.2$_{C2}$ and pAAV-MCS empty vectors were transfected into HEPG2 and HUH7 cells using a liposome transfection reagent (C0533, Beyotime, China), and co-transfected with the pmaxGFP plasmid. HUH7$^{\#}$ and HEPG2$^{\#}$ denote the HUH7 and HEPG2 cell lines after transfection with pAAV/HBV1.2$_{C2}$. The supernatants were collected after 48 h. HBsAg expression in the supernatant of the cells transfected with pAAV/HBV1.2$_{C2}$ was detected using a HBsAg diagnostic kit (S10980090, Wantai, China).

2.9 LGALS3 gene knockdown

pAAV/HBV1.2 or pAAV-MCS were transfected using Lipo8000™ transfection reagent (C0533, Beyotime, China) into HEPG2 and HUH7 cells in six-well plates. After 24 h, the

transfection medium was replaced with the culture medium. Lipo8000™ transfection reagent was used to transfect siRNA with LGALS3 knockdown (si-LGALS3) or negative control (si-NC) (General, Anhui, China) into HEPG2# and HUH7#. After 48h, supernatants and cells were collected. Cells were added to 0.5 mL Seven RNAkey™ Reagent (SM129-02, Seven, China), frozen at -80°C, and stored until further use. siRNA sequences are listed in Supplementary Table 4.

2.10 RT-qPCR

Supplementary Table 4 presents the Primer sequences. Total RNA was extracted from cells using Seven RNAkey™ Reagent (SM129-02, Seven, China), and cDNA was synthesized using a reverse transcription kit (SM131-02A, Seven, China) according to the manufacturer's instructions. The SYBR Green Master Mix kit (SM133-02, Seven, China) and an RT-qPCR instrument (SLAN-96p, Shanghai Hongshi, China) were used for RT-qPCR. Additionally, GAPDH was used as an internal control, and quantification was based on the 2-ΔΔCt method.

2.11 CCK8 assay

The CCK8 assay kit (SC119-01, Seven, China) was used to detect the proliferation of HBV-HCC cells after si-LGALS3 transfection. Cells (1×10^5 cells/well) were cultured with DMEM medium containing 10% fetal bovine serum. Additionally, after culture at 37°C for 24 h, 10 μL CCK8 was added to each well. Subsequently, cells were incubated at 37°C for 1 h, and absorbance was measured at 450 nm.

2.12 Transwell assay

Transwell chambers were used to determine the migration efficiency of the HBV-HCC cells after si-LGALS3 transfection. Cells (2×10^4 cells/well) were added to the upper chamber with serum-free DMEM medium (aperture 8.0 μm), with lower chamber containing DMEM supplemented with 20% fetal bovine serum. After culturing at 37°C for 24 h, the upper chamber cells were wiped with cotton swabs, the cells that passed through the membrane were fixed with 4% paraformaldehyde, stained with 0.1% crystal violet, and then washed with PBS thrice. Finally, the migrating cells in five random fields were counted using an optical microscope.

2.13 Statistical analysis

R (version 4.1.2), GraphPad Prism (version 8.0) and SPSS statistical software (version 25.0.0) were used for statistical analysis. Mann–Whitney U test was used for comparison between the two groups. Kruskal-Wallis H test was used to compare multiple groups. Spearman correlation statistics was used to test the association. $p < 0.05$ was considered statistically significant, with the symbols ***, **, *, and ns representing $p < 0.001$, $p < 0.01$, $p < 0.05$, and not significant, respectively.

3 Results

3.1 Identification and functional enrichment analysis of DENRGs

In the GSE14520 dataset, we identified 24 DENRGs in tumor and adjacent non-tumor tissues. The expression levels of C9, C7, EGR1, ID1, C6, ZFP36, TXNIP, CCL2, PINK1, and AHSG were significantly higher in adjacent non-tumor tissues than those in tumor tissues. The expression levels of CDC7, G6PD, CCT6A, BAG2, IRAK1, LGALS3, AURKA, VIL1, TP53I3, CCT3, SNRPE, NQO1, BUB1B, and IGF2BP3 in normal tissues were significantly lower than those in tumor tissues (Figures 2A, B). Figure 2C shows the chromosomal localization of the 24 DENRGs. To explore the role of DENRGs in HBV-HCC, we performed gene ontology pathway enrichment analysis on the 24 DENRGs. Particularly, gene ontology analysis was used to determine the 24 DENRGs enrichment pathways in HBV-HCC in terms of biological process, cellular component, and molecular function. The results showed positive regulation of cellular protein localization, protein stability, neuronal death, and enhanced protein serine kinase activity (Figure 2D).

3.2 Development and validation of the NRGPS system

The prognostic value of the 24 candidate DENRGs was investigated by conducting univariable Cox regression analysis on 129 patients with HBV-HCC in the training set. The results showed that CCT3, CCT6A, VIL1, PINK1, LGALS3, and G6PD significantly correlated with HBV-HCC prognosis (Figures 3A–G). Multivariable Cox regression analysis was performed for the above mentioned six genes, and the NRGPS estimated using PINK1, LGALS3, and G6PD was obtained according to the following formula: NRGPS = (-0.4025× PINK1 expression)+ (0.2081× LGALS3 expression)+ (0.2371× G6PD expression). Our results revealed that higher NRGPS scores were associated with higher risk, higher mortality, and shorter survival time in the training, test, and TCGA cohorts (Figures 4A–F). The heat map shows that G6PD and LGALS3 were highly expressed in the high NRGPS group, and PINK1 was highly expressed in the low NRGPS group (Figures 4G–I). In the principal component analysis, patients with different NRGPS scores could be clearly divided into two clusters (Figures 4J–L). Kaplan–Meier curves showed that the overall survival of the high-NRGPS group was shorter than that of the low-NRGPS group (Figures 4M–O). In the time-dependent receiver operating characteristic analysis, the areas under the 1-year curve of the training, test, and TCGA cohorts were 0.637, 0.687, and 0.696, respectively. Additionally, the 2-year area under the curve were 0.699, 0.686, and 0.683, respectively, and those of the 3-year area under the curve were 0.695, 0.659, and 0.685, respectively (Figures 5A–

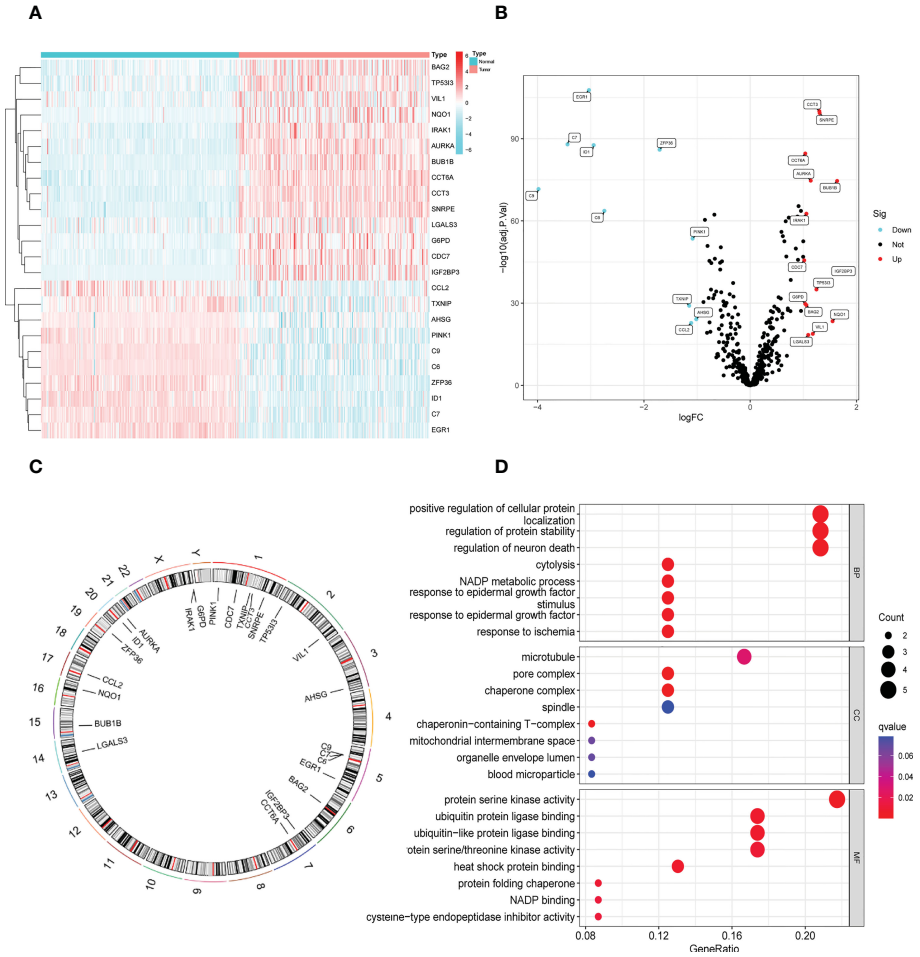


FIGURE 2
Identification and functional enrichment analysis of DENRGs. **(A)** The heatmap of 24 DENRGs in tumor and normal tissues. **(B)** The volcano plots of 24 DENRGs in tumor and normal tissues. **(C)** Chromosome schema maps of 24 DENRGs. **(D)** The GO enrichment analysis of the DENRGs.

C). These results indicated that the NRGPS has high specificity and sensitivity for predicting the prognosis of HBV-HCC.

3.3 Independent prognostic analysis of NRGPS and GSEA

This study evaluated whether NRGPS can be an independent prognostic indicator. Univariable and multivariable Cox regression analyses were performed to determine the associations between prognosis and age, gender, TNM stage, and NRGPS. Similarly, the correlation between age, gender, T stage, and NRGPS was analyzed in TCGA cohort. Univariable Cox regression analysis showed that NRGPS was correlated with patient prognosis in the GSE14520 and TCGA cohorts (Figures 5D–F). Multivariable Cox regression analysis confirmed that the NRGPS was an independent predictor of survival after adjusting for other clinical confounding factors in the GSE14520 and TCGA cohorts (Figures 5G–I). The Kyoto Encyclopedia of Genes and Genomes sets from the high- and low-NRGPS groups were used for GSEA to investigate the differences in biological characteristics between the two groups. High NRGPS enriched pathways included "cell cycle," "spliceosome," "ecm receptor interaction," "DNA replication," and "ribosome." The enriched pathways in the low NRGPS group were "metabolism of cytochrome P450," "metabolism of cytochrome P450 xenogeneic organisms," "retinol metabolism," "complement," "steroid hormone biosynthesis," "valine leucine," and "isoleucine degradation" (Figures 6A, B). These results suggest that high NRGPS expression is associated with tumor progression and metastasis. We combined the NRGPS with clinical variables to construct a nomogram to be more suitable for clinical application. Receiver operating characteristic curves showed that the nomogram had a good predictive performance for the 1-, 1.5-, and 2-year overall survival of patients with HBV-HCC (Figures 6C, D).

3.4 Applicability of the HCC model in HBV-HCC data

The results showed that the prognostic model for HCC constructed using the available literature did not predict the survival time of patients with HBV-HCC (Figure S1). In contrast, the NRGPS had a better predictive ability.

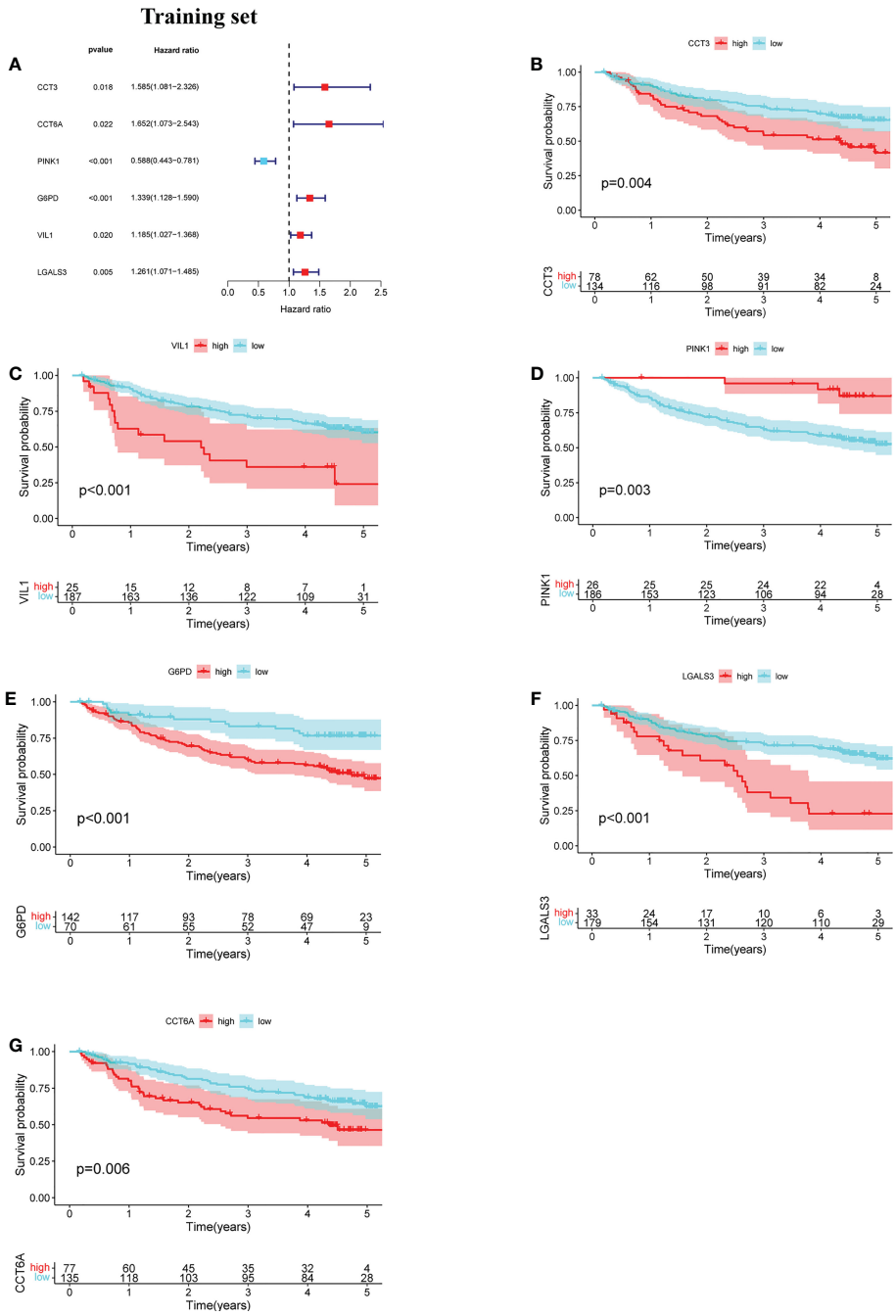


FIGURE 3
Development of the NRGPS System. **(A)** The forest map of 6 prognostic genes were obtained by Univariable Cox regression analysis. **(B–G)** Kaplan-Meier curve of 6 prognostic genes.

3.5 Drug sensitivity analysis

The differences in drug sensitivity between high and low NRGPS and the two- and three-dimensional structures of the drugs are shown in Figure S2.

3.6 Analysis of immune microenvironment

Using the ssGSEA algorithm, the infiltrating states of 16 types of immune cells and the activities of 13 types of immune-related functions in the GSE83148, GSE84044, GSE14520, and TCGA cohorts were studied. The results showed that regulatory T cells (Tregs), activated dendritic cells, immature dendritic cells, macrophages, Th2 cells, and neutrophils were significantly increased in patients with chronic HBV infection, and high fibrosis score and NRGPS were observed in the GSE83148, GSE84044, and GSE14520 cohorts (Figures 7A–C). The same trend was observed for immune checkpoints and cytokine-cytokine receptors (CCR) (Figures 7D–F). High expression of Tregs, macrophages, and Th2 cells are often associated with a

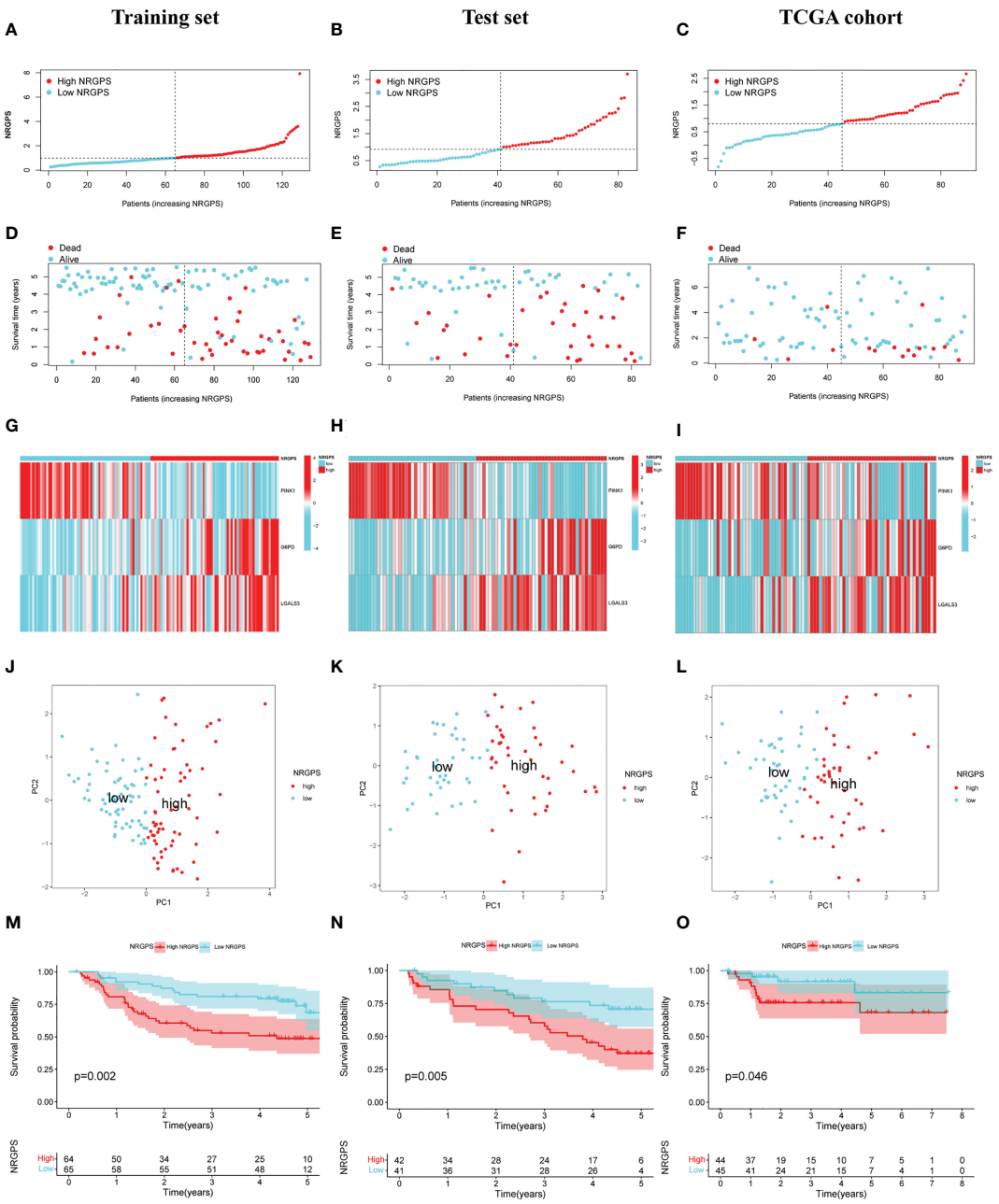


FIGURE 4
Validation of the NRGPS System. **(A–C)** The graphs of Training set, Test set and TCGA cohort. **(D–F)** The scatter plots of Training, Test set and TCGA cohort. **(G–I)** The expression of 3 DENRGs in high and low NRGPS. **(J–L)** PCA analysis in Training set, Test set and TCGA cohort. **(M–O)** Kaplan-meier survival analysis in Training set, Test set and TCGA cohort. NRGPS, necroptosis-related gene prognostic score; DENRGs, differentially expressed necroptosis-related genes.

poor tumor prognosis. We also analyzed the relationship between infiltrating immune cells and the immune function of the three model genes in the immune microenvironment of the GSE83148, GSE84044, GSE14520, and TCGA cohorts. The results showed that LGALS3 was more associated with increased infiltration of immune cells and expression of immune pathway compared to G6PD and PINK1.

Furthermore, LGALS3 expression was significantly and positively correlated with the expression of CCR and immune checkpoint genes, as well as Treg cell infiltration (Figures 8A–C). As important components of the CCR, chemokines and chemokine receptors regulate cell migration, adhesion, localization, and intercellular interactions and are highly involved in tumor development. Among them, CCR2-CCL2, CXCR4-CXCL12, and

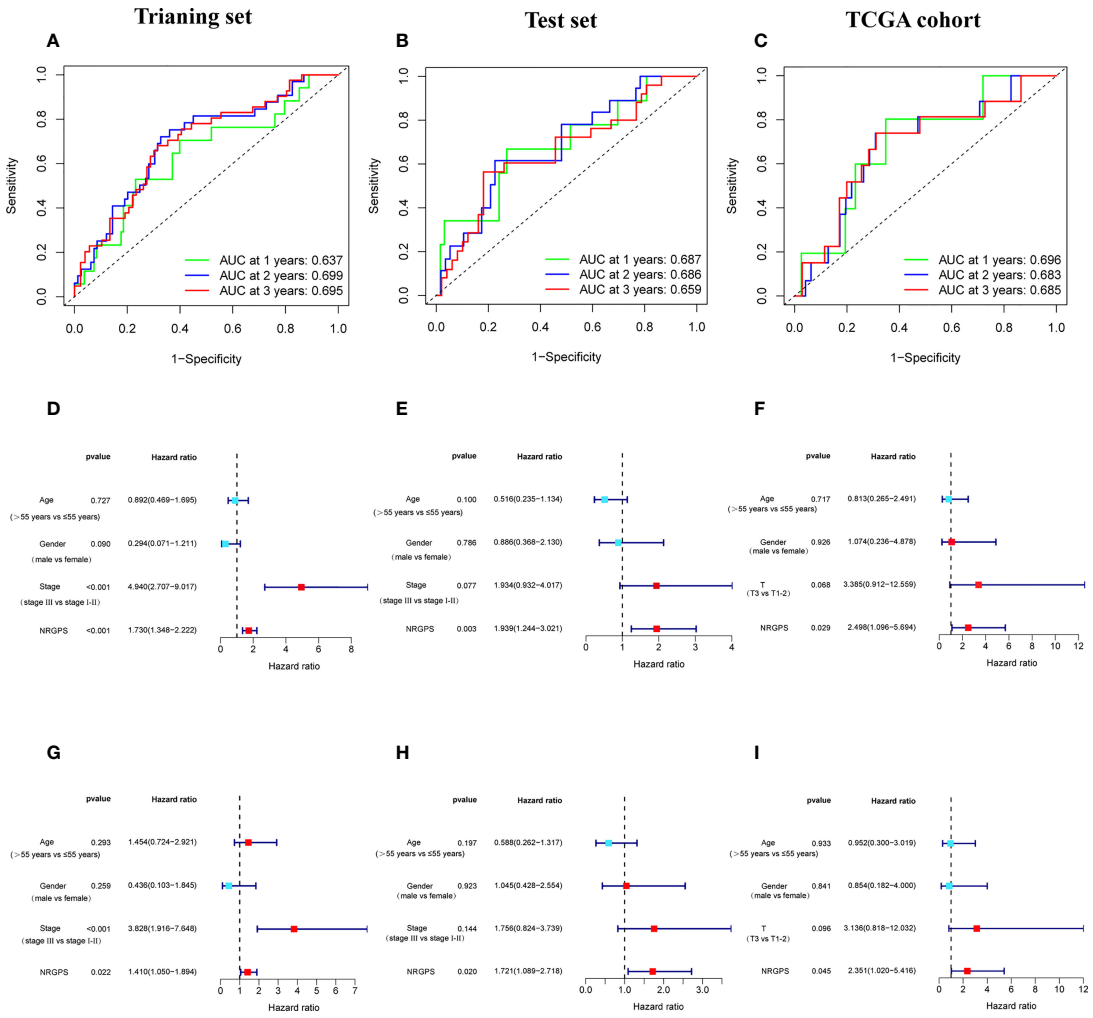


FIGURE 5
Validation of the NRGPS System. **(A–C)** The ROC curve of Training set, Test set and TCGA cohort verifies the prediction ability of this prediction model. **(D–I)** The Univariable and Multivariable COX regression analyses results of NRGPS.

CCR6-CCL20, three pairs of chemokines and their receptors, contribute to forming the inhibitory immune microenvironment and have an obvious effect on promoting tumor progression. Further analysis of the correlation between LGALS3 and chemokines and their receptors showed that LGALS3 expression was significantly and positively correlated with the expression of CCR6 and CCL20 in the CHI, HBV-HF, and HBV-HCC microenvironments (Figures 8D–F). Figure S3 shows analysis of the TCGA cohort immune microenvironment.

3.7 Changes in model genes with disease progression

LGALS3 and G6PD were highly expressed in the CHI, S2-S4, and tumor groups, and PINK1 was highly expressed in the normal, S0-S1 and paracancerous tissues (Figures 9A–C). Among the PBMCs, PINK1 was the most expressed in the normal group (Figure 9D). Compared with the normal group, LGALS3 and G6PD were significantly upregulated in all stages of HBV infection (Figures 9E, F).

3.8 Expression of CCR6 and FOXP3 in PBMCs in different patients

The expression of CCR6 and FOXP3 genes was significantly increased after HBV infection (Figures S4A–B); however, there was no significant correlation with the expression of LGALS3 in PBMCs (Figures S4C–H).

3.9 Analysis of transfection efficiency

The transfection efficiency of HEPG2 was 61.38 ± 2.44% and that of HUH7 was 81.85 ± 2.68% (Figure S5A). ELISA results showed that HEPG2 and HUH7 cells produced HBsAg after transfection with pAAV/HBV1.2$_{C2}$ (Figure S5B).

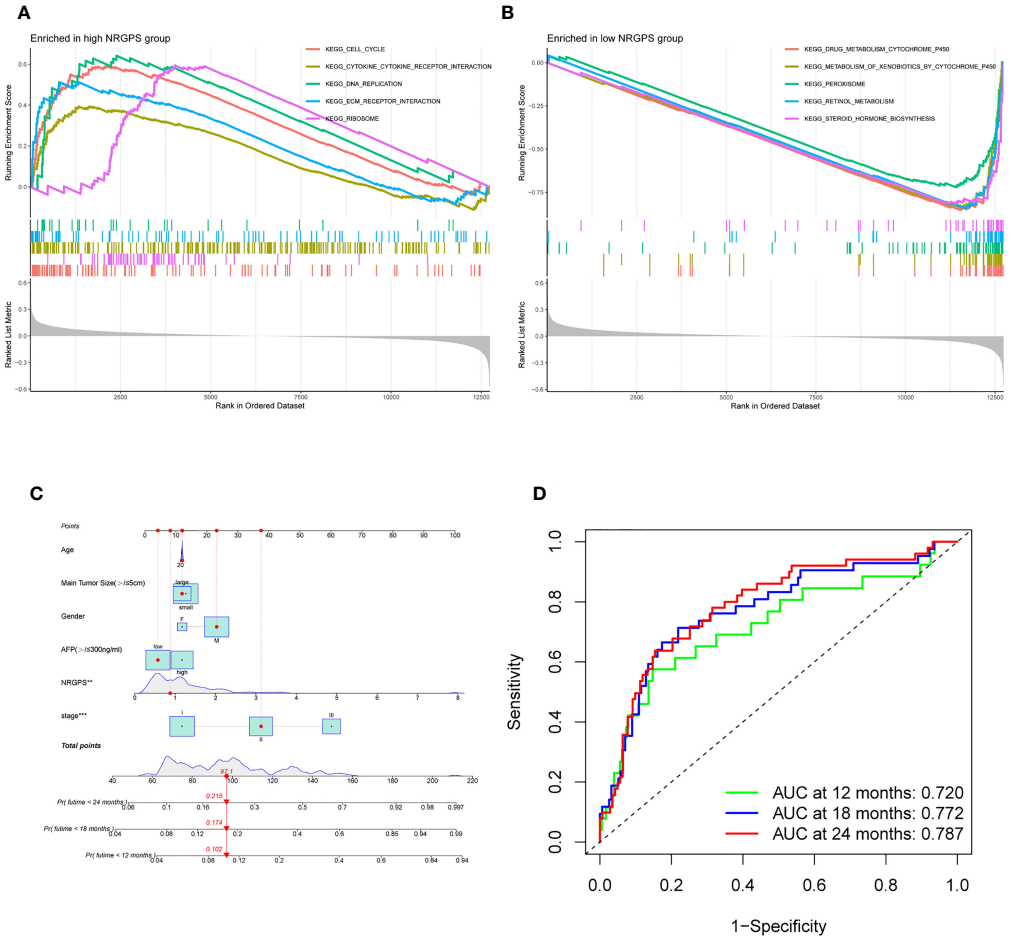


FIGURE 6
GSEA analysis and Nomogram. **(A)** GSEA in the high-NRGPS group. **(B)** GSEA in the low-NGPS group. **(C)** Nomogram. **(D)** The ROC curve of Nomogram.

3.10 Verification of gene expression

qRT-PCR was used to detect the expression of each gene in the cells. After transfection with pAAV-MCV and pAAV/HBV1.2$_{C2}$, G6PD expression in HEPG2 was significantly increased, whereas PINK1 and LGALS3 expression showed no significant changes (Figures 10A–C). Compared to pAAV-MCV plasmid transfection, the expression of the G6PD gene was increased, and PINK1 gene expression was decreased in HUH7 cells transfected with pAAV/HBV1.2$_{C2}$; however, there was no significant difference in LGALS3 expression (Figures 10D–F). The expression level of the G6PD gene in HEPG2$^{\#}$ and HUH7$^{\#}$ was significantly higher than that in LX2 (Figure 10G), the expression level of the PINK1 gene in LX2 was significantly higher than that in HEPG2$^{\#}$ and HUH7$^{\#}$ (Figure 10H), and the expression level of the LGALS3 gene in HEPG2$^{\#}$ was significantly higher than that in LX2 (Figure 10I).

3.11 Analysis of *in vitro* LGALS3 function

After LGALS3 knockdown in the HEPG2$^{\#}$ and HUH7$^{\#}$ cell lines, LGALS3 gene expression was downregulated and CCL20 expression was also significantly decreased (Figures 11A, B). Moreover, the results of the CCK8 cell proliferation assay showed that the proliferation of cancer cells decreased significantly after LGALS3 knockdown (Figure 11C), and the results of the transwell assay showed that LGALS3 knockdown significantly affected the migratory ability of cancer cells (Figures 11D, E).

4 Discussion

NRGs play a vital role in the development of HCC, and many studies have used them to establish risk-prediction models for prognosis prediction of patients with HCC (10–15). However, currently established prognostic models using NRGs for HCC are not applicable to patients with HBV-HCC, and the role of NRGs in the adverse progression of chronic HBV infection remains unclear. In this study, we systematically analyzed the expression level and correlation of NRGs in HBV-HCC and paracarcinoma tissues. Gene ontology analysis revealed 24 DENRGs that were enriched for pathways related to the proliferation and migration of tumor cells (17, 18). In addition to HBV-HCC tissues, LGALS3 expression was also upregulated in CHI tissues and tissues with high fibrosis

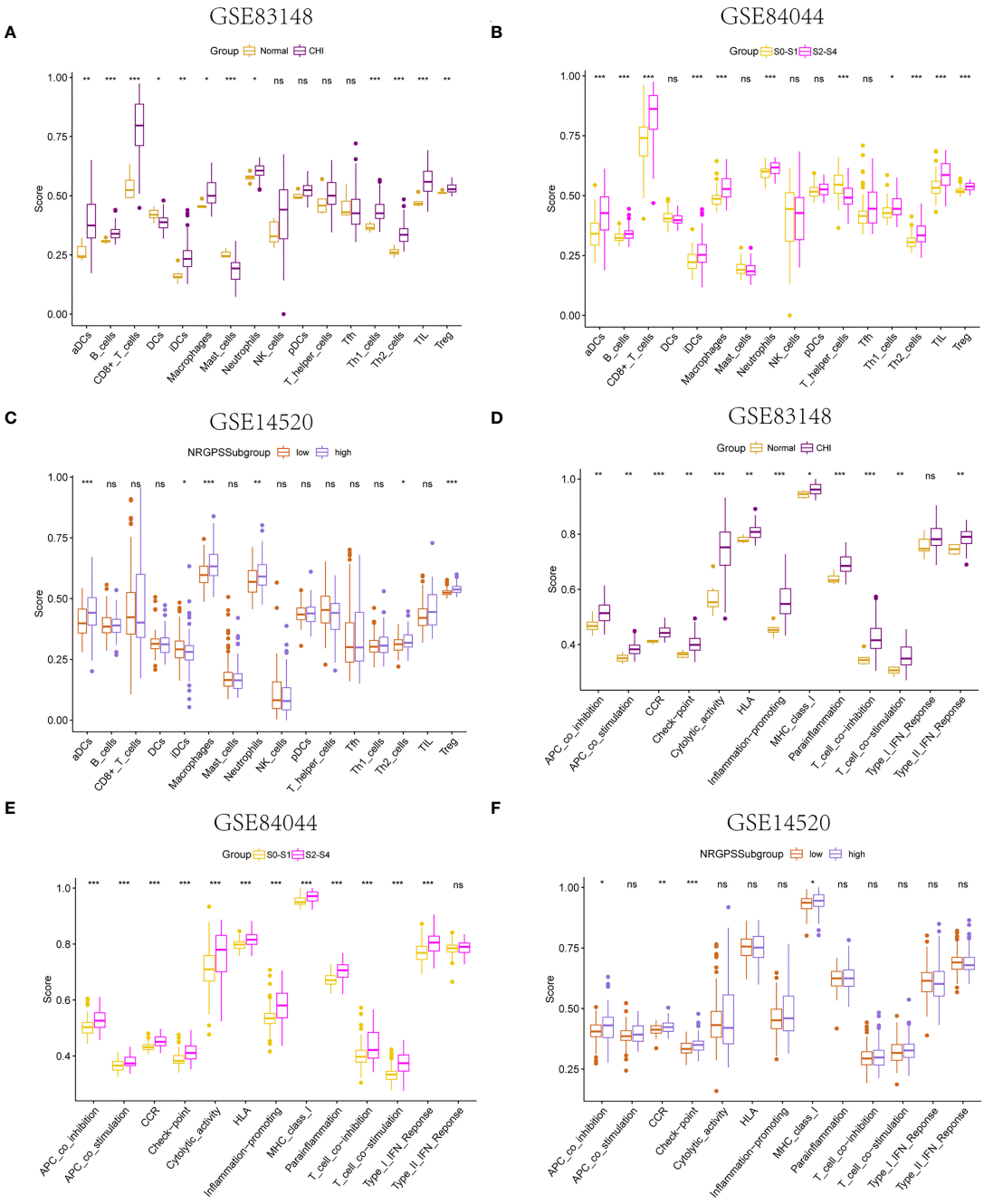


FIGURE 7
Immune microenvironment analysis. The boxplot of 16 immune cell differences in **(A)** GSE83148, **(B)** GSE84044, **(C)** GSE14520. The boxplot of 13 immune signaling pathway differences in **(D)** GSE83148, **(E)** GSE84044, **(F)** GSE14520. *P <0.05, **P <0.01, ***P < 0.001, ns, not significant

scores. Subsequently, univariable and multivariable Cox regression analyses established a prognostic model comprising G6PD, PINK1, and LGALS3. Univariable and multivariable Cox, Kaplan–Meier, and receiver operating characteristic curve analyses all indicated that the NRGPS had good predictive performance and could be used as an independent prognostic indicator for patients with HBV-HCC. The gene expression and predictive power of the model were verified in our constructed HBV-HCC cell model and TCGA dataset.

LGALS3, also known as Galectin-3, encodes the carbohydrate-binding protein Galectin-3, which is a member of the galactoagglutinin family. It is mainly localized in the cytoplasm but is also expressed in the nucleus, cell surface, and extracellular (19). Galectin-3 expression in peripheral blood and liver is associated with the progression of chronic and acute liver failure, liver fibrosis, HCC, and other liver diseases (20–22). Moreover, our experimental data revealed persistent upregulation of the LGALS3 gene in the HBV-infected group. It has been suggested that the

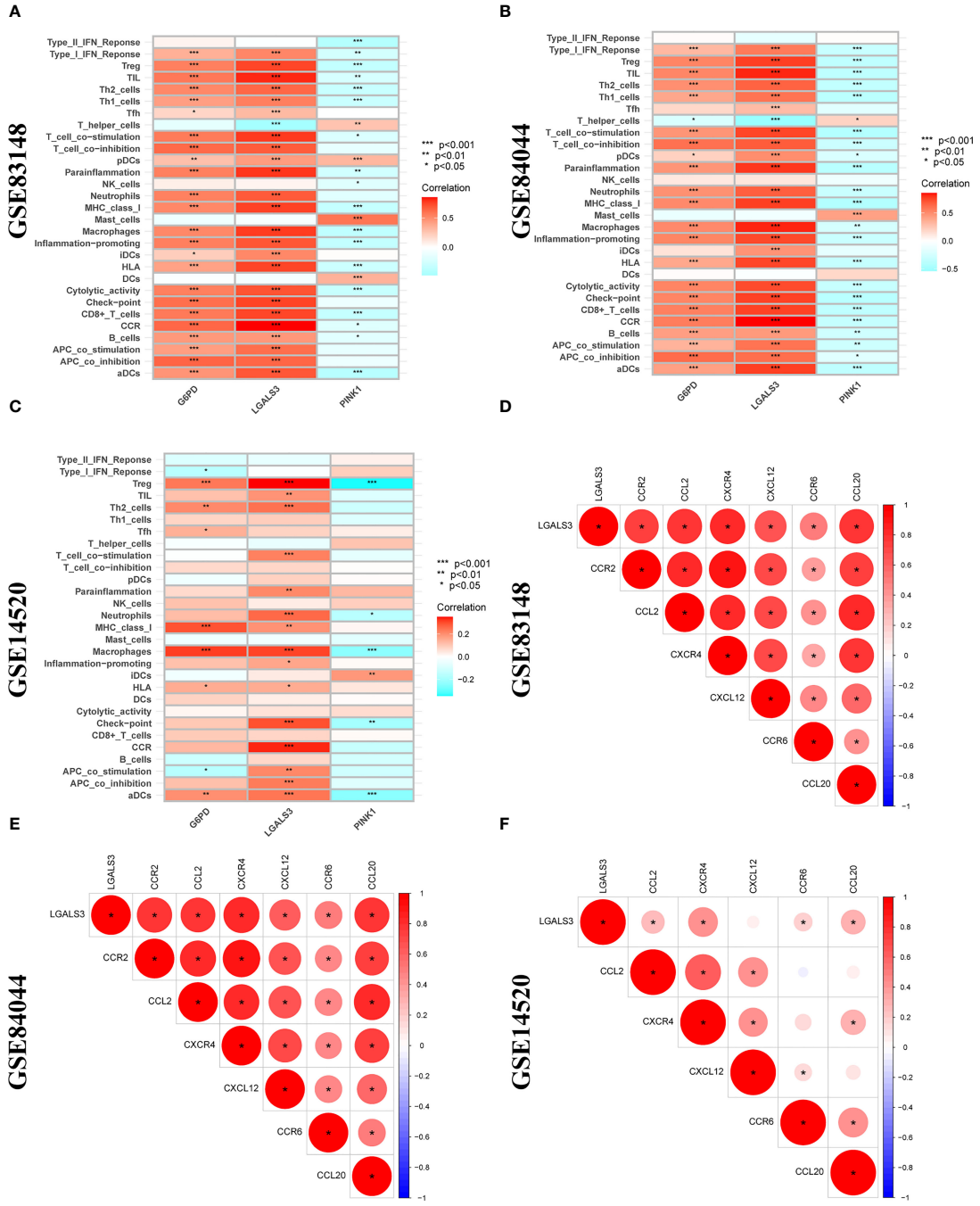


FIGURE 8
Correlation analysis between 3 DENRGs and immune microenvironment. The relationship between 3 DENRGs and immune microenvironment in **(A)** GSE83148, **(B)** GSE84044, **(C)** GSE14520. Correlation analysis of 3 DENRGs with chemokines and chemokine receptors in **(D)** GSE83148, **(E)** GSE84044, **(F)** GSE14520.

HBV-X protein produced by HBV in hepatocytes can transactivate the Galectin-3 promoter or upregulate Galectin-3 expression through the CREB/ATF-transcription pathway after HBV infection, and this phenomenon is more obvious in normal liver cells than in hepatoma cells (23). As reported in the literature, LGALS3 expression in HEPG2 and HUH7 cells did not change significantly after transfection with the HBV plasmid, which results from the high expression of LGALS3 in HCC cells that reduces the influence of HBV on LGALS3 expression.

The changes in the immune microenvironment are closely related to the occurrence and progression of the disease. Galectin-3 plays a role in many cellular functions, including apoptosis, innate immunity, and T-cell regulation, and is a vital component of the immune microenvironment (24, 25). HBV has multiple immunosuppressive effects, a key factor in the progression of HBV infection from chronic infection to HCC (26). In ssGSEA analysis, we found that the expression of Treg cells and immune checkpoints was upregulated in the chronic HBV infection, high

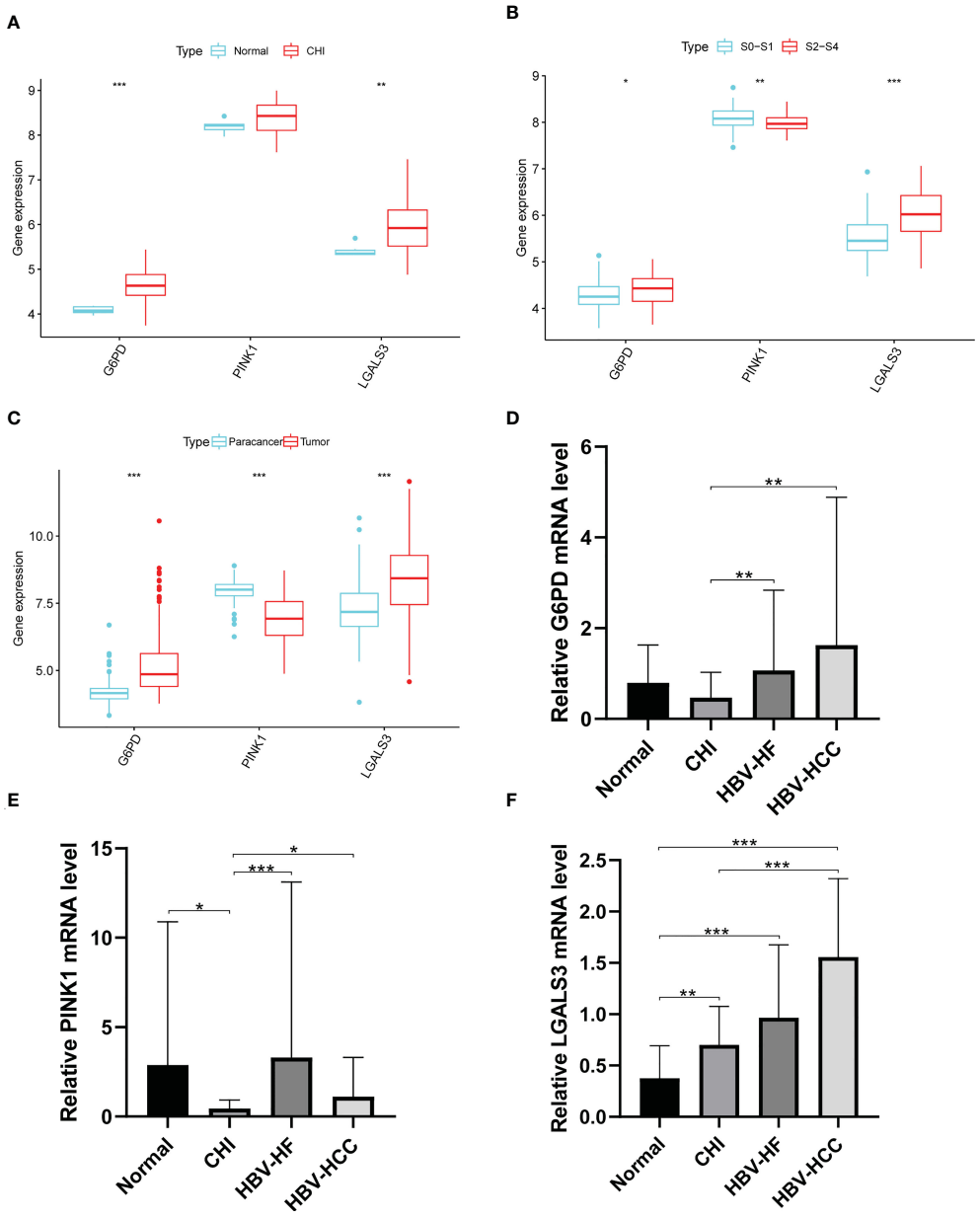


FIGURE 9
Expression of G6PD,PINK1 and LGALS3 in disease progression. Expression of G6PD, PINK1 and LGALS3 in **(A)** GSE83148, **(B)** GSE84044 and **(C)** GSE14520. **(D–F)**, The expression of G6PD, PINK1 and LGALS3 in PBMCs. *P <0.05, **P <0.01, ***P < 0.001, ns, not significant.

fibrosis score, and high-NRGPS groups. Additionally, our population analysis results showed that the expression of FOXP3, a key transcription factor of Treg cells, was higher in the experimental group than in the control group. Tregs can inhibit the proliferation of CD8+T cells and the production of granzyme A and B and perforin, ultimately leading to a decline in the normal immune defense and surveillance functions of CD8+T cells in the microenvironment (27, 28). High infiltration of Tregs and upregulation of immune checkpoints promoted the formation of an immunosuppressive microenvironment (6, 29), which is conducive to the persistency of the HBV infection and to the escape of tumor cells from the surveillance of the immune system.

Further analysis of the immune microenvironment showed that high expression of LGALS3 was positively correlated with immune checkpoint gene expression, CCL20, CCR6, and Treg cells in patients with CHI, HBV-HF, or HBV-HCC. Chemokines are essential for immune cell transport and promote the recruitment of immune cells to immunoreactive sites during inflammation (30). Currently, it is known that CCL20 exclusively combines with CCR6 to form the CCR6-CCL20 axis, involved in regulating immune system homeostasis (31). CCL20 is a cytokine that can promote Treg cell infiltration, and the CCR6-CCL20 axis regulated Tregs migrate into the tumor microenvironment, thereby leading to tumor progression and poor prognosis in patients with HCC (32–

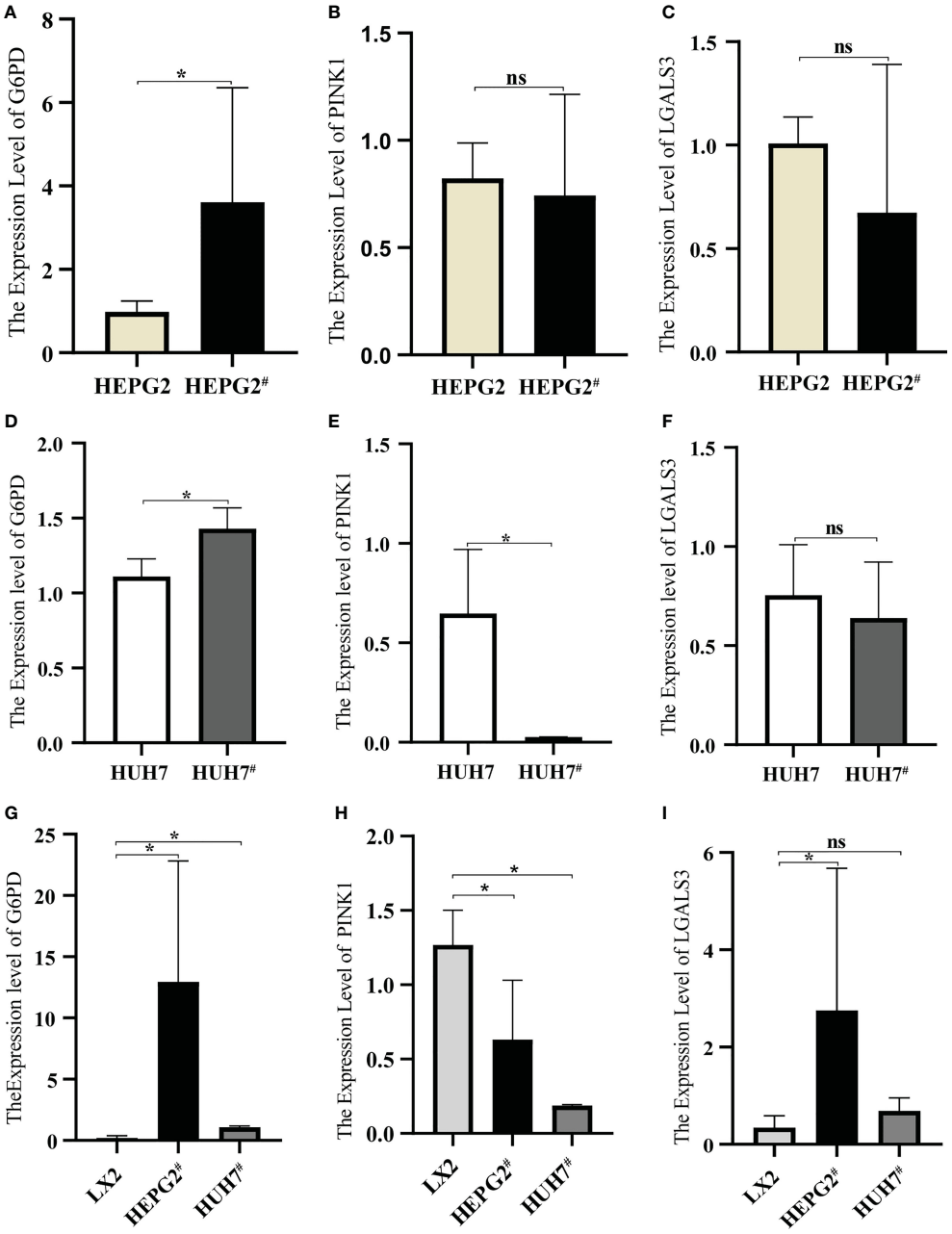


FIGURE 10
Expression of model genes in hepatocellular carcinoma cell lines. **(A)** Expression of G6PD in HEPG2 and HEPG2#. **(B)** Expression of PINK1in HEPG2 and HEPG2#. **(C)** Expression of LGALS3 in HEPG2 and HEPG2#. **(D)** Expression of G6PD in HUH7 and HUH7#. **(E)** Expression of PINK1in HUH7 and HUH7#, **(F)** Expression of LGALS3 in HUH7 and HUH7#. **(G)** Expression of G6PD in LX2, HEPG2# and HUH7#. **(H)** Expression of PINK1in LX2, HEPG2# and HUH7#. **(I)** Expression of LGALS3 in LX2, HEPG2# and HUH7#. **P* <0.05, ns, not significant.

35). Moreover, the oversecretion of CCL20 by myoblasts in cirrhotic HCC promotes the production of HCC by regulating aerobic glycolysis through the CCR6-receptor and the ERK/PKM2-signaling pathway (36). All of these highlight the important role of CCL20 in disease worsening after HBV infection. Although there was no significant correlation between LGALS3 and the expression of CCR6 and FOXP3 in PBMCs, the expression of CCL20 in HEPG2# and HUH7# decreased significantly after LGALS3 knockdown. LGALS3 may be involved in the formation of an immunosuppressive microenvironment by influencing the expression of CCL20, leading to adverse disease progression. In addition, LGALS3 is associated with the metabolism of HCC and lymph node metastasis, which is a key regulatory factor for tumor cell proliferation and migration (37, 38). Furthermore, the proliferation and migration of HEPG2# and HUH7# cells decreased significantly after the LGALS3 knockdown.

In conclusion, our findings identified the important role of a key gene, LGALS3, in disease progression after HBV persistence infection. However, this study has some limitations. First, we could not collect HBV-HCC liver tissue samples to verify the

10.3389/fimmu.2023.1142319

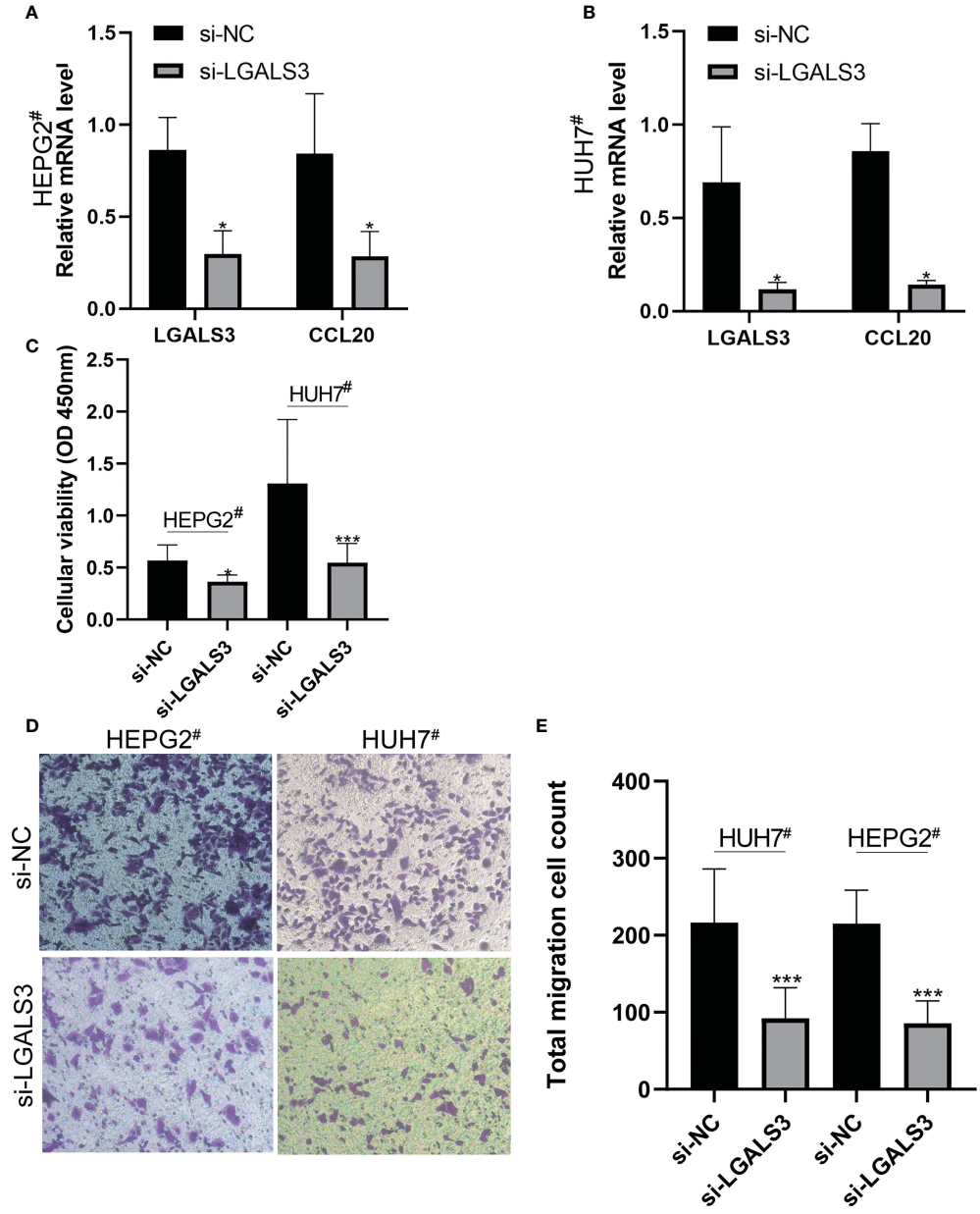


FIGURE 11
LGALS3 knockdown affects the proliferation and migration of hepatocellular carcinoma cells and the expression of CCL20. **(A)** Changes of CCL20 before and after LGALS3 knockdown in HEPG2#. **(B)** Changes of CCL20 before and after LGALS3 knockdown in HUH7#. **(C)** Cell proliferation before and after knockdown was detected by CCK8. **(D, E)** Transwell assay showed that cell migration ability decreased after LGALS3 knockdown. *P <0.05, ***P <0.001.

predictive power of our model. Second, the prognostic model established in this study still needs to be further verified in multicenter, large-scale clinical studies.

5 Conclusion

This study successfully constructed a prognostic model for HBV-HCC comprising G6PD, PINK1, and LGALS3, and analyzed the key role of LGALS3 in adverse disease progression after HBV persistence infection. Moreover, LGALS3 was demonstrated to be a potential therapeutic target for the adverse progression of HBV persistence infection.

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 the Second Affiliated Hospital of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JD and RZ designed the study. YX and KZ wrote and revised the manuscript. XJ and JL helped to perform the statistical analysis. MG and XC downloaded and collated the data. SZ was responsible for supervising the study. All authors contributed to the article and approved the submitted version.

Funding

The present study was supported by Special Fund for Clinical Research of Wu Jieping Medical Foundation (320.6750.18230).

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/fimmu.2023.1142319/full#supplementary-material

References

1. Iannacone M, Guidotti LG. Immunobiology and pathogenesis of hepatitis b virus infection. *Nat Rev Immunol* (2022) 22(1):19–32. doi: 10.1038/s41577-021-00549-4

2. Gong J, Chen Y, Cao J, Wang Y, Chen J, Li D, et al. The immune landscape of hepatitis b virus-related acute liver failure by integration analysis. *J Immunol Res* (2022) 2022:6764379. doi: 10.1155/2022/6764379

3. Murphy JM, Czabotar PE, Hildebrand JM, Lucet IS, Zhang JG, Alvarez-Diaz S, et al. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity* (2013) 39(3):443–53. doi: 10.1016/j.immuni.2013.06.018

4. Zhang X, Fan C, Zhang H, Zhao Q, Liu Y, Xu C, et al. MLKL and FADD are critical for suppressing progressive lymphoproliferative disease and activating the NLRP3 inflammasome. *Cell Rep* (2016) 16(12):3247–59. doi: 10.1016/j.celrep.2016.06.103

5. Nailwal H, Chan FK. Necroptosis in anti-viral inflammation. *Cell Death Differ* (2019) 26(1):4–13. doi: 10.1038/s41418-018-0172-x

6. Demetriou M, Granovsky M, Quaggin S, Dennis JW. Negative regulation of T-cell activation and autoimmunity by Mgat5 n-glycosylation. *Nature* (2001) 409(6821):733–9. doi: 10.1038/35055582

7. Kim HR, Lin HM, Biliran H, Raz A. Cell cycle arrest and inhibition of anoikis by galectin-3 in human breast epithelial cells. *Cancer Res* (1999) 59(16):4148–54.

8. An Y, Xu S, Liu Y, Xu X, Philips CA, Chen J, et al. Role of galectins in the liver diseases: A systematic review and meta-analysis. *Front Med (Lausanne)* (2021) 8:744518. doi: 10.3389/fmed.2021.744518

9. Kong F, Jin M, Cao D, Jia Z, Liu Y, Jiang J. Galectin-3 not galectin-9 as a candidate prognosis marker for hepatocellular carcinoma. *PeerJ* (2020) 8:e9949. doi: 10.7717/peerj.9949

10. Yang H, Jiang Q. A multi-omics-based investigation of the immunological and prognostic impact of necroptosis-related genes in patients with hepatocellular carcinoma. *J Clin Lab Anal* (2022) 36(4):e24346. doi: 10.1002/jcla.24346

11. Chen J, Wang H, Zhou L, Liu Z, Chen H, Tan X. A necroptosis-related gene signature for predicting prognosis, immune landscape, and drug sensitivity in hepatocellular carcinoma. *Cancer Med* (2022) 11(24):5079–96. doi: 10.1002/cam4.4812

12. Gao X, Huang D, Li SG, Wang WX, Sun DL, Qian JM, et al. Identification and validation of prognosis-related necroptosis genes for prognostic prediction in hepatocellular carcinoma. *J Oncol* (2022) 2022:3172099. doi: 10.1155/2022/3172099

13. Ren H, Zheng J, Cheng Q, Yang X, Fu Q. Establishment of a necroptosis-related prognostic signature to reveal immune infiltration and predict drug sensitivity in hepatocellular carcinoma. *Front Genet* (2022) 13:900713. doi: 10.3389/fgene.2022.900713

14. Xiang Z, Mranda GM, Zhou X, Xue Y, Wang Y, Wei T, et al. Identification and validation of the necroptosis-related gene signature related to prognosis and tumor immune in hepatocellular carcinoma. *Med (Baltimore)* (2022) 101(36):e30219. doi: 10.1097/MD.0000000000030219

15. Li J, Wu Z, Wang S, Li C, Zhuang X, He Y, et al. A necroptosis-related prognostic model for predicting prognosis, immune landscape, and drug sensitivity in hepatocellular carcinoma based on single-cell sequencing analysis and weighted co-expression network. *Front Genet* (2022) 21:984297. doi: 10.3389/fgene.2022.984297

16. Huang LR, Wu HL, Chen PJ, Chen DS. An immunocompetent mouse model for the tolerance of human chronic hepatitis b virus infection. *Proc Natl Acad Sci U.S.A.* (2006) 103(47):17862–7. doi: 10.1073/pnas.0608578103

17. Yao D, Lin S, Chen S, Wang Z. circHIPK3 regulates cell proliferation and migration by sponging microRNA-124 and regulating serine/threonine kinase 3 expression in esophageal squamous cell carcinoma. *Bioengineered* (2022) 13(4):9767–80. doi: 10.1080/21655979.2022.2060776

18. Fan F, Roszik J, Xia L, Ghosh S, Wang R, Ye X, et al. Cytospin-a regulates colorectal cancer cell division and migration by modulating stability of microtubules and actin filaments. *Cancers (Basel)* (2022) 14(8):1977. doi: 10.3390/cancers14081977

19. Dumic J, Dabelic S, Flögel M. Galectin-3: an open-ended story. *Biochim Biophys Acta* (2006) 1760(4):616–35. doi: 10.1016/j.bbagen.2005.12.020

20. Bai L, Lu W, Tang S, Tang H, Xu M, Liang C, et al. Galectin-3 critically mediates the hepatoprotection conferred by M2-like macrophages in ACLF by inhibiting pyroptosis but not necroptosis signalling. *Cell Death Dis* (2022) 13(9):775. doi: 10.1038/s41419-022-05181-1

21. Traber PG, Chou H, Zomer E, Hong F, Klyosov A, Fiel MI, et al. Regression of fibrosis and reversal of cirrhosis in rats by galectin inhibitors in thioacetamide-induced liver disease. *PloS One* (2013) 8(10):e75361. doi: 10.1371/journal.pone.0075361

22. Setayesh T, Colquhoun SD, Wan YY. Overexpression of galectin-1 and galectin-3 in hepatocellular carcinoma. *Liver Res* (2020) 4(4):173–9. doi: 10.1016/j.livres.2020.11.001

23. Hsu DK, Dowling CA, Jeng KC, Chen JT, Yang RY, Liu FT. Galectin-3 expression is induced in cirrhotic liver and hepatocellular carcinoma. *Int J Cancer* (1999) 81(4):519–26. doi: 10.1002/(sici)1097-0215(19990517)81:4<519::aid-ijc3>3.0.co;2-0

24. Ruvolo PP. Galectins as regulators of cell survival in the leukemia niche. *Adv Biol Regul* (2019) 71:41–54. doi: 10.1016/j.jbior.2018.09.003

25. Shi Y, Tang D, Li X, Xie X, Ye Y, Wang L. Galectin family members: Emerging novel targets for lymphoma therapy? *Front Oncol* (2022) 12:889034. doi: 10.3389/fonc.2022.889034

26. Loomba R, Liang TJ. Hepatitis b reactivation associated with immune suppressive and biological modifier therapies: Current concepts, management strategies, and future directions. *Gastroenterology* (2017) 152(6):1297–309. doi: 10.1053/j.gastro.2017.02.009

27. Yang G, Liu A, Xie Q, Guo TB, Wan B, Zhou B, et al. Association of CD4+CD25+Foxp3+ regulatory T cells with chronic activity and viral clearance in patients with hepatitis b. *Int Immunol* (2007) 19(2):133–40. doi: 10.1093/intimm/dxl130

28. Fu J, Xu D, Liu Z, Shi M, Zhao P, Fu B, et al. Increased regulatory T cells correlate with CD8 T-cell impairment and poor survival in hepatocellular carcinoma patients. *Gastroenterology* (2007) 132(7):2328–39. doi: 10.1053/j.gastro.2007.03.102

29. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* (2015) 15(8):486–99. doi: 10.1038/nri3862

30. Sallusto F, Baggiolini M. Chemokines and leukocyte traffic. *Nat Immunol* (2008) 9(9):949–52. doi: 10.1038/ni.f.214

31. Lee AY, Phan TK, Hulett MD, Körner H. The relationship between CCR6 and its binding partners: does the CCR6-CCL20 axis have to be extended? *Cytokine* (2015) 72(1):97–101. doi: 10.1016/j.cyto.2014.11.029

32. Yang P, Markowitz GJ, Wang XF. The hepatitis b virus-associated tumor microenvironment in hepatocellular carcinoma. *Natl Sci Rev* (2014) 1(3):396–412. doi: 10.1093/nsr/nwu038

33. Chen KJ, Lin SZ, Zhou L, Xie HY, Zhou WH, Taki-Eldin A, et al. Selective recruitment of regulatory T cell through CCR6-CCL20 in hepatocellular carcinoma fosters tumor progression and predicts poor prognosis. *PloS One* (2011) 6(9):e24671. doi: 10.1371/journal.pone.0024671

34. Shen X, Li N, Li H, Zhang T, Wang F, Li Q. Increased prevalence of regulatory T cells in the tumor microenvironment and its correlation with TNM stage of hepatocellular carcinoma. *J Cancer Res Clin Oncol* (2010) 136(11):1745–54. doi: 10.1007/s00432-010-0833-8

35. Stross L, Günther J, Gasteiger G, Asen T, Graf S, Aichler M, et al. Foxp3+ regulatory T cells protect the liver from immune damage and compromise virus control during acute experimental hepatitis b virus infection in mice. *Hepatology* (2012) 56 (3):873–83. doi: 10.1002/hep.25765

36. Yuan Q, Zhang J, Liu Y, Chen H, Liu H, Wang J, et al. MyD88 in myofibroblasts regulates aerobic glycolysis-driven hepatocarcinogenesis *via* ERK-dependent PKM2 nuclear relocalization and activation. *J Pathol* (2022) 256(4):414–26. doi: 10.1002/path.5856

37. Chen X, Yu C, Liu X, Liu B, Wu X, Wu J, et al. Intracellular galectin-3 is a lipopolysaccharide sensor that promotes glycolysis through mTORC1 activation. *Nat Commun* (2022) 13(1):7578. doi: 10.1038/s41467-022-35334-x

38. Wang X, Yuegao, Bai L, Ibrahim MM, Ma W, Zhang J, et al. Evaluation of annexin A7, galectin-3 and gelsolin as possible biomarkers of hepatocarcinoma lymphatic metastasis. *BioMed Pharmacother* (2014) 68(3):259–65. doi: 10.1016/j.biopha.2013.12.009

Frontiers in Immunology

Explores novel approaches and diagnoses to treat immune disorders.

The official journal of the International Union of Immunological Societies (IUIS) and the most cited in its field, leading the way for research across basic, translational and clinical immunology.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact