frontiers | Research Topics

Insights in cancer genetics and oncogenomics 2022

Edited by
Anton A. Buzdin, Tugba Önal-Süzek and Xin Hong

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-3717-6
DOI 10.3389/978-2-8325-3717-6

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

Insights in cancer genetics and oncogenomics: 2022

Topic editors

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

Tugba Önal-Süzek — Mugla University, Türkiye

Xin Hong — Southern University of Science and Technology, China

Citation

Buzdin, A. A., Önal-Süzek, T., Hong, X., eds. (2023). *Insights in cancer genetics and oncogenomics: 2022*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3717-6

Table of contents

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fgene.2022.966896

OPEN ACCESS

EDITED BY
Xin Hong,
Southern University of Science and Technology, China

REVIEWED BY
Yizhe Wang,
Cornell University, United States
Chao Ma,
First Affiliated Hospital of Zhengzhou University, China
Ting Xiao,
Nankai University, China

*CORRESPONDENCE
Hong Ren,
renh_med@163.com
Boxiang Zhang,
zhangboxiang@xjtufh.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 11 June 2022
ACCEPTED 30 August 2022
PUBLISHED 15 September 2022

CITATION
Wu J, Song D, Zhao G, Chen S, Ren H and Zhang B (2022), Cross-talk between necroptosis-related lncRNAs to construct a novel signature and predict the immune landscape of lung adenocarcinoma patients. *Front. Genet.* 13:966896. doi: 10.3389/fgene.2022.966896

Cross-talk between necroptosis-related lncRNAs to construct a novel signature and predict the immune landscape of lung adenocarcinoma patients

Jie Wu[1†], Dingli Song[1†], Guang Zhao[1], Sisi Chen[2], Hong Ren[1*] and Boxiang Zhang[1*]

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2]Department of Oncology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Background: As a new style of cell death, necroptosis plays a crucial role in tumor immune microenvironment. LncRNAs have been identified to act as competitive RNAs to influence genes involved in necroptosis. Therefore, we aim to create a signature based on necroptosis-related lncRNAs to predict the prognosis and immune landscape of lung adenocarcinoma (LUAD) patients in this study.

Methods: TCGA database was used to acquire RNA sequencing (RNA-Seq) data and clinical information for 59 lung normal samples and 535 lung adenocarcinoma samples. The Pearson correlation analysis, univariate cox regression analysis and least absolute shrinkage and selection operator (LASSO) cox regression were performed to construct the prognostic NRlncRNAs signature. Then we used Kaplan-Meier (K-M) analysis, time-dependent ROC curves, univariate and multivariate cox regression analysis, and nomogram to validate this signature. In addition, GO, KEGG, and GSVA were analyzed to investigate the potential molecular mechanism. Moreover, we analyzed the relationship between our identified signature and immune microenvironment, TMB, and some clinical characteristics. Finally, we detected the expression of the six necroptosis-related lncRNAs in cells and tissues.

Results: We constructed a NRlncRNAs signature consisting of six lncRNAs (FRMD6-AS1, LINC01480, FAM83A-AS1, FRMD6-AS1, MED4-AS1, and LINC01415) in LUAD. LUAD patients with high risk scores had lower chance

Abbreviations: LUAD, lung adenocarcinoma; RIPK1/RIPK3, receptor-interacting protein kinases1/3; MLKL, mixed lineage kinase domain like pseudokinase; lncRNAs, long non-coding RNAs; ncRNAs non-proteincoding RNAs; NRGs, necroptosis-related genes; NRlncRNAs, necroptosis-related lncRNAs; TIME, tumor immune microenvironment; TCGA, The Cancer Genome Atlas; GRCh38, the genome reference consortium human build 38; LASSO, the least absolute shrinkage and selection operator; OS, overall survival; K-M curves, Kaplan-Meier curves; DCA, decision curve analysis; ROC curve, receiver operating characteristic curve; AUC, area under curve; GSVA, gene set variant analysis; GO, Gene Oncology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TMB, tumor mutation burden; ATCC, American Type Culture Collection; RT-qPCR, quantitative real-time PCR.

of survival with an AUC of 0.739, 0.709, and 0.733 for 1-year, 3-year, and 5-year respectively. The results based on GO, KEGG, and GSVA enrichment analysis demonstrated that NRlncRNAs signature-related genes were mainly correlated with immune pathways, metabolic-and cell growth-related pathways, cell cycle, and apoptosis. Moreover, the risk score was correlated with the immune status of LUAD patients. Patients with higher risk scores had lower ESTIMATE scores and higher TIDE scores. The risk score was positively correlated with TMB. LINC01415, FRMD6-AS1 and FAM83A-AS1 were significantly overexpressed in lung adenocarcinoma, while the expression levels of MED4-AS1 and LINC01480 were lower in lung adenocarcinoma.

Conclusion: Overall, an innovative prognostic signature based on NRlncRNAs was developed for LUAD through comprehensive bioinformatics analysis, which can act as a predictor of immunotherapy and may provide guidance for clinicians.

KEYWORDS

lung adenocarcinoma, necroptosis, lncRNA, gene signature, immune landscape

Introduction

Lung adenocarcinoma (LUAD) is the most common type of lung cancer nowadays (Siegel et al., 2021). Although there are a variety of comprehensive therapeutic styles such as surgery, chemotherapy, radiotherapy and immunotherapy, some patients cannot be effectively treated and have a low 5-year overall survival rate due to the lack of specific targets (Nasim et al., 2019; Franzi et al., 2022). Therefore, to improve the prognosis and treatment of LUAD patients, it is vital to establish novel, efficient biomarkers and therapeutic approaches.

As a new style of cell death, necroptosis is universally mediated by receptor-interacting protein kinases1/3 of the receptor family (RIPK1/RIPK3) and is mainly governed by the effector protein Mixed Lineage Kinase Domain Like Pseudokinase (MLKL) (Frank and Vince, 2019; Yuan et al., 2019). Furthermore, necroptosis is able to modulate tumor immune responses which may lead to potential immunotherapeutic benefits (Gong et al., 2019; Tang et al., 2020a; Galluzzi and Garg, 2021). In one aspect, during necroptosis, cancer cells release cytokines and chemokines that stimulate inflammatory and tumor-modulating effects in tumor microenvironment. By luring macrophages and dendritic cells, on the other hand, necroptotic tumor cells encourage effector T cells to penetrate tumor tissues, which strengthens the immunosuppression of the tumor. Therefore, targeting necroptosis could lead to novel cancer therapies, especially immunotherapy. Nevertheless, the specific regulatory mechanism of necroptosis in lung cancer remains unclear.

Long non-coding RNAs (lncRNAs) are a class of non-proteincoding RNAs (ncRNAs) whose length is more than 200 nucleotides (Grote and Boon, 2018). A growing body of evidence has demonstrated that lncRNAs play a crucial role in lung cancer progression and the immune pathway (Chen et al., 2017; Bocchetti et al., 2021; Park et al., 2022). LncRNAS can alter cancer cells' resistance to immune responses, leading to immune evasion. In addition, several studies have demonstrated that lncRNAs can also act as competitive RNAs to influence genes involved in necroptosis (Jiang et al., 2021; Zhao et al., 2021; Chen et al., 2022). However, few studies have been done on necroptosis-related lncRNAs (NRlncRNAs) and tumor immune microenvironment (TIME) in LUAD.

In this study, an innovative prognostic signature based on NRlncRNAs was developed for LUAD. Additionally, we validated its clinical significance, confirming that this signature can act as a predictor of immunotherapy and may provide guidance for clinicians.

Methods and materials

Data sources

We downloaded the RNA sequencing data (59 normal tissues and 535 tumor tissues) and corresponding clinical information of LUAD samples from the TCGA database (https:/portal.gdc.cancer.gov/). Necroptosis-related genes were extracted from previous studies (Fan et al., 2014; Frank and Vince, 2019; Gong et al., 2019; Molnár et al., 2019; Yuan et al., 2019; Tang et al., 2020a). The mutation data in MAF format of LUAD samples was also obtained from TCGA. Next, we downloaded the Genome Reference Consortium Human Build 38 (GRCh38) to annotate lncRNAs and mRNAs by preforming Perl scripts.

Identification of Differentially Expressed Necroptosis-related LncRNAs in LUAD.

4191 lncRNAs and 19116 mRNAs were identified from the TCGA-LUAD RNA-seq data. The co-expression correlations between NRGs and lncRNAs in LUAD samples were

investigated through Pearson correlation analysis. The cutoffs for this study were |Coefficient| >0.4 and *p*-value <0.001. The "igraph" R package was used to get the co-expression network of NRGs and predictive lncRNAs. Lastly, we analyzed differentially expressed NRlncRNAs using the "limma" package (|logFc = 1|, FDR <0.05).

Construction and validation the prognostic NRlncRNAs signature

First, the prognosis-related NRlncRNAs were identified in TCGA-LUAD through univariate Cox proportional regression analysis ($p < 0.05$). Significant lncRNAs were visualized in heatmap by using "heatmap" package. The Sankey plot was created by the "limma", "dplyr," "ggalluvial," and "ggplot2" packages to visualize the correlation between NRGs and NRlncRNAs. Then, "caret" package was utilized to allocate all patients into the training and the testing sets. The "glmnt" was performed to select significant NRG-lncRNA into the Least absolute shrinkage and selection operator (LASSO) cox regression. The LASSO Cox regression approach was used to find the best panel of prognostic lncRNAs and create an optimum signature. The standardized expression levels of NRLs and the related regression coefficients produced from the LASSO regression analysis were then used to calculate each LUAD patient's survival risk score. The formula is given:

Risk score = Σ(Coef (lncRNAi)×Exp (lncRNAi)).

Coef and Exp denote the coefficient and the standardized expression levels of each NRL. The training set's median risk score was used as the demarcation point to divide LUAD samples into low- or high-risk subgroups. To compare the overall survival (OS) of the high-risk and low-risk subgroups among the training and testing sets, Kaplan-Meier (K-M) curves were generated by performing the "survival" package. A heatmap was utilized to display the significant lncRNA in this model. Time-dependent ROC curves were generated to assess the survival predictive ability of the NRlncRNAs signature. Univariate and multivariate cox regression analysis were performed to detect the independence of this prognostic risk model. We also contrasted the differences among different risk groups and clinical characteristics by using "limma" R package (Ma et al., 2020). To determine if our NRlncRNAs signature risk model is superior to previously reported signatures in LUAD, we compared its predictive power to that of other signatures, including two five-lncRNA signatures (Song et al., 2021; Wang et al., 2022) and a seven-lncRNA signature (Yao et al., 2021). The lncRNAs in these signatures were obtained from the corresponding published literature, and the AUC of 1-, 3-, and 5-year ROC curves, as well as the OS, were calculated for each signature.

Development of a nomogram score system

Subsequently, a nomogram score system based on age, stage, gender, and risk score of each patient with LUAD was constructed to predict the prognosis of individual patient outcomes. The procedure was run by using "survival" and "rms" packages. The calibration curve was conducted to assess the consistency between the actual outcomes and the predicted prognosis. The AUC of the ROC curve was used to compare the prediction abilities of the nomogram with other prognostic factors. Moreover, decision curve analysis (DCA) was used to evaluate the net clinical utility of the nomogram.

Function enrichment analysis and gene set variant analysis (GSVA)

To investigate the difference in potential molecular function and cancer-associated signaling, we analyzed the DEGs between high-risk and low-risk groups. Then Gene Oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted via R packages including clusterProfiler, org. Hs.eg.db and enrichplot. In addition, we performed GSVA to explore the NRlncRNAs signature in KEGG and GO. "c2. cp.kegg.v7.2. symbols.gmt" and "c5. go.v7.4. symbols.gmt" gene sets were downloaded from the MSigDB database. The procedure was conducted by using R packages including GSVA, limma, GSEABase and heatmap. Adjusted $p < 0.05$ was considered as statistical significance.

Immune microenvironment, immune check-point and immune therapy response analysis

To estimate the connection between the NRlncRNAs signature and immune microenvironment of LUAD samples, a gene expression matrix-based ESTIMATE algorithm was utilized to determine the infiltration levels of stromal cells and immune cells in tumors (Yoshihara et al., 2013). The immune and stromal scores reflected the infiltration levels of immune cells and stromal cells, respectively, while the ESTIMATE score was a stroma-immune composite score. Tumor-infiltrating immune cell dataset was obtained from TIMER2.0 (http://timer.cistrome.org) database. We applied TIMER (Li et al., 2017), CIBERSORT (Chen et al., 2018), QUANTIseq (Plattner et al., 2020), MCP-counter (Dienstmann et al., 2019), xCELL (Aran et al., 2017), and EPIC (Racle et al., 2017) algorithms to compare immune cell abundance between high-risk and low-risk groups based on the NRlncRNAs signature. The expression of immune checkpoint genes between different risk groups was examined to assess the potential effects on immunotherapy. To predict

immunotherapy response in patients with malignant tumors, tumor immune dysfunction and exclusion (TIDE) score (http://tide.dfci.harvard.edu/) was calculated.

Analysis of tumor mutation burden, prediction of the effect of chemotherapy and target therapy

The "maftools" R package was used to analyze the tumor mutation burden (TMB) of LUAD samples (Mayakonda et al., 2018). By comparing TMB between high- and low-risk groups, the top 20 genes with the highest mutation rate and their mutation types were obtained. Then the Kaplan–Meier survival curves were used to assess the effect of TMB on the OS of LUAD patients. The "pRRophetic" R package was performed to predict the IC50 of commonly used chemotherapeutic drugs (Geeleher et al., 2014). Wilcoxon signed-rank test was used to determine the difference between the groups.

Consensus clustering for NRlncRNAs signature

Unsupervised consensus clustering was conducted on 490 LUAD patients using "ConsensusClusterPlus" based on the expression of the NRlncRNAs signature to find potential molecular subgroups (Wilkerson and Hayes, 2010). The "survival" and "survminer" packages were used to perform Kaplan–Meier survival analysis between distinct clusters in R software. Principle component analysis was performed to explore the discrimination among different clusters and risk groups. A Sankey diagram was plotted to display the molecular subtypes and survival status of patients in different risk groups. The "heatmap" R package was used to examine the differences in molecular subtypes for diverse clinicopathological characteristics. GSVA analysis was performed to identify the potential KEGG pathways associated with different clusters. TMB value, the abundance of infiltrating immune cells, the expression of checkpoints and drug sensitivity analysis of different clusters were evaluated as mentioned earlier.

Cell culture

Immortalized lung epithelial cells (BEAS-2B) and lung adenocarcinoma cells (A549/PC9) were all obtained from American Type Culture Collection (ATCC). BEAS-2B and A549 cells were incubated in DMEM high glucose medium with 10% fresh fetal bovine serum. PC9 cells were incubated in RPMI 1640 medium with 10% fresh fetal bovine serum. All the cells were cultured in a constant-temperature incubator (37°C, 5% CO2) for proper time to get the total RNA.

Quantitative real-time PCR (RT-qPCR)

Total RNA of cells (BEAS-2B, A549 and PC9) and tissues was extracted by using an RNA extraction kit (Tiangen) following the protocol. 1000ng of total RNA was reversely transcribed and then PCR amplification of obtained cDNA was processed by using the kit (Cowin Bio.) and right primers. Sequences of primers we used in this study were designed by Primer-BLAST and were listed in supplemental table (Supplementary Table S1).

Statistical analysis

R 4.0.4 (https://www.r-project.org/) was used for all statistical analyses. The chi-square test or Fisher's exact test were used to test categorical variables. On continuous variables, the t-test or Wilcoxon test was used. Statistical significance was defined as $p < 0.05$.

Results

Identification of differentially expressed necroptosis-related LncRNAs in patients with LUAD

In this study, 490 LUAD samples with comprehensive clinical data were included for further analysis. A flowchart of the study is presented in Figure 1. Firstly, we identified 67 NRGs from previous studies (Supplementary Table S2). Then, we analyzed the expression matrix of NRGs and lncRNAs by performing Pearson correlation analysis and obtained 586 NRlncRNAs (Supplementary Table S3). The network of mRNA-lncRNA co-expression showed a potential connection of 46 NRGs and 586 NRlncRNAs (Figure 2A). Among the 586 NRlncRNAs, 249 were found to be differentially expressed using the criteria: |logFc = 1|, FDR < 0.05 (Figure 2B). The heatmap displayed the 249 NRG-lncRNAs expression landscape between tumor and normal tissues (Figure 2C).

Development and validation of the prognostic risk model

First, a total of 15 NRlncRNAs were identified as the prognostic lncRNAs for patients with LUAD by univariate Cox regression analysis. (Figure 3A). The expression heatmap of 15 prognostic NRlncRNAs was presented in Figure 3B. The

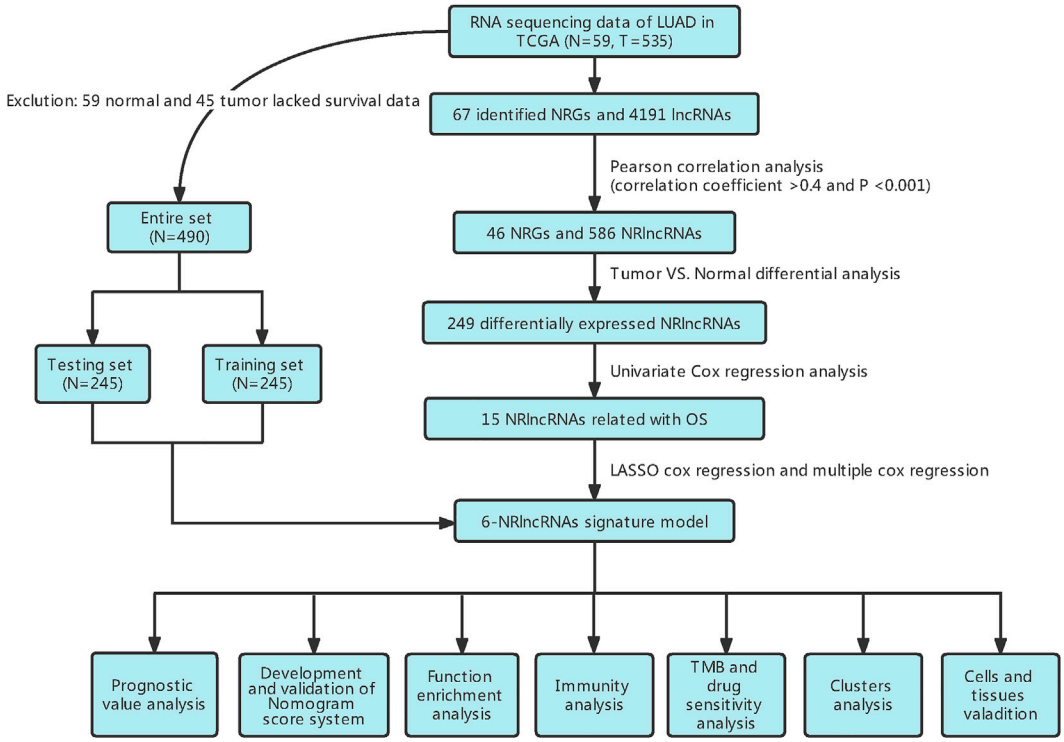


FIGURE 1
Flow diagram of the study design.

Sankey diagram showed the positive regulatory relationship between 16 NRGs and 15 NRlncRNAs (Figure 3C). Next, we randomly divided 490 LUAD samples into a training dataset (n = 245) and a testing dataset (n = 245), and we performed LASSO cox regression to select 13 NRlncRNAs for multiple cox regression (Figures 3D,E; Supplementary Table S4). Finally, the risk model was updated using multivariate Cox regression analysis and included six NRlncRNAs. Risk score = TRMT2B-AS1 * (-1.5285) + LINC01480 * (-0.7024) + FRMD6-AS1*1.3853 + FAM83A-AS1 * 0.3043 + MED4-AS1 * (-1.2861) + LINC01415 * 2.0938. The expression of six NRlncRNAs between high- and low-risk groups was presented in Supplementary Figure S1. Based on the median risk score of the training set, we divided samples into high-risk group and low-risk group in the training, testing and entire datasets (Supplementary Figure S2). And we discovered that the majority of the dead patients belonged to the high-risk category (Supplementary Figure S2). The expression of six NRlncRNAs in high- and low-risk groups was shown in Figures 4A–C. Additionally, the K-M curves indicated that patients with a high-risk score had a lower chance of survival than patients with a low-risk score. (Figures 4D–F). NRlncRNAs To further investigate the prognosis value of this signature, we performed the time-dependent receiver operating characteristic (ROC) analysis. The area under curve (AUC) values for the 1-year, 2-year, and 3-year survival rates showed good specificity and sensitivity of this signature in predicting OS either in the training, testing, or entire group (Figures 4G–I). Next, we investigated whether the signature was independent of other clinical characteristics using univariate and multivariate cox regression analysis. The results indicated that the risk score of our signature was correlated with the OS and it acted as an independent prognostic predictor for patients with LUAD (Figures 5A–F). To explore the differences in risk score among different subgroups of patients, we discovered that male and advanced patients had a higher risk score than female and early-stage patients in the entire TCGA cohort (Figures 5G–I). Furthmore, the survival rate was significantly lower in the high-risk group than in the low-risk group in the age ≥ 60 ($p < 0.001$), age < 60 ($p = 0.033$), male ($p < 0.001$), female ($p < 0.001$), stage I and stage II ($p = 0.001$), and stage III and stage IV ($p = 0.002$) subgroups of patients, according to stratified survival analysis in combination with clinical characteristics (Supplementary Figure S3). Additionally, compared with other previously identified signatures, our risk model based on the 6 lncRNAs signature had better predictive power (Supplementary Figure S4).

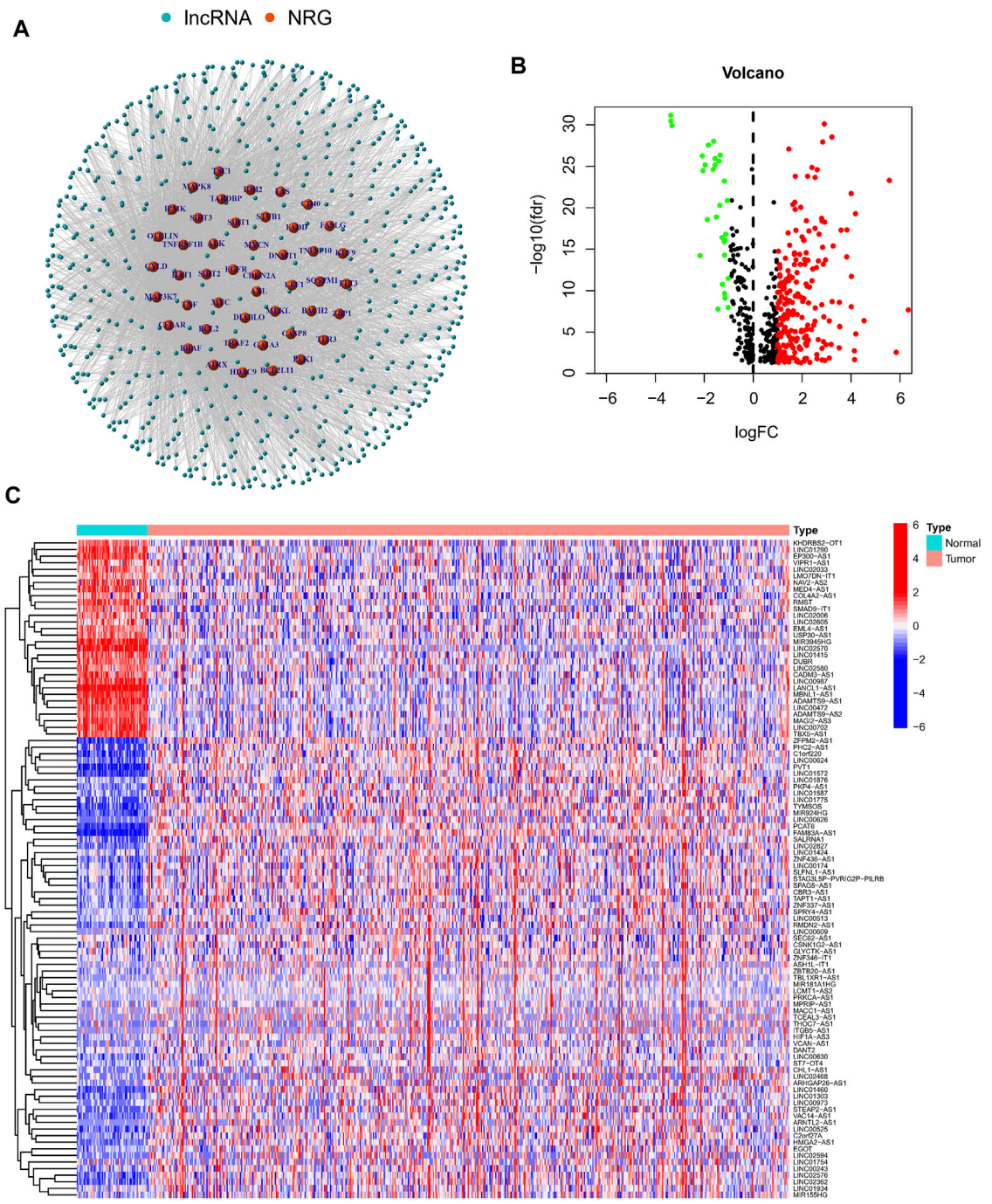


FIGURE 2
Identification of differentially expressed necroptosisNRlncRNAs in Lung adenocarcinoma. **(A)** A co-expression network of the 586 NRlncRNAs-mRNA was constructed and visualized by using "igraph" R package. **(B,C)**. The volcano plot and heatmap showed the differentially expressed NRlncRNAs in tumor and normal tissues.

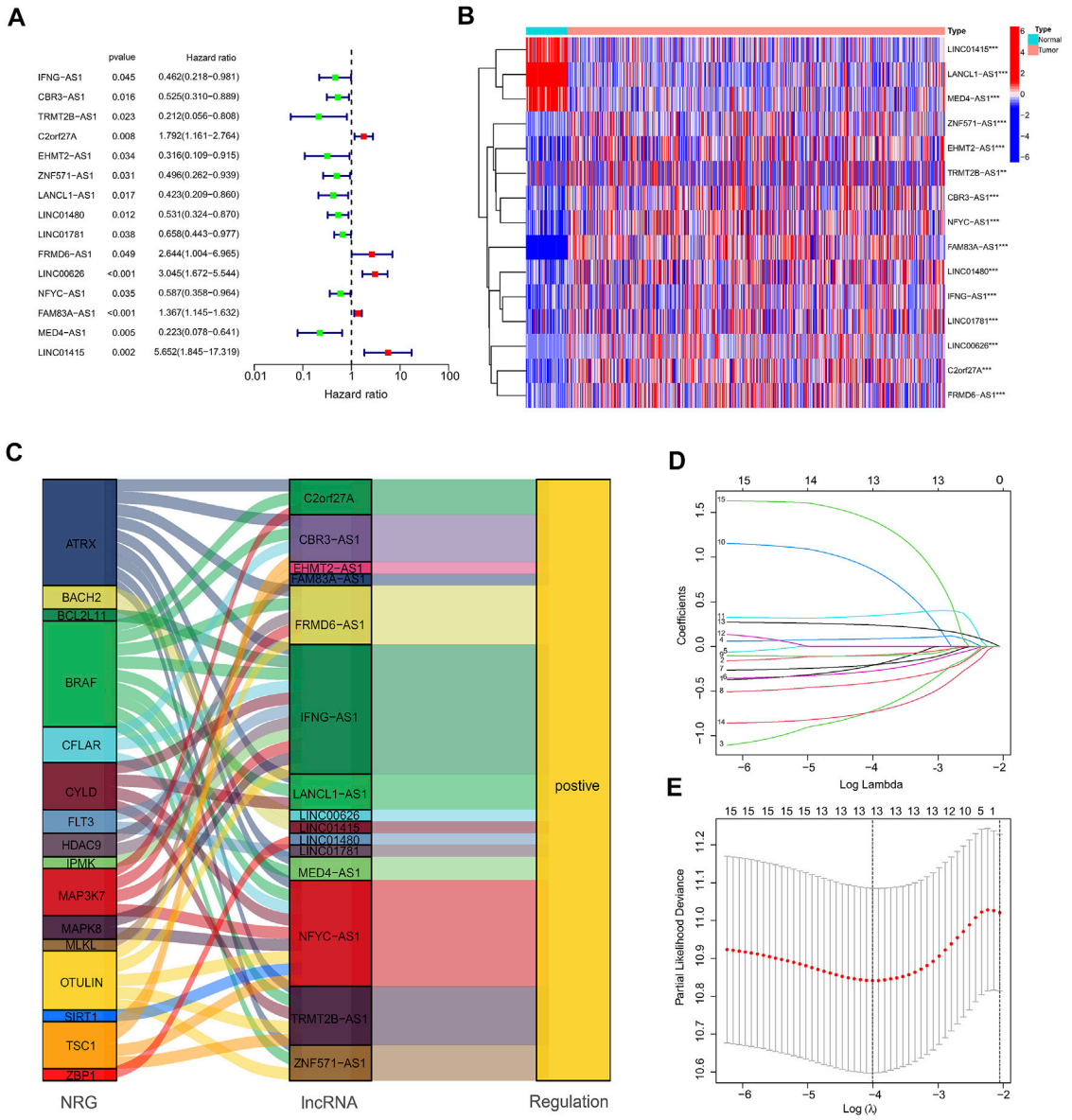


FIGURE 3
Construction of the NRlncRNAs signature in the TCGA. **(A,B)**. The forest plot and heatmap showed 15 NRlncRNAs with prognostic value in LUAD. **(C)**. The Sankey diagram of 16 NRGs and 15 NRlncRNAs. **(D,E)**. The LASSO coefficient and the 10-fold cross-validation for variable selection in the LASSO model.

The signature-based nomogram score system for predicting the prognosis of LUAD patients

In order to create a more reliable model for prognosis prediction, a nomogram comprised of the identified NRlncRNAs signature and several clinical characteristics was found to be effective for predicting 1-, 3-, and 5-year survival probabilities in the entire dataset (Figure 6A). Furthermore, the calibration curve revealed an accordant agreement between observation and prediction for 1-, 3-, and 5-year OS in LUAD (Figure 6B). In addition, the nomogram score system had superior predictive power (1-year AUC = 0.735, 3-year AUC = 0.724, and 5-year AUC = 0.697) than the risk score, age, gender and stage (Figures 6C–E). The DCA curves suggested that the nomogram and risk score had good consistency in forecasting survival rate at 1-, 3-, and 5-year (Figures 6F–H).

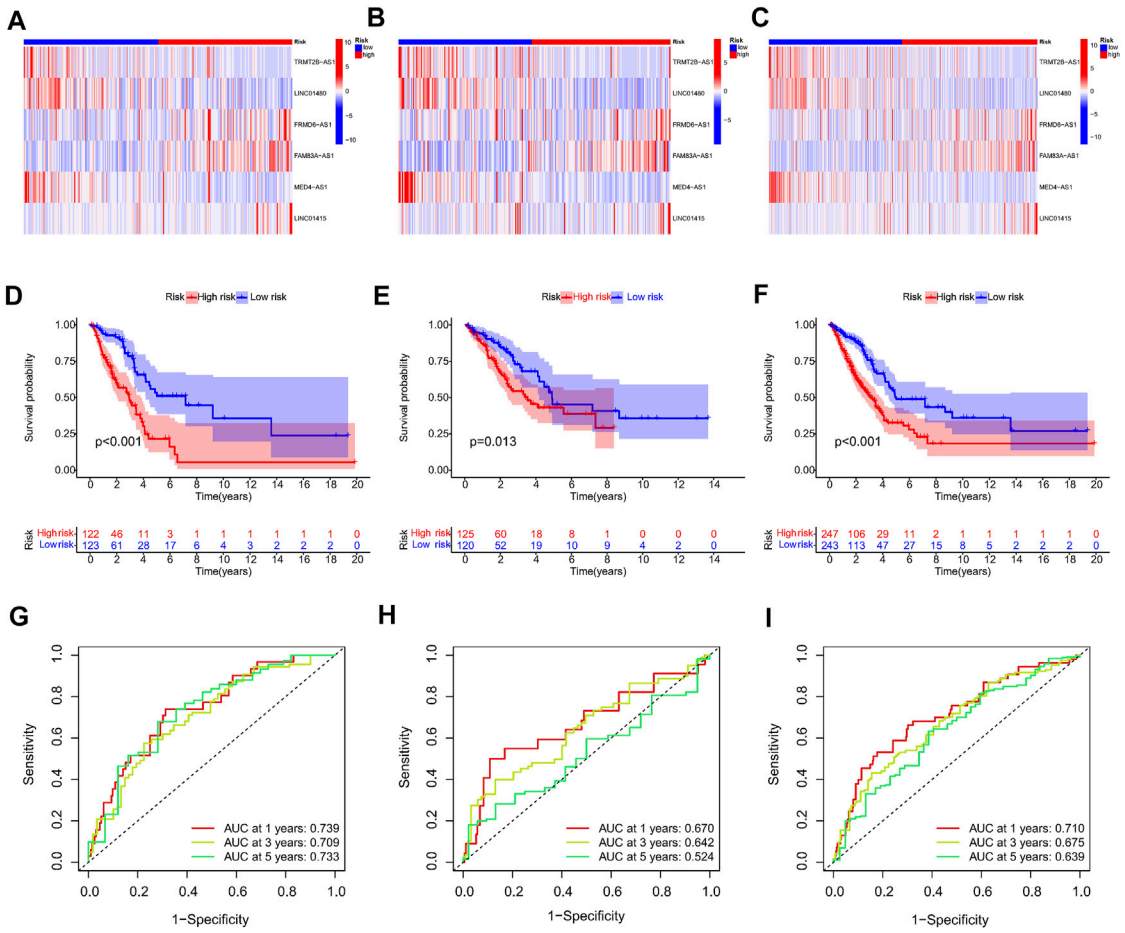


FIGURE 4
Prognosis value of the 6 NRlncRNAs model in the train, test, and entire sets. **(A–C)** The heatmap of six NRlncRNAs between two groups in the train, test, and entire set, respectively. **(D–F)** Survival status and time of patients between two groups in the train, test, and entire set, respectively, **(G–I)** The time-dependent ROC curve of patients between two groups in the train, test, and entire set, respectively.

Function enrichment analysis

We performed GO and KEGG to investigate the probable mechanisms of the NRlncRNAs signature-related genes' expression. The results revealed that these genes were associated with immune pathways, such as MHC class II protein complex binding, Th17 cell or Th1 and Th2 cell differentiation, and antigen processing and presentation. Moreover, cancer-associated biological functions such as cell cycle, apoptosis, and glycolysis were found to be significant in LUAD patients (Supplementary Figure S4). We then run a GSVA enrichment analysis to delve deeper into the underlying differences in biological characteristics that underlie the different risk groups. As shown in Figures 7A,B, the high-risk group had a considerable enrichment of metabolic- and necroptosis-related pathways, such as pyruvate metabolism, glucose metabolism, DNA replication, cell cycle and P53 signal pathway.

Tumor microenvironment, immune cell infiltration, and immunotherapy response of LUAD patients

As above, we discovered that the risk score was associated with immune-associated pathways. Next, we compared the difference in immunological state among different risk categories. The results based on ESTIMATE algorithm revealed that the immune and ESTIMATE scores were lower in the high-risk group than in the low-risk group (Figures 8A–C). The correlation examination of immune cell infiltration in two risk groups indicated that $CD4^+$ T cells, NK cells, and B cells were negatively regulated in multiple algorithms, whereas neutrophil cells were positively regulated (Figure 8D). The $CD4^+$ Th2 cells and B cells had the most obvious coefficient connection with the risk score

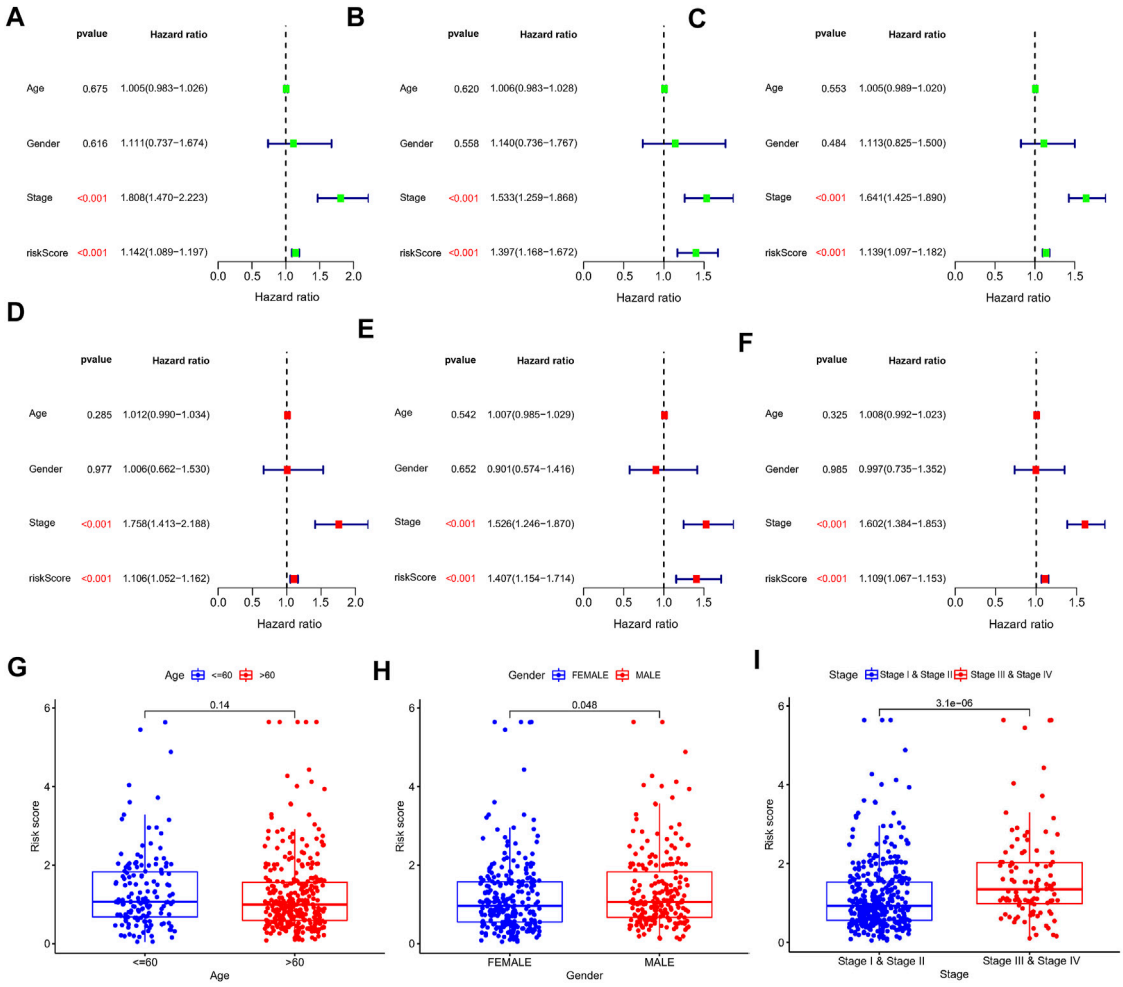


FIGURE 5
Univariate and multivariate Cox regression analysis in the train, test, and entire set, respectively. **(A–C)** Univariate Cox regression analysis in the train, test, and entire set, respectively. **(D–F)** Multivariate Cox regression analysis in the train, test, and entire set, respectively. **(G–I)** Distribution of risk score stratified by age, gender and clinical stage in entire set.

(Figures 8E,F; Supplementary Table S5). Then, to investigate the difference in immunological status between the high-risk and low-risk groups, we examined their TIDE scores, exclusion scores, and dysfunction scores. According to the findings, the TIDE and exclusion score were higher in the high-risk group than in the low-risk group (Supplementary Figure S6). Low-risk patients had a greater dysfunction score than high-risk patients, whereas the exclusion score had the opposite pattern (Supplementary Figure S6). These data indicated that the high-risk group had higher immune escape risk and poor prognosis. The different expression of 28 immune check points between the two risk groups showed that the expression of TNFSF4, CD276 was higher in the high-risk group, while the other multiple checkpoints presented higher expression in the low-risk group (Figure 8G).

Analysis of tumor burden mutation and drug sensitivity

Previous studies have reported that TMB plays a crucial role in immunotherapy, chemotherapy, and target therapy responses. Hence, we further investigated the correlation between risk score and TMB. It showed that the high-risk group had higher TMB than low-risk group, and the risk score was positively correlated with TMB (Figures 9A,B). Patients with lower TMB and higher risk score had worse prognosis (Figure 9C). According

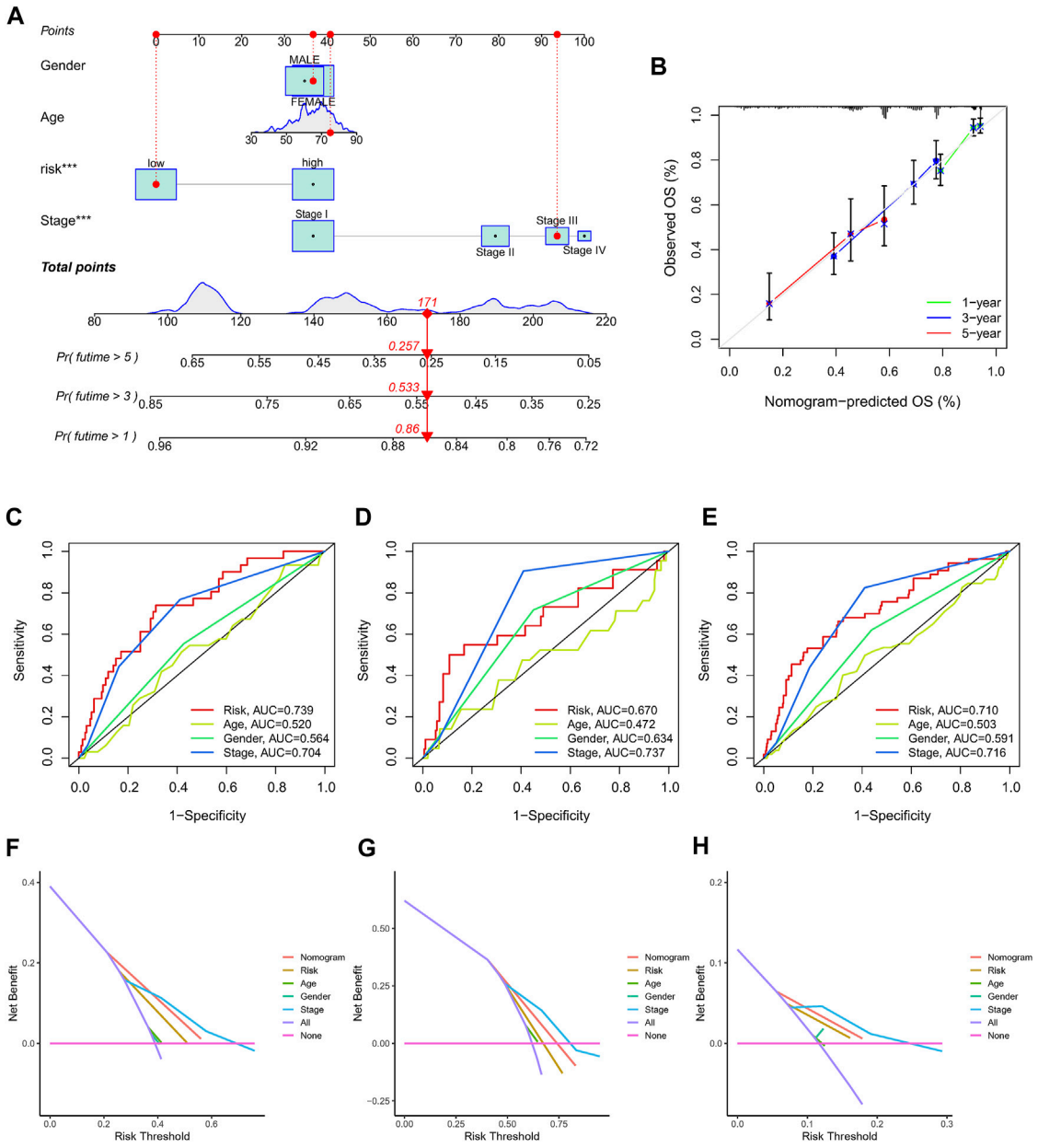


FIGURE 6
The nomogram used to predict the OS prognosis of LUAD patients at 1, 3, and 5 years. **(A)** A nomogram based on characteristics and clinical information. **(B)**. Calibration curve of nomogram. **(C–E)** ROC analysis of risk score, nomogram, age, gender and clinical stage predicting OS at 1, 3, and 5 years. **(F–H)** The DCA of NRlncRNAs prognostic risk scores and several clinicopathological factors such as gender, age, and stage.

to the status of TP53, KRAS and EGFR, the patients with TP53 and KRAS mutations had higher risk scores, while patients with EGFR mutation showed lower risk scores (Figures 9D–F). The waterfall plot presented the top 20 genes with the most mutations in different risk groups. The high-risk group presented higher genetic alteration rate than the low-risk group (92.28 vs. 84.19%), and TP53 and TTN were the top two genes in each risk group (Figures 9G,H). Drug sensitivity analysis between different risk groups showed that patients with high risk scores were more sensitive to docetaxel, doxorubicin, erlotinib and gefitinib et al. (Supplementary Figure S7).

Characterization of clusters in different risk groups

To further assess the distinct molecular patterns based on the expression of the NRlncRNAs signature, we performed

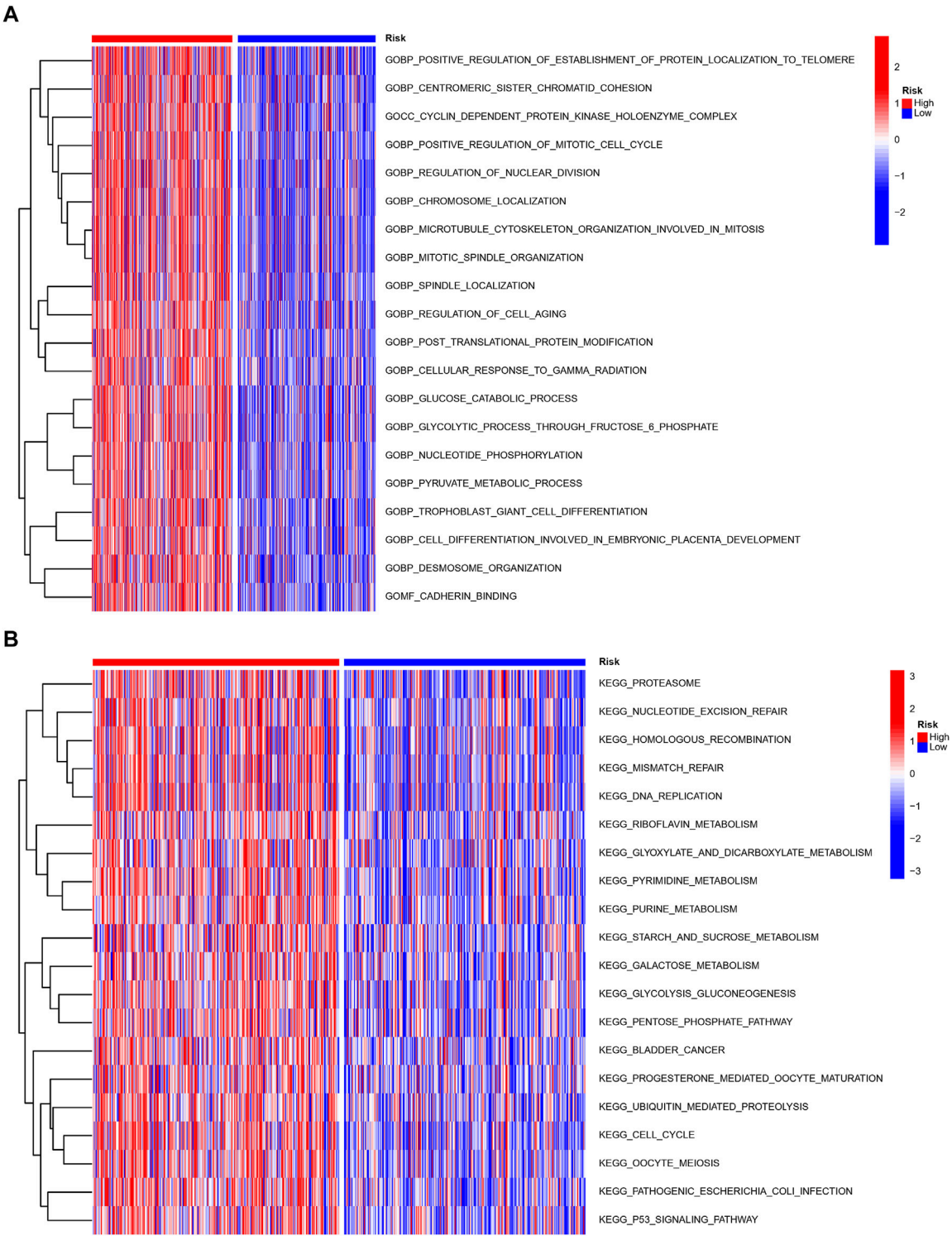


FIGURE 7
Molecular characteristics of patients in the high- and low-risk groups. The GSVA analysis in the high- and low-risk group to enrich GO characteristic gene sets **(A)** and KEGG gene sets **(B)**.

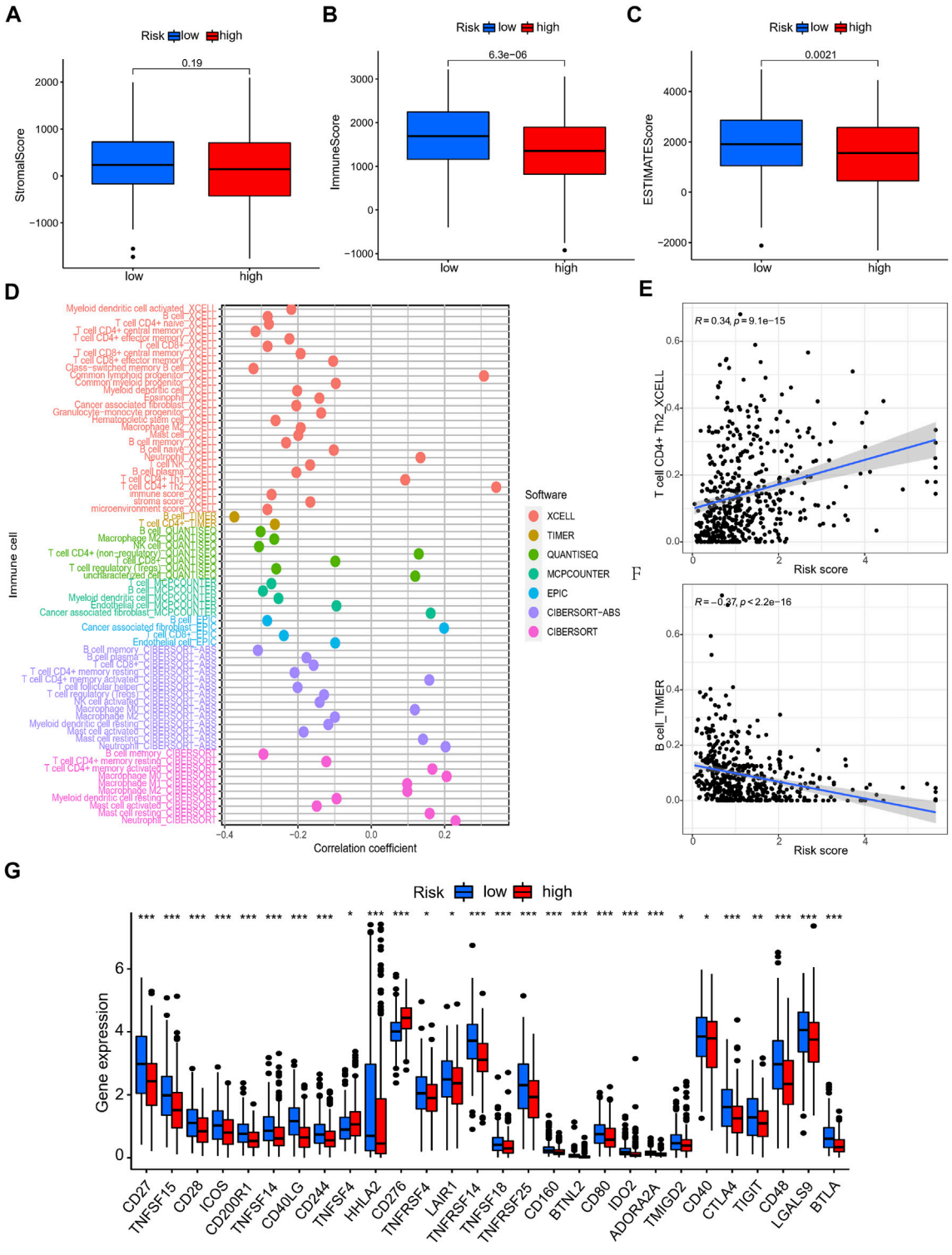


FIGURE 8
The immune landscape analysis of NRlncRNAs signature in LUAD patients. **(A–C)** The TME score (immune score, stromal score, and ESTIMATE score) between different risk group. **(D–F)** Relationship between the NRlncRNAs risk score and immune cells infiltration. **(G)** The differential expression of immune checkpoint between high- and low-risk group.

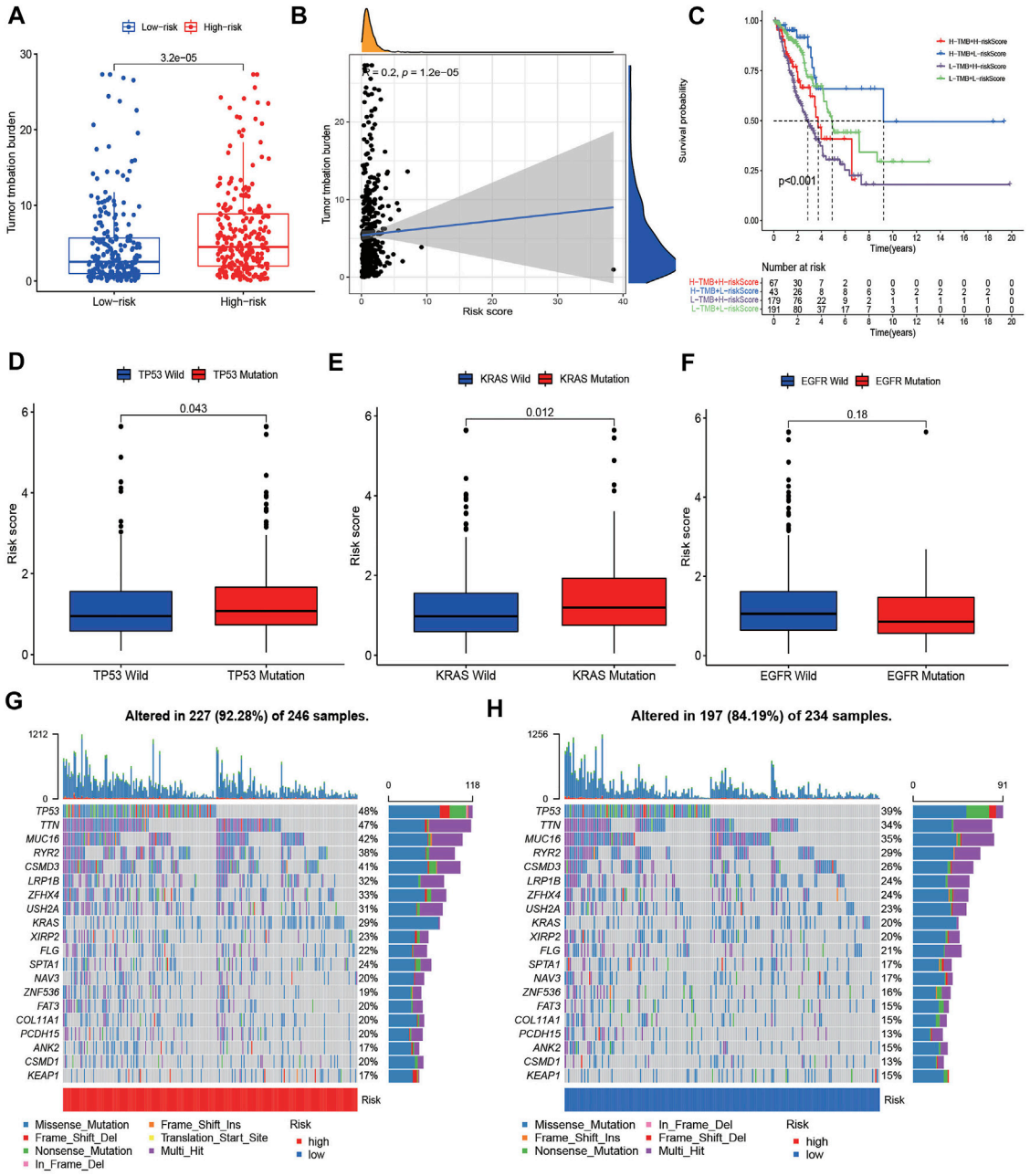


FIGURE 9
The TMB level and Somatic mutation analysis of risk score. **(A)** Differences of tumor mutation burden (TMB) in the high-risk and low-risk groups. **(B)** Relationship between TMB and risk score. **(C)** Survival analysis between patients with high- and low-TMB. **(C)** Two-factor survival analyses of risk score and TMB levels. **(D–F)** Differences of risk score between different status of TP53, KRAS, and EGFR. **(G,H)** Landscape of top 20 mutated gene mutation profiles between high- and low-risk group.

unsupervised consensus clustering to divided patients with different risk scores into two clusters, with k = 2 shown to be the best for clustering stability (Figure 10A). As presented in Figure 10B, the patients in cluster 1 had a poorer prognosis than those in cluster 2. Furthermore, the results of PCA suggested that clusters and risk groups can be completely distinguished (Figures 10C,D). The Sankey plot revealed the distribution of patients in different risk groups, clusters and survival status, with cluster 1 accounting for the majority of high-risk patients (Figure 10E). Furthermore, the heatmap presented the expression landscape of six NRG-LncRNAs in two clusters, which indicated that two clusters had significant differences in clinical stage, age, and

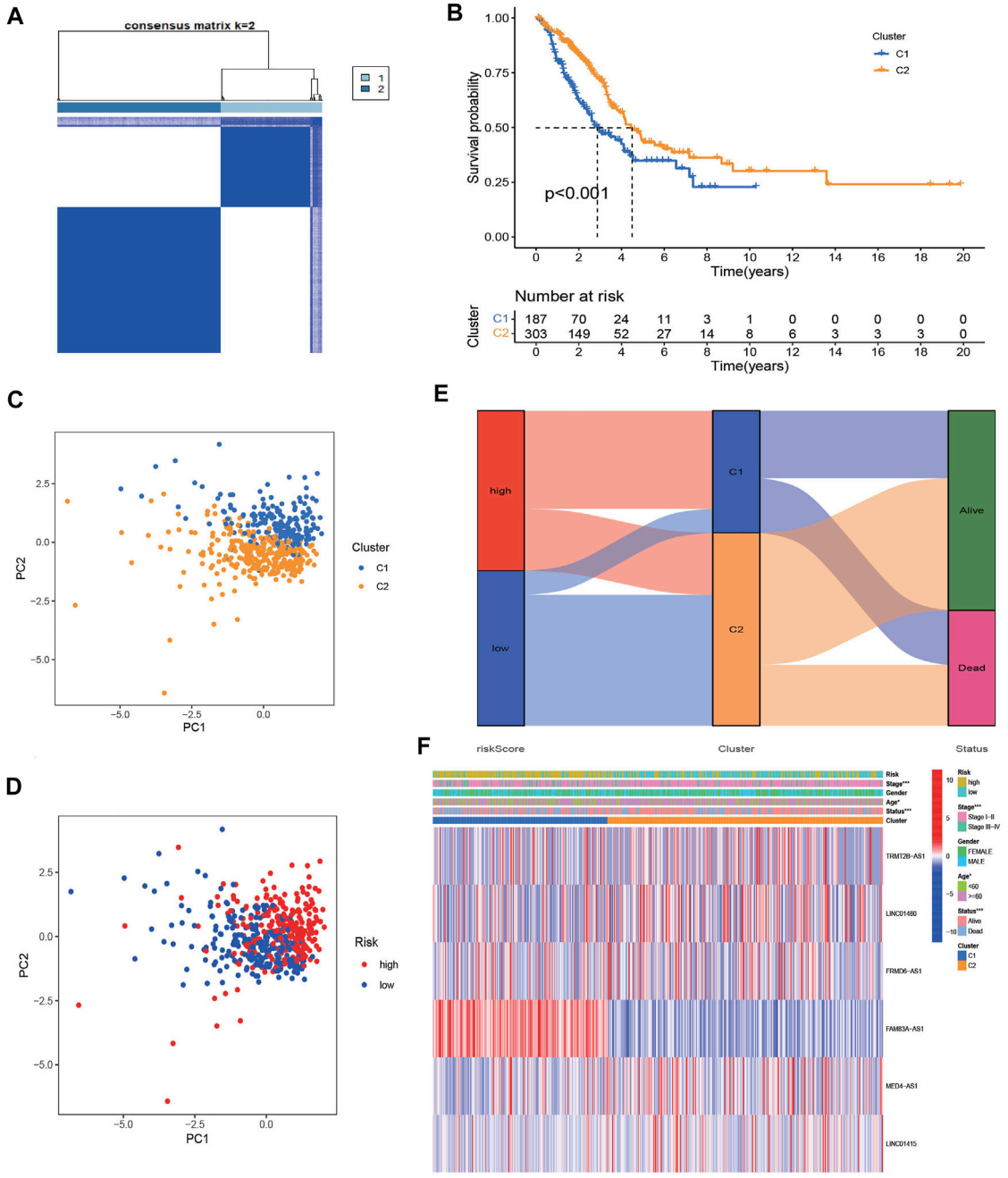


FIGURE 10
Identification of two clusters based on prognostic NRlncRNAs signature. **(A)** Consensus clustering (K-means) algorithm was performed for overall patients. Consensus matrix plots. K = 2 was determined as optimal clustering number. **(B)** K–M survival analysis in clusters C1 and C2. **(C,D)** The distribution of different patients from two clusters and risk groups. **(E)**. The Sankey diagram showed distribution of patients in different risk groups, two clusters and survival status. **(F)**. Heatmap of the lncRNA expression of prognostic NRlncRNAs signature between two clusters.

survival status (Figure 10F). The GSVA results revealed that cluster1 was associated with cell cycle, mismatch repair and some cancer-associated pathways (Supplementary Figure S8). Similar to the risk model, cluster 2 had a higher immune score and ESTIMATE score (Supplementary Figure S9). The status of multiple immune cells infiltration in two clusters exhibited significant differences. The heatmap of the immune infiltration landscape revealed that cluster 1 had a positive correlation with B cells, $CD4^+$ T cells, macrophages cells, mast cells activated and cancer-associated fibroblasts, while it had a

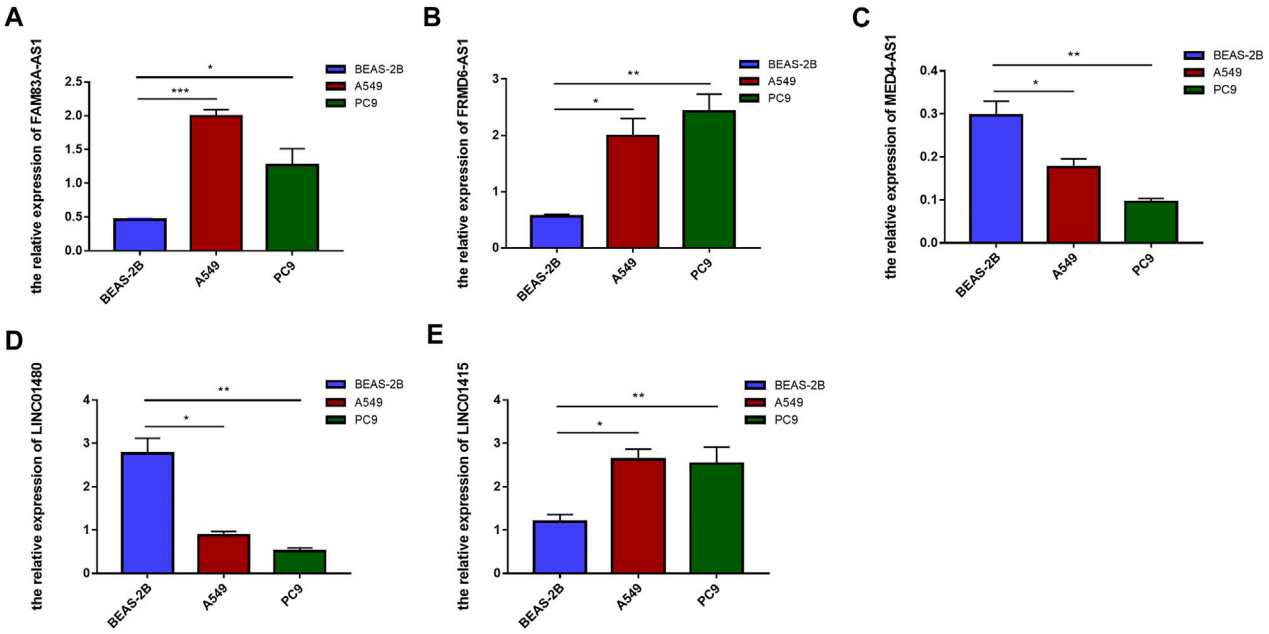


FIGURE 11
The expression level of six NRlncRNAs in BEAS-2B, A549, and PC9 cell lines. Expression of FAM83A-AS1 **(A)**, FRMD6-AS1 **(B)**, MED4-AS1 **(C)**, LINC01480 **(D)**, and LINC01415 **(E)** in normal lung epithelial cells (BEAS-2B) and lung adenocarcinoma cells (A549/PC9) detected by RT-qPCR. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

negative correlation with monocytes (Supplementary Figure S9). The expression level of these immune checkpoints, such as TNFSF4, CD274, CD276, NRP1, TNFRSF18 was higher in cluster 1, while the expression levels of CD44, IDO2, HHLA2 et al. were higher in cluster 2 (Supplementary Figure S9). The drug prediction analysis showed similar results to the risk model. As shown in Supplementary Figure S10, cluster 1 had a lower IC50 of gefitinib, erlotinib, docetaxel and paclitaxel. These results revealed that individuals in cluster 1 were more susceptible to chemotherapy and target therapy.

The expression level of six prognostic LncRNAs

To further verify the expression of these screened lncRNAs in lung adenocarcinoma cells, total RNA of BEAS-2B, A549, and PC9 were extracted and real-time quantitative PCR was conducted. Surprisingly, our results were consistent with the database. Compared with normal lung-epithelial cells (BEAS-2B), the expression level of LINC01415, FRMD6-AS1 and FAM83A-AS1 was significantly higher in lung adenocarcinoma cells (A549/PC9), while the expression level of MED4-AS1 and LINC01480 was lower in lung adenocarcinoma cells (Figure 11). We also detected the expression of our signature in lung adenocarcinoma tissues and adjacent normal lung tissues, and we obtained consistent results with our observations in cells (Figure 12).

Discussion

Lung adenocarcinoma remains a serious health problem worldwide with its high mortality and morbidity rates (Mao et al., 2016). In order to improve prognosis for LUAD patients, it is vital to identify a precise and reliable prognostic signature. Recent studies indicated that tumor cells resistant to apoptosis may be sensitive to the necroptosis (Su et al., 2016; He et al., 2017), suggesting that necroptosis may be a potential therapeutic target for lung adenocarcinoma. Besides, lncRNAs also play important roles in tumor genesis and metastasis. Therefore, a necroptosis-related lncRNAs signature was constructed here for the prognosis and treatment of LUAD patients.

In this study, we first identified 15 NRlncRNAs correlated with the OS of patients with LUAD by performing Pearson correlation analysis and univariate Cox proportional regression analysis. Subsequently, a signature including 6 NRlncRNAs (TRMT2B-AS1, LINC01480, FRMD6-AS1, FAM83A-AS1, MED4-AS1, LINC01415) was developed by LASSO analyses and multiple cox regression. Furthermore, LUAD patients were classified into high- and low-risk groups according to their risk scores based on the signature. We found that

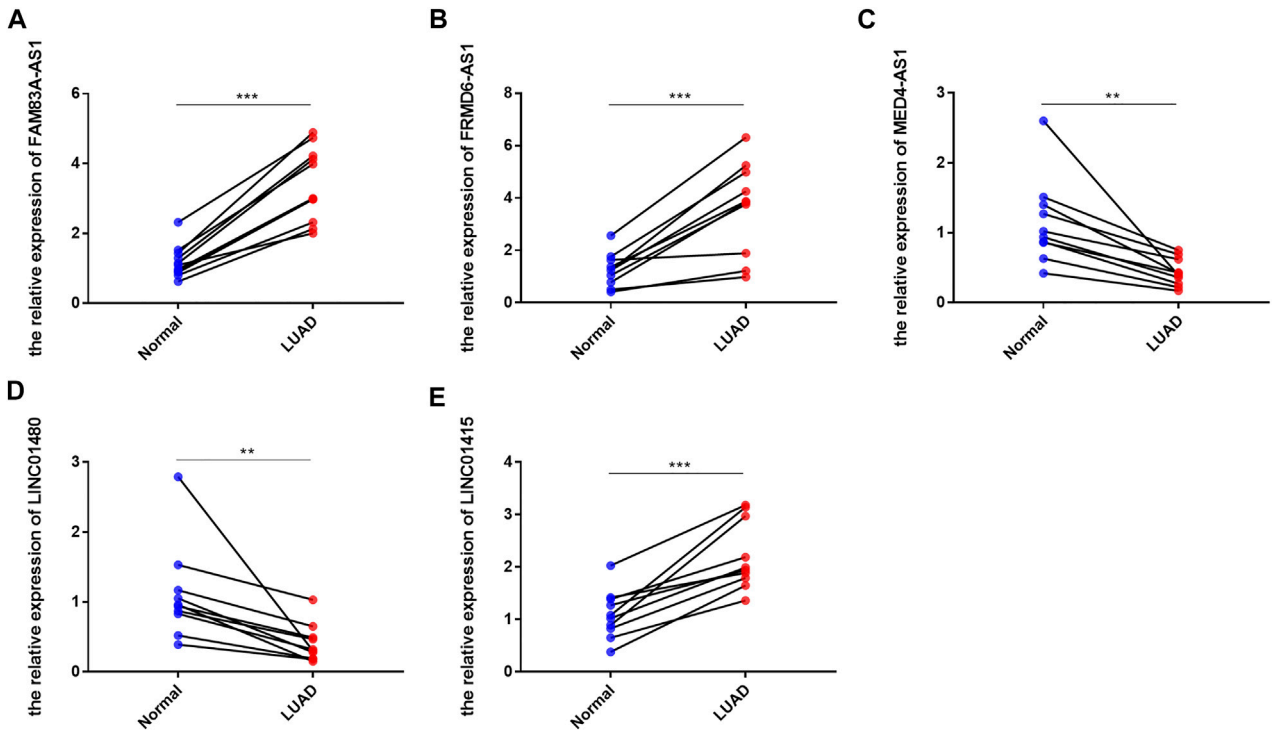


FIGURE 12
The expression level of six NRlncRNAs in lung adenocarcinoma tissues and adjacent normal lung tissues. Expression of FAM83A-AS1 **(A)**, FRMD6-AS1 **(B)**, MED4-AS1 **(C)**, LINC01480 **(D)**, and LINC01415 **(E)** in lung adenocarcinoma tissues and adjacent normal lung tissues detected by RT-qPCR. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

patients with high risk scores have lower probabilities of survival than those with low risk scores. The ROC analysis showed that our signature had significant prognostic value both in training and testing sets and it showed superiority over previous lncRNA signatures in predicting survival of LUAD patients. Besides, the nomogram system based on our signature and some clinical characteristics also showed steady predictive performance. To sum up, our results above indicated that our NRlncRNAs signature is a good prognostic predictor for LUAD patients.

A mount of studies illustrated the vital role of lncRNAs in tumor progression, but the relationship between the 6 lncRNAs and LUAD is still unclear to us. Among these 6 NRlncRNAs included in our signature, FAM83A-AS1 was found to promote the development of LUAD and suggesting a novel therapeutic approach for LUAD by sponging miR-141-3p (Huang et al., 2022). MED4-AS1 was associated with the prognostic of lung adenocarcinoma (Tang et al., 2020b). LINC01415 was associated with a poor prognosis in ESCC (Tang et al., 2020c). However, the other lncRNAs have not been reported in tumors, which may give us insight into the mechanism of development of lung adenocarcinoma. Therefore, we performed function enrichment analysis based on the differentially expressed genes between different risk groups to investigate the potential mechanism. GSVA showed that metabolism- and cell growth-associated pathways including pyruvate metabolism, glucose metabolism, mismatch repair, and cell cycle was observed in the high-risk group. Nowadays, more and more studies have found that metabolic proteome is involved in tumor development and immune response, which is likely to be a new target for future tumor therapy (Malireddi et al., 2020; Liu et al., 2021). In breast cancer cells glucose deprivation triggers ZBP1-depedent necroptosis (Yu et al., 2021). In colorectal cancer cells, by scavenging free radicals in the mitochondria, glucose confers resistance to 5-FU-induced necroptosis (Zhu et al., 2021). In summary, the necroptosis-related signature may be involved in glucose metabolism in patients with LUAD, thereby influence the necroptosis of lung adenocarcinoma cells, which may be an important direction for future research in lung adenocarcinoma treatment.

Previously, a large number of studies have reported that necroptosis is related to tumor development and immune response (Malireddi et al., 2020; Liu et al., 2021; Yan et al., 2022). In our study, the GO and KEGG analysis indicated that the NRlncRNAs signature-related genes were associated with immune pathways, especially MHC class II protein complex binding. Previous reports indicated that MHC class II protein

may be involved in the polarization of tumor-associated macrophage (TAM) to M1 phenotype, which could inhibit the lung cancer cells proliferation and promote the apoptosis (Yu et al., 2021). In addition, comparing immune status among different risk groups, we found that the high-risk group had a higher risk of immune escape and a poorer prognosis. $CD4^+$ T cell, NK cell, and B cells were mainly active among the high-risk groups, some of which were closely linked to necroptosis. As studied previously, Rac-1 related-necrotic cells could enhance proinflammatory NK cell killing (Zhu et al., 2021). The inhibition of RIP3 can increase the proportion of $CD4^+$ T cells and inhibit the secretion of inflammatory cytokines (IFN-γ, IL-16 and IL-17) (Duan et al., 2022). These results indicated that our NRlncRNAs signature was significantly related with tumor immune microenvironment and it can predict the immune landscape of patients with LUAD.

Moreover, our findings demonstrated that patients in high-risk group had higher TMB, and the risk score was positively correlated with TMB. Besides, patients with lower TMB and high-risk had worse prognosis. TMB is considered a potential biomarker for discriminating NSCLC patients who might benefit more from immunotherapy (Pan et al., 2022). This also suggests that patients with high risk may be more sensitive to immunotherapy. In addition, our results showed that patients in high-risk group were more sensitive to docetaxel, doxorubicin, erlotinib and gefitinib. These results allowed us to select these populations that are more sensitive to drugs and receive better treatment.

Finally, we used RT-qPCR to detect he expression of these signature NRlncRNAs in lung adenocarcinoma cells and tissues. Surprisingly, the results confirmed that the expression levels of these NRlncRNAs were abnormal in LUAD, implying that these lncRNAs may play distinct roles in LUAD. Although our findings shed light on the mechanisms of necroptosis in LUAD, the limitations and flaws still exist. First, although we conducted some experiments to validate it, additional clinical evidence is required to confirm the findings. Second, the underlying mechanism of NRlncRNAs in LUAD is still unknown, and more research is needed in the future.

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

HR and BX were in charge of designing the research. JW and DL were responsible for data analysis and manuscript writing. GZ and SS mainly revised the article. All the authors gave final approval of the published version, and agreed to be responsible for all aspects of the work.

Funding

The present study was supported by the National Natural Science Foundation of China (Program No: 82102801), Natural Science Foundation of Shaanxi Province (Program No: 2022JQ-818&2019JM-559).

Acknowledgments

We would like to thank all teammate for contributing 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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.966896/full#supplementary-material

References

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

Bocchetti, M., Scrima, M., Melisi, F., Luce, A., Sperlongano, R., Caraglia, M., et al. (2021). LncRNAs and immunity: coding the immune system with noncoding oligonucleotides. *Int. J. Mol. Sci.* 22 (4), 1741. doi:10.3390/ijms22041741

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, F., Yang, J., Fang, M., Wu, Y., Su, D., and Sheng, Y. (2022). Necroptosis-related lncRNA to establish novel prognostic signature and predict the immunotherapy response in breast cancer. *J. Clin. Lab. Anal.* 36 (4), e24302. doi:10.1002/jcla.24302

Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017). Gene regulation in the immune system by long noncoding RNAs. *Nat. Immunol.* 18 (9), 962–972. doi:10.1038/ni.3771

Dienstmann, R., Villacampa, G., Sveen, A., Mason, M. J., NieDzwiecki, D., NesbAkken, A., et al. (2019). Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. *Ann. Oncol.* 30 (10), 1622–1629. doi:10.1093/annonc/mdz287

Duan, C., Xu, X., Lu, X., Wang, L., and Lu, Z. (2022). RIP3 knockdown inhibits necroptosis of human intestinal epithelial cells via TLR4/MyD88/NF-κB signaling and ameliorates murine colitis. *BMC Gastroenterol.* 22 (1), 137. doi:10.1186/s12876-022-02208-x

Fan, H., LiuF.DonG, G., Ren, D., Xu, Y., Dou, J., et al. (2014). Activation-induced necroptosis contributes to B-cell lymphopenia in active systemic lupus erythematosus. *Cell. Death Dis.* 5 (9), e1416. doi:10.1038/cddis.2014.375

Frank, D., and Vince, J. E. (2019). Pyroptosis versus necroptosis: Similarities, differences, and crosstalk. *Cell. Death Differ.* 26 (1), 99–114. doi:10.1038/s41418-018-0212-6

Franzi, S., Mattioni, G., Rijavec, E., Croci, G. A., and Tosi, D. (2022). Neoadjuvant chemo-immunotherapy for locally advanced non-small-cell lung cancer: A review of the literature. *J. Clin. Med.* 11 (9), 2629. doi:10.3390/jcm11092629

Galluzzi, L., and Garg, A. D. (2021). Immunology of cell death in cancer immunotherapy. *Cells* 10 (5), 1208. doi:10.3390/cells10051208

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

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The role of necroptosis in cancer biology and therapy. *Mol. Cancer* 18 (1), 100. doi:10.1186/s12943-019-1029-8

Grote, P., and Boon, R. A. (2018). LncRNAs coming of age. *Circ. Res.* 123 (5), 535–537. doi:10.1161/CIRCRESAHA.118.313447

He, G. W., Gunther, C., Thonn, V., Yu, Y. Q., Martini, E., Buchen, B., et al. (2017). Regression of apoptosis-resistant colorectal tumors by induction of necroptosis in mice. *J. Exp. Med.* 214 (6), 1655–1662. doi:10.1084/jem.20160442

Huang, H., Yang, C., Zhang, Q., Zhuo, T., Li, X., Li, N., et al. (2022). Long non-coding RNA FAM83A antisense RNA 1 (lncRNA FAM83A-AS1) targets microRNA-141-3p to regulate lung adenocarcinoma cell proliferation, migration, invasion, and epithelial-mesenchymal transition progression. *Bioengineered* 13 (3), 4964–4977. doi:10.1080/21655979.2022.2037871

Jiang, N., Zhang, X., Gu, X., Li, X., and Shang, L. (2021). Progress in understanding the role of lncRNA in programmed cell death. *Cell. Death Discov.* 7 (1), 30. doi:10.1038/s41420-021-00407-1

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). Timer: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Liu, L., Tang, Z., Zeng, Y., Liu, Y., Zhou, L., Yang, S., et al. (2021). Role of necroptosis in infection-related, immune-mediated, and autoimmune skin diseases. *J. Dermatol.* 48 (8), 1129–1138. doi:10.1111/1346-8138.15929

Ma, C., Luo, H., Cao, J., Zheng, X., Zhang, J., Zhang, Y., et al. (2020). Identification of a novel tumor microenvironment-associated eight-gene signature for prognosis prediction in lung adenocarcinoma. *Front. Mol. Biosci.* 7, 571641. doi:10.3389/fmolb.2020.571641

Malireddi, R. K. S., Gurung, P., Kesavardhana, S., Samir, P., Burton, A., Mummareddy, H., et al. (2020). Innate immune priming in the absence of TAK1 drives RIPK1 kinase activity-independent pyroptosis, apoptosis, necroptosis, and inflammatory disease. *J. Exp. Med.* 217 (3), e20191644. doi:10.1084/jem.20191644

Mao, Y., Yang, D., He, J., and Krasna, M. J. (2016). Epidemiology of lung cancer. *Surg. Oncol. Clin. N. Am.* 25 (3), 439–445. doi:10.1016/j.soc.2016.02.001

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Molnár, T., Mazlo, A., Tslaf, V., Szollosi, A. G., Emri, G., and Koncz, G. (2019). Current translational potential and underlying molecular mechanisms of necroptosis. *Cell. Death Dis.* 10 (11), 860. doi:10.1038/s41419-019-2094-z

Nasim, F., Sabath, B. F., and Eapen, G. A. (2019). Lung cancer. *Med. Clin. North Am.* 103 (3), 463–473. doi:10.1016/j.mcna.2018.12.006

Pan, Y., Fu, Y., Zeng, Y., Liu, X., Peng, Y., Hu, C., et al. (2022). The key to immunotherapy: How to choose better therapeutic biomarkers for patients with non-small cell lung cancer. *Biomark. Res.* 10 (1), 9. doi:10.1186/s40364-022-00355-7

Park, E. G., Pyo, S. J., Cui, Y., Yoon, S. H., and Nam, J. W. (2022). Tumor immune microenvironment lncRNAs. *Brief. Bioinform.* 23 (1), bbab504. doi:10.1093/bib/bbab504

Plattner, C., Finotello, F., and Rieder, D. (2020). Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. *Methods Enzymol.* 636, 261–285. doi:10.1016/bs.mie.2019.05.056

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 6, e26476. doi:10.7554/eLife.26476

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *Ca. Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654

Song, J., Sun, Y., Cao, H., Liu, Z., Xi, L., Dong, C., et al. (2021). A novel pyroptosis-related lncRNA signature for prognostic prediction in patients with lung adenocarcinoma. *Bioengineered* 12 (1), 5932–5949. doi:10.1080/21655979.2021.1972078

Su, Z., Yang, Z., Xie, L., DeWitt, J. P., and Chen, Y. (2016). Cancer therapy in the necroptosis era. *Cell. Death Differ.* 23 (5), 748–756. doi:10.1038/cdd.2016.8

Tang, H., Wang, Z., Shao, Q., Wang, Y., and Yang, Q. (2020). Comprehensive analysis of competing endogenous RNA (ceRNA) network based on RNAs differentially expressed in lung adenocarcinoma using the cancer genome Atlas (TCGA) database. *Med. Sci. Monit.* 26, e922676. doi:10.12659/MSM.922676

Tang, L., Chen, Y., Peng, X., Zhou, Y., Jiang, H., Wang, G., et al. (2020). Identification and validation of potential pathogenic genes and prognostic markers in ESCC by integrated bioinformatics analysis. *Front. Genet.* 11, 521004. doi:10.3389/fgene.2020.521004

Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J., et al. (2020). Ferroptosis, necroptosis, and pyroptosis in anticancer immunity. *J. Hematol. Oncol.* 13 (1), 110. doi:10.1186/s13045-020-00946-7

Wang, H., Li, M., Wang, Y., and Wang, L. (2022). Construction of a nomogram based on lncRNA and patient's clinical characteristics to improve the prognosis of non-small cell lung cancer. *Technol. Cancer Res. Treat.* 21, 15330338221097215. doi:10.1177/15330338221097215

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Yan, J., Wan, P., Choksi, S., and Liu, Z. G. (2022). Necroptosis and tumor progression. *Trends Cancer* 8 (1), 21–27. doi:10.1016/j.trecan.2021.09.003

Yao, J., Chen, X., Liu, X., Li, R., Zhou, X., and Qu, Y. (2021). Characterization of a ferroptosis and iron-metabolism related lncRNA signature in lung adenocarcinoma. *Cancer Cell. Int.* 21 (1), 340. doi:10.1186/s12935-021-02027-2

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

Yu, T. T., Sang, X. Y., Han, N., Peng, X. C., Li, Q. R., Xu, X., et al. (2021). Macrophages mediated delivery of chlorin e6 and treatment of lung cancer by photodynamic reprogramming. *Int. Immunopharmacol.* 100, 108164. doi:10.1016/j.intimp.2021.108164

Yuan, J., Amin, P., and Ofengeim, D. (2019). Necroptosis and RIPK1-mediated neuroinflammation in CNS diseases. *Nat. Rev. Neurosci.* 20 (1), 19–33. doi:10.1038/s41583-018-0093-1

Zhao, Z., Liu, H., Zhou, X., Fang, D., Ou, X., Ye, J., et al. (2021). Necroptosis-related lncRNAs: Predicting prognosis and the distinction between the cold and hot tumors in gastric cancer. *J. Oncol.* 2021, 6718443. doi:10.1155/2021/6718443

Zhu, Y., Xie, J., and Shi, J. (2021). Rac1/ROCK-driven membrane dynamics promote natural killer cell cytotoxicity via granzyme-induced necroptosis. *BMC Biol.* 19 (1), 140. doi:10.1186/s12915-021-01068-3

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fgene.2022.953322

OPEN ACCESS

EDITED BY
Tugba Önal-Süzek,
Muğla University, Turkey

REVIEWED BY
Jian Chen,
BioAtla, Inc., United States
Shaogang Wang,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Guangzhen Wu,
wuguang0613@hotmail.com
Qifei Wang,
wangqifei6008@hotmail.com

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

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

RECEIVED 26 May 2022
ACCEPTED 24 August 2022
PUBLISHED 15 September 2022

CITATION
Che X, Li J, Xu Y, Wang Q and Wu G (2022), Analysis of genomes and transcriptomes of clear cell renal cell carcinomas identifies mutations and gene expression changes in the TGF-beta pathway.
Front. Genet. 13:953322.
doi: 10.3389/fgene.2022.953322

Analysis of genomes and transcriptomes of clear cell renal cell carcinomas identifies mutations and gene expression changes in the TGF-beta pathway

Xiangyu Che[1†], Jianyi Li[2†], Yingkun Xu[3], Qifei Wang[1*] and Guangzhen Wu[1*]

[1]Department of Urology, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, China, [2]Organ Transplant Center, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [3]Department of Endocrine and Breast Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

The occurrence of clear cell renal cell carcinoma (ccRCC) is related to changes in the transforming growth factor-β (TGF-β) signaling pathway. In this study, we adopted an integrated approach to identify and verify the effects of changes in this pathway on ccRCC and provide a guide for identifying new therapeutic targets. We performed transcriptome analysis of 539 ccRCC cases from The Cancer Genome Atlas (TCGA) and divided the samples into different TGF-β clusters according to unsupervised hierarchical clustering. We found that 76 of the 85 TGF-β pathway genes were dysregulated, and 55 genes were either protective or risk factors affecting the prognosis of ccRCC. The survival time of patients with tumors with low TGF-β scores was shorter than that of patients with tumors with high TGF-β scores. The overall survival (OS) of patients with ccRCC with high TGF-β scores was better than that of patients with low TGF-β scores. The TGF-β score correlated with the expression of key ccRCC and deacetylation genes. The sensitivity of tumor patients to targeted drugs differed between the high and low TGF-β score groups. Therefore, a prognostic model based on the TGF-β gene pathway can predict the prognosis of ccRCC patients. Grouping patients with ccRCC according to their TGF-β score is of great significance for evaluating the prognosis of patients, selecting targeted drugs, and identifying new therapeutic targets.

KEYWORDS

TGF-β, ccRCC, TCGA, prognosis, treatment

Introduction

As one of the most common cancers, the number of new cases and deaths of renal cell carcinoma (RCC) has remained high in recent years (Siegel et al., 2018; Siegel et al., 2019; Siegel et al., 2020). ccRCC accounts for 75%–80% of RCCs (Leibovich et al., 2010) and often warrants radiotherapy or chemotherapy (Coppin et al., 2011). At the time of preliminary diagnosis, 20%–30% of patients with RCC have local or distant metastases (Gong et al., 2016). Targeted drug therapy is a common treatment for these patients; however, one of the leading causes of treatment failure is drug resistance. Several studies have revealed the molecular mechanisms and signaling pathways of RCC, including TGF-β, Wnt-β-catenin, and angiogenesis signal transduction. The TGF-β signaling pathway, as one of the common signaling pathways, plays an important yet complex role in the occurrence and development of RCC and has a significant impact on tumor metastasis and prognosis (Kominsky et al., 2007; Wang et al., 2020a). At present, most studies suggest that the TGF-β pathway is involved in the regulation of tumors as a cancer-promoting factor (Bao et al., 2020; Du et al., 2020). However, some studies have shown that TGF-β can induce apoptosis in renal cancer cells, and c-Ski can weaken the anti-tumor effect of TGF-β by inhibiting TGF-β signal transduction (Taguchi et al., 2019).

Here, we used data from TCGA to systematically analyze the genetic changes, prognosis, and treatment-related information on TGF-β-related genes in RCC and to explore the role of the TGF-β signaling pathway in RCC.

Materials and methods

Data acquisition and analysis

Through the R/BioConductor package of TCGAbiolinks, we downloaded the RNA-seq transcriptome data on the ccRCC group from the Genomic Data Commons (GDC) portal (Colaprico et al., 2016); the data included 539 cases of ccRCC tissue and 72 cases of normal renal tissue. A total of 85 genes related to the TGF pathway were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) TGF-β signaling pathway on the Gene Set Enrichment Analysis (GSEA) website. The clinical information on the cancer patients was obtained from TCGAbiolinks, including the tumor size (T) status, metastatic (M) status, tumor grade, tumor stage, age, and survival status. The Lasso regression analysis was carried out with "glmnet" and "Survival" packages. Univariate and multivariate COX risk analyses of clinical features were performed with the "Survival" package. The correlation of immune infiltration was analyzed with the "ggstatsplot" package.

Genetic alteration and survival analysis

The differential expression of TGF-β pathway genes in ccRCC and normal renal tissues was analyzed using the "limma" package, and the effect of TGF-β pathway genes on the prognosis of patients with RCC was analyzed using the "Survival" package. We downloaded the single-nucleotide variation (SNV) data and expression data on TGF-β pathway genes in different cancer types from TCGA database (Tomczak et al., 2015), analyzed them using Perl language, and visualized them with TBtools software (Chen et al., 2020).

Cluster analysis based on transforming growth factor-β scores

We constructed a TGF-β scoring model to show the differences between samples. Based on the expression characteristics of normal renal tissues, we divided the renal carcinoma tissues into three categories: the TGF-β normal-score group (cluster 1), TGF-β high-score group (cluster 2), and TGF-β low-score group (cluster 3 and cluster 4). We used violin plots to describe the relationship between normal tissues and gene expression levels in these three groups. We plotted survival curves for the three clusters using the "Survival" package. We used "pheatmap" to draw a heat map showing the relationship between these three clusters and the clinicopathological features of ccRCC patients.

Prediction of targeted drug response

We predicted the therapeutic response based on the Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang et al., 2012). The R package "pRRophetic" was used for the prediction process; ridge regression was used to estimate the half-maximal inhibitory concentration (IC50) of the sample, and the 10-fold cross-validation based on the GDSC training set was used to evaluate the prediction accuracy (Geeleher et al., 2014a; Lu et al., 2019). All parameters were set to default values with the removal of the batch effect of "combat" and tissue type of "allSoldTumours," and duplicate gene expression was summarized as the mean value (Geeleher et al., 2014b). The multiple testing correction used was a Bonferroni adjustment.

Immune cell infiltration and immunotherapy

We analyzed immune cells quantitatively using the single-sample gene set enrichment analysis (ssGSEA) combined with the expression of TGF-β-related genes (Zhang et al., 2018). The heat map was drawn using "ggplot2" and "dplyr" in R. We used

five R software packages, "ggplot2," "dplyr," "data.table," "tidyr," and "ggstatsplot" to analyze and plot the correlation between the TGF-β score and immune factors. We selected two classical immune regulators: type II interferon response and mast cells. Then, we used the "ggdisterstats" package to show their correlation with the TGF-β score in the form of a scatter plot. Then, we adopted the "ggdisterstats" package to make scatter diagrams to show both of their correlations with the TGF-β score. Visual correlation matrix analysis was used to show the relationship between programmed cell death 1(PD1 and PDCD1), cytotoxic T-lymphocyte-associated protein-4 (CTLA-4), and TGF-β scores. We used the subclass mapping and tumor immune dysfunction and rejection (TIDE) algorithm to predict the clinical response of RCC to block immune checkpoints PD-1 and CTLA-4 (Hoshida et al., 2007; Jiang et al., 2018). Bonferroni correction was used in this process.

Establishment of the Lasso regression prognostic model

Based on its statistical significance, we first selected the TGF-β gene related to survival ($p < 0.05$). We then used a Lasso regression analysis to delete genes that may overfit the model. Finally, a multivariate analysis was used to determine the optimal predictive factors of the model. The analysis used the following formula: Risk score = Σ Ni = 1 (Expi × Coei). The median was set as the cut-off value, according to which all patients with ccRCC were divided into two groups: the low-risk group and high-risk group. The overall survival time-dependent recipient operating characteristics were used to evaluate the accuracy of the prognostic model. Taking the median as the cut-off value, all patients with ccRCC were divided into the low-risk group and high-risk group. The overall survival time-dependent recipient operating characteristics were used to evaluate the accuracy of the prognostic model. To evaluate the accuracy of the prognostic model, we adopted the OS time-dependent receiver operating characteristic (ROC) curve.

Compounds targeting transforming growth factor-β pathway genes

The Baird Institute's public online Connectivity Map (CMAP) Build02 (Lamb et al., 2006) (https://portals.broadinstitute.org/cmap/) can predict compounds that activate or inhibit targets based on gene expression signatures, and we used this tool to explore which drugs may target TGF-β pathway genes. We further used the CmapTools to conduct a special analysis to explore the mechanism of action of the compound (https://clue.io/) (Subramanian et al., 2017). Similar to the GSEA analysis, CMAP uses the pattern-matching strategy of the Kolmogorov–Smirnov test to find the similarity between differentially expressed genes (DEGs). Then, we compared the results of the CMAP analysis to the DEG ranking list to determine the positive or negative regulatory relationship of genes and to generate enrichment scores (ES) from −1 to 1. Finally, we used the aforementioned scores to rank all the case data in the database. After we obtained two tables in each type of tumor, we applied the results of the connection map to the expression signatures of the TGF-β pathway and then used $p < 0.05$ as our inclusion criteria to determine the average meaningful compounds for each type of tumor.

Results

Transforming growth factor-β pathway genes were significantly differentially expressed in clear cell renal cell carcinoma samples compared with normal samples and were closely related to prognosis

From the KEGG TGF-β signaling pathway on the GSEA website, we found 85 genes related to the TGF-β signaling pathway (Supplementary Table S1) and analyzed the RNA sequencing data on 539 patients with human ccRCC and 72 normal kidney samples. We found that 76 genes of the TGF-β pathway were dysregulated in ccRCC compared with normal kidney tissues, of which 39 genes were upregulated and 37 were downregulated (Figure 1A; Supplementary Table S2). A total of 36 genes were upregulated and 40 were downregulated compared with those in normal kidney tissues. Furthermore, we analyzed the effects of these DEGs related to the TGF-β pathway on the overall survival of patients with ccRCC. The results showed that 55 genes significantly affected the prognosis of patients with ccRCC, of which 19 genes were risk factors (hazard ratio >1) and 36 genes were protective factors (hazard ratio <1) (Figure 1B; Supplementary Table S3).

Unsupervised hierarchical clustering and prognostic analysis

Unsupervised hierarchical clustering of the data revealed four different clusters of the TGF-β pathway (Figure 2A). Cluster 1 showed that the expression levels of 85 genes were similar to those in the normal samples, indicating that the TGF-β pathway was in a normal state. However, the expression levels of genes related to the TGF-β pathway in cluster 2 were increased, indicating high TGF-β scores, while cluster 3 and cluster 4 showed low expression of TGF-β pathway genes, indicating low TGF-β scores. Figure 2B depicts the TGF-β scores in five clusters (where cluster 5 represents normal kidney tissue) and further shows the differences in TGF-β scores between them. Supplementary Table S4 shows the TGF-β scores for each

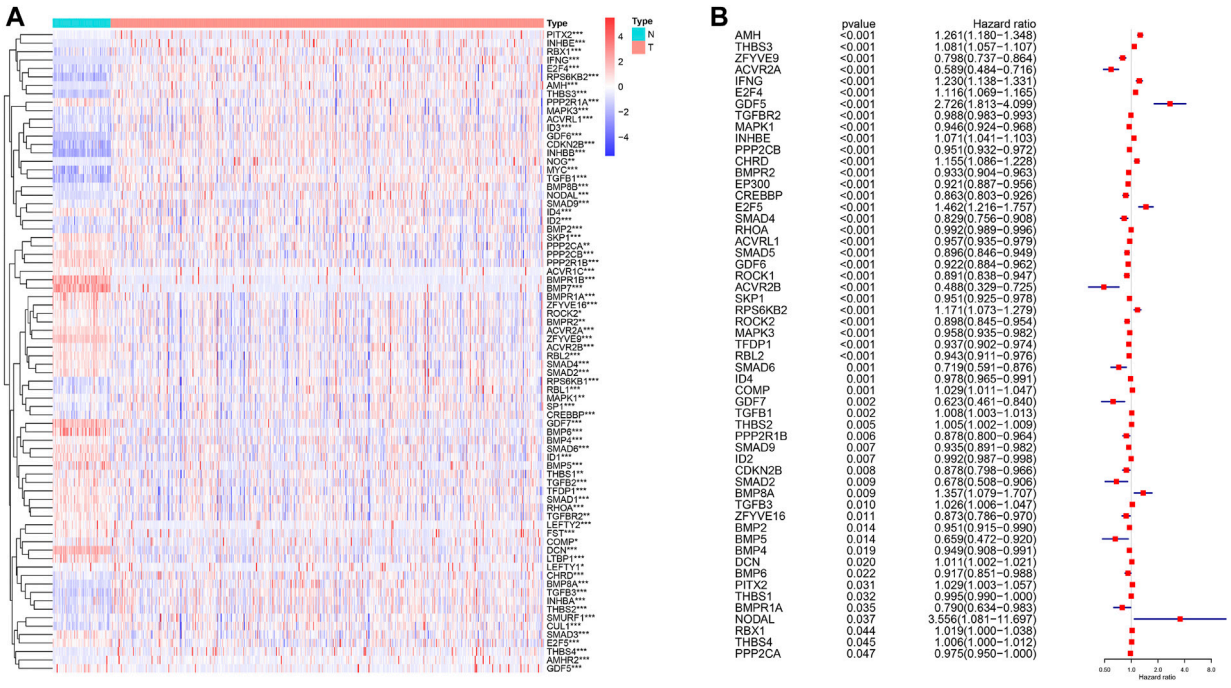


FIGURE 1
Differential expression of TGF-β pathway genes in ccRCC and their effect on prognosis. **(A)** Dysregulation of the TGF-β pathway genes in ccRCC. N: normal kidney tissue, T: kidney tumor tissue. Blue represents a low level of gene expression, red represents a high level of gene expression, and the color depth represents the level of gene expression. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. **(B)** TGF-β pathway genes that affect the prognosis of patients with ccRCC. A hazard ratio <1 represents a protective factor for prognosis, while a hazard ratio >1 represents a risk factor for prognosis. ccRCC: clear cell renal cell carcinoma.

sample. We then further classified all of the samples into three groups; the normal group, the KEGG-TGF-β high-score group, and the KEGG-TGF-β low-score group. In other words, clusters 3 and 4 were merged into the latter group. We then analyzed the correlation between the TGF-β score and clinicopathological features, and the results showed that the expression of TGF-β pathway genes was significantly correlated with the overall survival rate of patients with ccRCC (Figure 2C). Further survival analysis showed that the KEGG-TGF-β low-score group had the worst prognosis, while the high-score group had the best prognosis (Figure 2D).

Disruption of the transforming growth factor-β signaling pathway is closely related to dysregulation of key and deacetylated genes in clear cell renal cell carcinoma

We explored the relationship between the expression of various well-known key genes and the TGF-β pathway in ccRCC (Figure 3A). In other words, VHL, TP53, and PTEN were downregulated in the KEGG-TGF-β low-score group. These results suggest that disruption of the TGF-β signaling pathway is related to the promotion of tumors. *EGFR*, *MYC*, *VEGFA*, and other oncogenes were highly expressed in the KEGG-TGF-β high-score group, indicating that it may be more effective to target these genes in this group. Analysis of the transcriptome of ccRCC patients also showed a strong correlation between the abnormal expression of sirtuins and histone deacetylases (HDACs) and the abnormality of the TGF-β pathway (Figure 3B).

Prediction of IC50 of different targeted drugs based on the transforming growth factor-β score

Considering that targeted drugs are commonly used in the treatment of metastatic RCC, we evaluated the efficacy of different targeted drugs on the TGF-β signaling pathway in the KEGG-TGF-β high-score and low-score groups. We obtained a satisfactory prediction model using ridge regression on the GDSC cell-line dataset. Based on this model, we evaluated the IC50 values of 12 targeted drugs. The results of the analysis showed that there was no significant

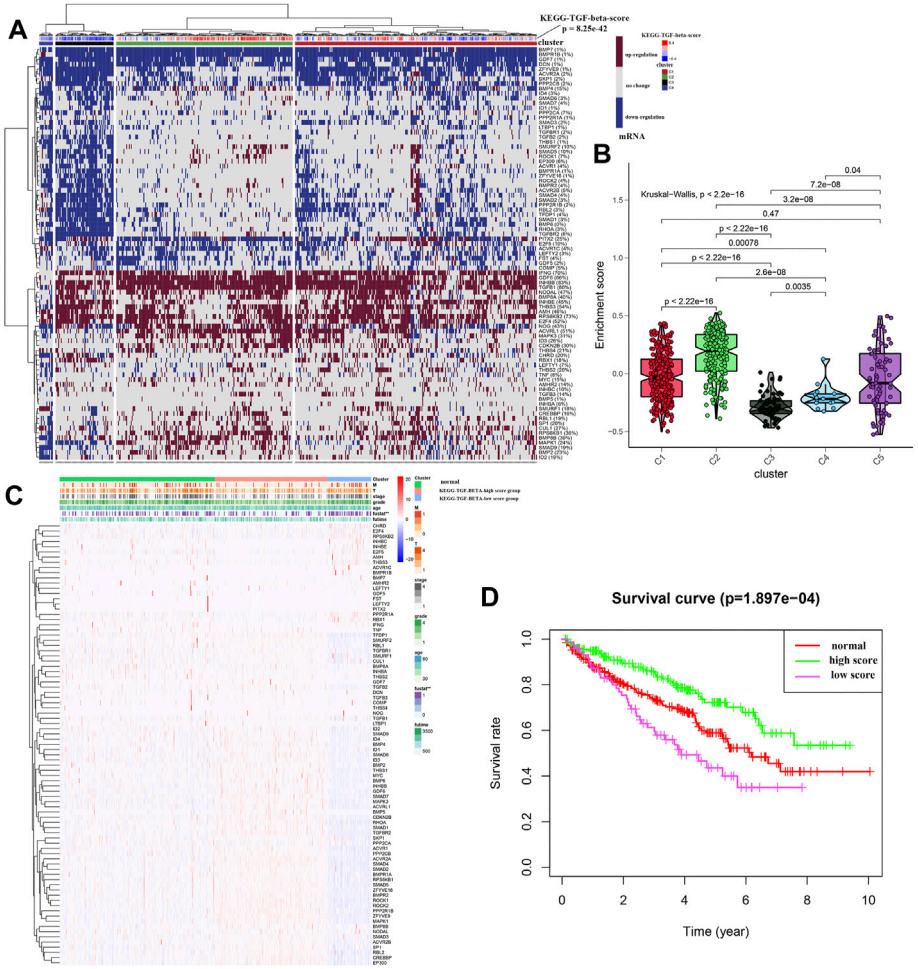


FIGURE 2
Unsupervised hierarchical clustering of ccRCC samples. **(A)** Cluster analysis of transcriptome data from 539 ccRCC samples from TCGA. **(B)** Box plot showing the activity score of the TGF-β pathway in each of the four clusters. **(C)** Relationship between the risk score and clinicopathological features. **(D)** Kaplan–Meier survival curves of the high-, low-, and normal-score clusters. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

difference in the IC50 values of pazopanib, gefitinib, bosutinib, and lapatinib among the three groups (Figures 4A,D,G,I). Compared with the KEGG-TGF-β high-score and low-score groups, IC50 of temsirolimus and sunitinib was lower in the normal group (Figures 4E,F). Therefore, it is recommended that these two drugs be used for the treatment of the normal group. The IC50 values of imatinib, nilotinib, and axitinib were lower in the KEGG-TGF-β high-score group than those in the normal and the KEGG-TGF-β low-score groups (Figures 4B,C,L). This indicates that these three drugs may have better results in the KEGG-TGF-β high-score group. Compared with the normal and KEGG-TGF-β high-score groups, the IC50 values of metformin, tipifarnib, and sorafenib were lower than those in the KEGG-TGF-β low-score group (Figures 4H,J,K). This indicates that three drugs may have better results in the KEGG-TGF-β low-score group.

The transforming growth factor-β pathway is related to immune regulation

Immune regulation plays a vital role in the tumor microenvironment. We identified that many TGF-β genes were associated with the infiltration of many types of immune cells (Figure 5A). We then analyzed the correlation between the immune infiltration and TGF-β score and found that there was a close relationship between them (Figure 5B), especially Type-II-IFN-response and mast cells, which were positively correlated with the TGF-β score (Figures 5C,D). Next, we analyzed the correlation between the TGF-β score and immune checkpoints, and the results showed that the former was negatively correlated with CTLA4 and PDCD1 (Figure 5E). We used the TIDE algorithm to predict the response of the KEGG-TGF-β low-score and high-score groups to immune checkpoint inhibitors;

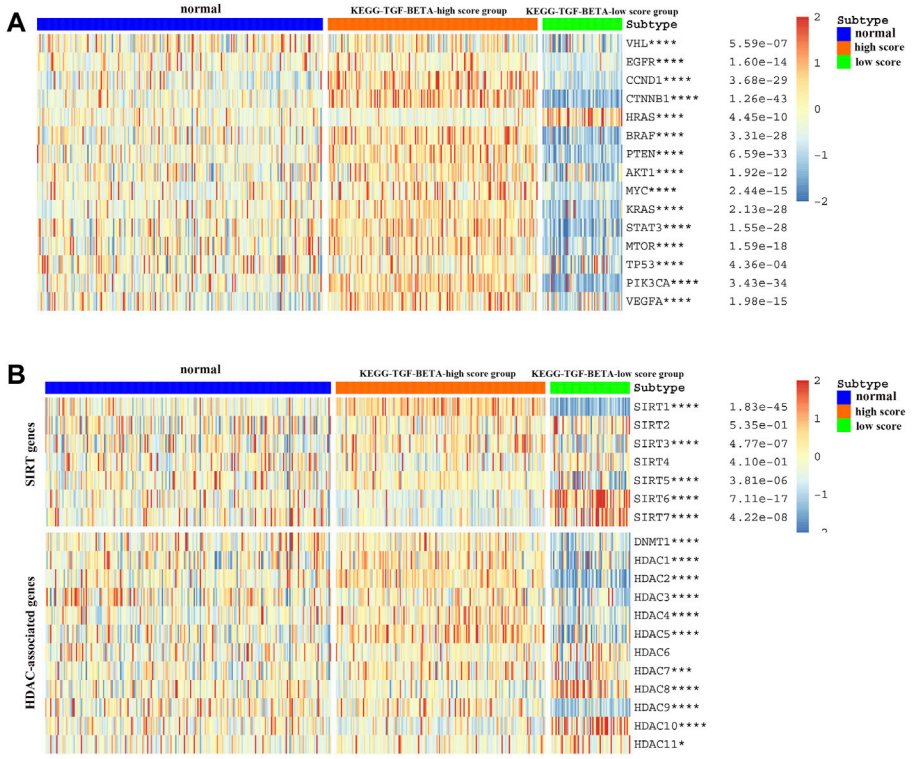


FIGURE 3
Relationship between the TGF-β score and the expression of other key genes. **(A)** Relationship between the activation level of the TGF-β pathway and the expression of oncogenes and tumor suppressor genes. **(B)** Relationship between the activation level of the TGF-β pathway and the expression of deacetylation genes. Blue represents the TGF-β normal-score group, orange represents the TGF-β high-score group, and green represents the TGF-β low-score group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

however, after Bonferroni correction, we did not find a significant difference in the response of the groups to immune checkpoint inhibitors (Figure 5F).

Construction and verification of a new transforming growth factor-β-based survival model

First, TGF-β genes related to survival were screened according to the survival analysis and significance value ($p < 0.05$). The Lasso regression model was then used to analyze and determine the most reliable prognostic markers. The number of points at which the vertical line intersects the curve at the same site in Figure 6A is the number of variables for when the fit is optimal, which indicates the number of selected genes. On this basis, five genes *ZFYVE9*, *ACVR2A*, *IFNG*, *AMH*, and *THBS3* were selected to establish a risk-characteristic model based on the minimum criterion (Figures 6A,B). Then, according to the median risk score, we divided ccRCC patients into low- and high-risk groups and studied the prognostic performance of the new survival model composed of five genetic risk characteristics. Kaplan–Meier survival analysis revealed that patients in high-risk groups had a significantly lower survival rate than patients in low-risk groups (Figure 6C). In addition, ROC curve analysis was performed to analyze the prognostic performance of the new survival model in patients with ccRCC. The area under the curve (AUC) value was 0.728 for the 5-year survival rate, 0.744 for the 7-year survival rate, and 0.752 for the 10-year survival rate. Therefore, the prognostic model of ccRCC based on the TGF-β pathway has a relatively high predictive value (Figures 6D–F). To better understand the relationship between clinicopathological characteristics of ccRCC patients and TGF-β pathway genes, we systematically analyzed the correlation between risk scores based on five TGF-β pathway genes and clinicopathological characteristics of ccRCC patients. We found that the risk score was closely related to (T) and (M) status, tumor grade, stage, and OS (Figure 6G). Univariate Cox regression analysis showed that these features along with the risk score were associated with overall survival in ccRCC patients (Figure 6H; Supplementary Table S5). Multivariate Cox regression analysis showed that

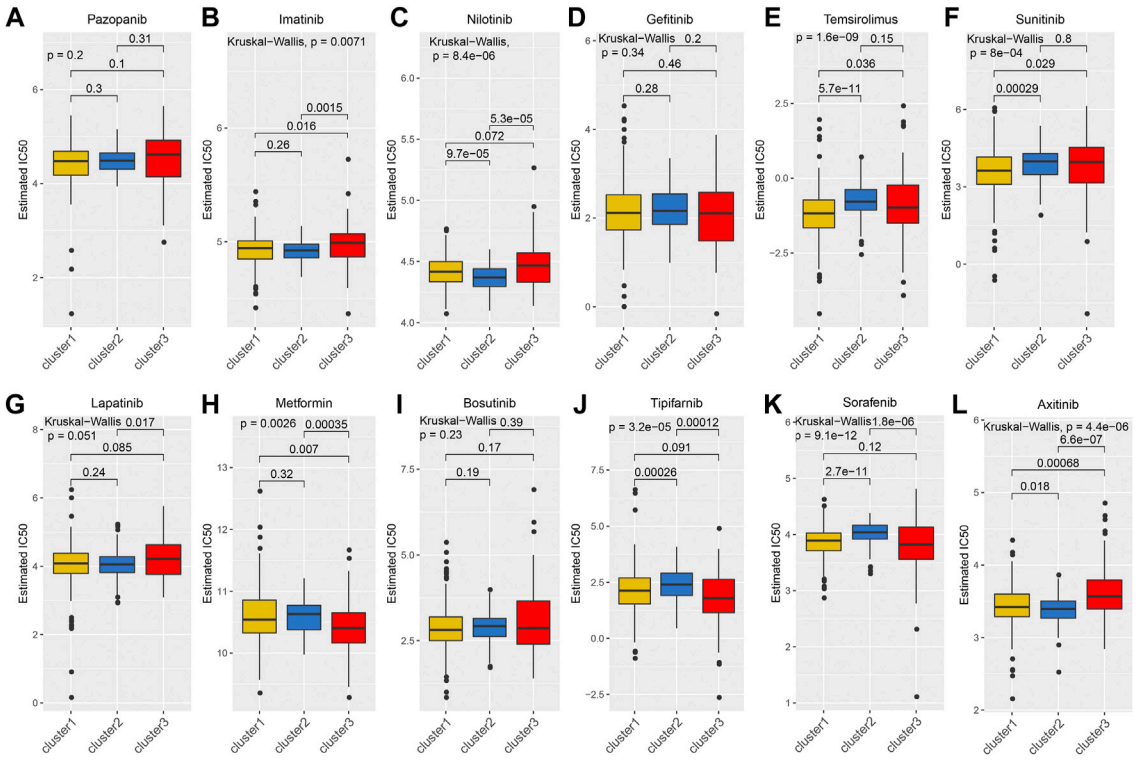


FIGURE 4
Prediction of IC50 values of targeted drugs. **(A)** pazopanib, **(B)** imatinib, **(C)** nilotinib, **(D)** gefitinib, **(E)** temsirolimus, **(F)** sunitinib, **(G)** lapatinib, **(H)** metformin, **(I)** bosutinib, **(J)** tipifarnib, **(K)** sorafenib, and **(L)** axitinib.

age, grade, stage, and risk score were independent risk factors affecting the prognosis of patients with ccRCC (Figure 6I; Supplementary Table S6). We used a nomogram to predict the ccRCC risk; 5-year, 7-year, and 10-year survival rates were estimated based on the patient's age, grade, stage, and risk score (Figure 6J).

Transforming growth factor-β pathway genes undergo a wide range of genetic changes across cancer types and affect the prognosis of many cancers

To further explore the genetic changes in TGF-β pathway genes in pan-tumors, we analyzed SNV and gene expression changes of TGF-β pathway genes across multiple cancer types. The results showed that TGF-β pathway genes had a wide range of SNVs (Figure 7A; Supplementary Table S7) and gene expression differences (Figure 7B; Supplementary Table S8) across the different cancer types. We then analyzed the effect of TGF-β pathway genes on the prognosis of cancer patients (Figure 7C; Supplementary Table S9). The results showed that most of the TGF-β pathway genes were risk factors in other types of tumors, but were protective factors in ccRCC, which was consistent with our previous analysis that the KEGG-TGF-β low-score group had a worse prognosis.

Connectivity map analysis identified potential compounds/inhibitors targeting TGF-β

Considering that most TGF-β pathway genes are risk factors in tumors, we aimed to identify compounds that can inhibit TGF-β pathway genes. We used a data-driven systematic method, CMAP, to explore the relationship between genes, compounds, and biological conditions to identify candidate compounds that may target TGF-β pathway genes. We found 54 compounds inhibiting TGF-β pathway genes that were enriched in different tumors (Figure 8A). Simultaneously, we explored the possible action mechanisms of 19 small molecular compounds and found that the compounds involved 18 mechanisms, where two compounds had the same mechanism (Figure 8B). Therefore, this suggests that we can select different compounds in different tumors and suppress the TGF-β

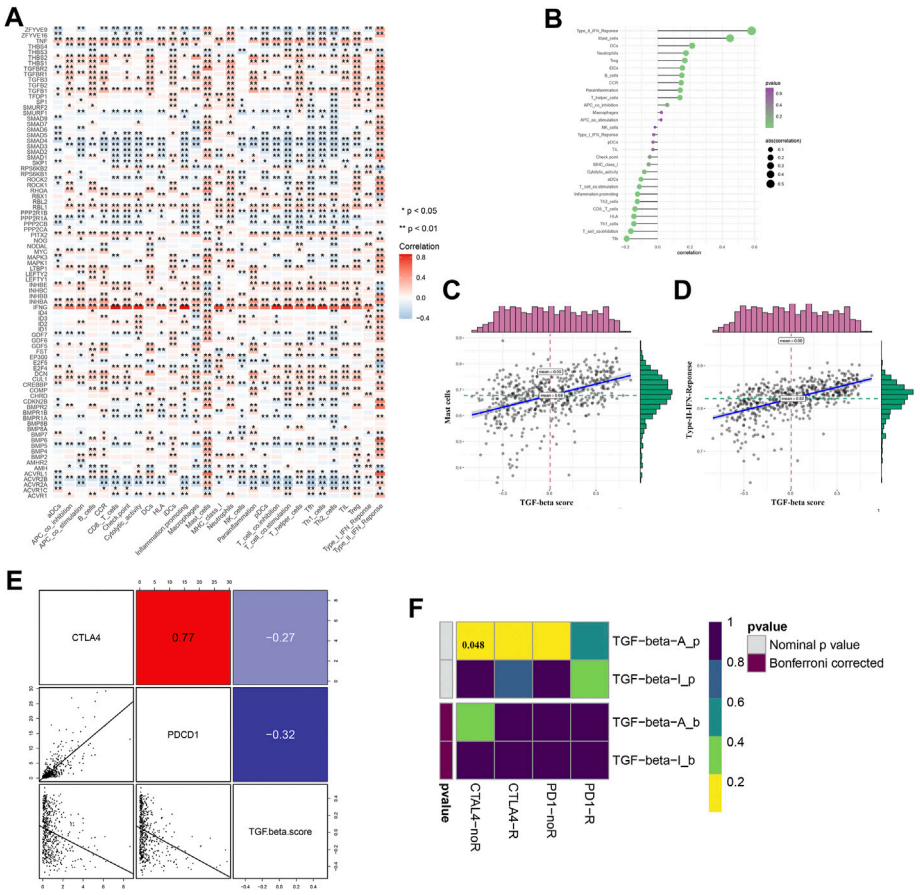


FIGURE 5
Relationship between TGF-β pathway genes and immune regulation. **(A)** Relationship between TGF-β pathway genes and immune cell infiltration. **(B–D)** Relationship between the TGF-β score and immune infiltration. **(E)** Relationship between the TGF-β score and immune checkpoint expression (Pearson's correlation analysis). **(F)** Prediction of immune checkpoint inhibitor response in TGF-β activation and inactivation groups. Blue represents a positive correlation, and red represents a negative correlation. Pearson's correlation analysis was used. *$p < 0.05$ and **$p < 0.01$.

pathway genes, according to different mechanisms to achieve tumor inhibition.

Discussion

Our comprehensive analysis of a large number of open access RCC cases provides new insights into the key role of the TGF-β pathway in the occurrence and development of RCC. Previous studies have reported the effects of various TGF-β pathway genes on RCC. For example, TGF-β1 enhances the proliferation and metastatic potential of RCC cells by upregulating lymphoid enhancer-binding factor 1/integrin αMβ2 (Liu and Shang, 2020), and MUC12 relies on TGF-β1 signaling to mediate the growth and invasion of renal cancer cells (Gao et al., 2020).

However, we found that when we analyzed all genes of the TGF-β pathway as a whole, the results were surprising. Analysis of TCGA database revealed that 76 of the 85 TGF-β pathway genes were significantly differentially expressed between RCC and normal renal tissues, and 55 genes could play a pivotal role in the prognosis of patients with RCC. The results of this analysis piqued our interest in the role of TGF-β pathway genes in RCC.

In the three groups that we divided the ccRCC samples into (normal, KEGG-TGF-β high-score, and KEGG-TGF-β low-score), the degree of TGF-β gene expression, prognosis, and response to drugs differed. Similar to hepatocellular carcinoma (Chen et al., 2018), the prognosis of the TGF-β high-score group was better and that of the low-score was poorer in RCC.

We observed a correlation between the expression of many well-known genes related to RCC and the expression of TGF-β pathway genes. The VHL gene was expressed at significantly lower levels in the KEGG-TGF-β low-score group, and the loss of VHL gene function often leads to the pathogenesis of RCC. Similarly, the expression levels of well-known tumor suppressor genes such as PTEN and P53 were also low in the KEGG-TGF-β low-score group, which explains the poor prognosis of this

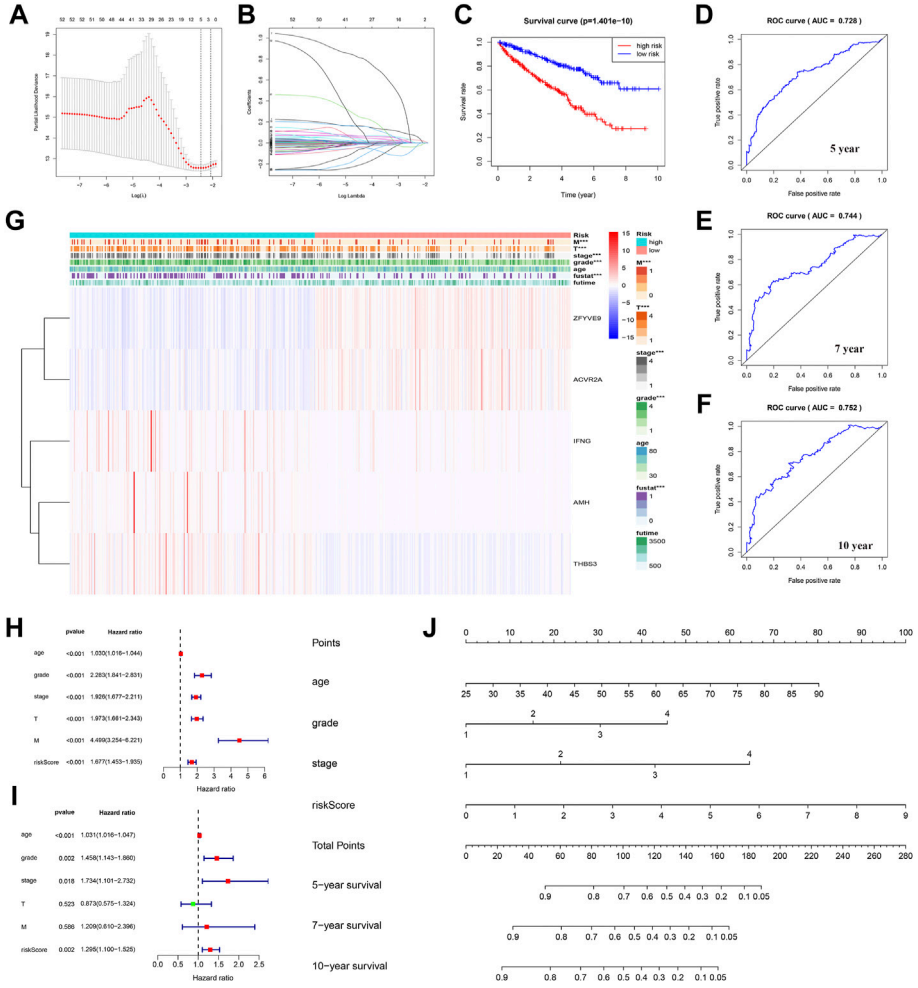


FIGURE 6
Establishment of a prognostic model based on TGF-β in ccRCC. **(A)** Partial likelihood of deviance was plotted against log(lambda). **(B)** Lasso coefficient profiles of TGF-β in ccRCC. (**C**) Grouped according to the risk scores calculated by the new survival model based on TGF-β expression; the Kaplan–Meier survival curve shows the overall survival rate of ccRCC patients in both the high- and low-risk groups. **(D–F)** ROC curve analysis showed that the new survival model was efficient in predicting prognosis, and the AUC values for 5-, 7-, and 10-year survival were 0.728, 0.744, and 0.752, respectively. **(G)** Relationship between the risk score and clinicopathological features. **(H)** Univariate Cox regression analysis showed that clinicopathological parameters such as age, grade, stage, tumor size (T), tumor metastasis (M), and the risk score of the new survival model were correlated with OS in patients with renal cell carcinoma (RCC). **(I)** Multivariate Cox regression analysis showed that risk score, age, grade, and stage were independent risk factors affecting the prognosis of patients with ccRCC. **(J)** New nomogram predicted 5-, 7-, or 10-year survival rates in patients with ccRCC. The value of each variable is a fraction along the dotted axis. The nomogram has nine lines. The second, third, fourth, and fifth lines represent the age, grade, stage, and risk score, respectively. The total score in the sixth row is derived from the sum of each score of age, grade, stage, and risk score. Based on the total score of ccRCC patients, the 5-, 7-, and 10-year survival rates of patients can be estimated.

group. However, a related study showed that the synergistic effect of TGF-β type I receptor and hypoxia-inducible factor-α (HIF-α) promotes the progression of RCC (Mallikarjuna et al., 2019). There are some obvious differences between this study and our research results, which may warrant further study. In the KEGG-TGF-β low-score group, we found some highly expressed oncogenes, such as *EGFR*, *MYC*, *MTOR*, and *VEGFA*, which play a key role in the occurrence and development of RCC. Therefore, compared with the KEGG-TGF-β low-score group, patients in the KEGG-TGF-β high-score group may have a better therapeutic outcome if these oncogenes are used as therapeutic targets.

Acetylation and deacetylation are common epigenetic modifications that play vital roles in the formation and development of tumors. Analysis of TCGA database transcriptome also showed a strong correlation between the aberrant expression of sirtuins and HDACs and the abnormal expression of the TGF-β pathway in patients with RCC. SIRT1, SIRT3, and SIRT5 were significantly correlated with a high TGF-β score, while the expression levels of SIRT6 and SIRT7 were

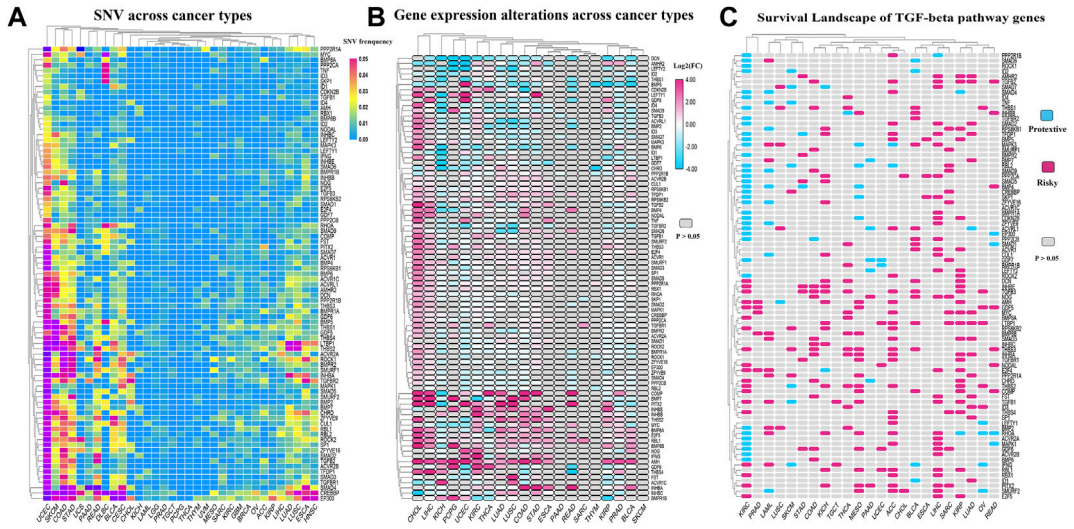


FIGURE 7
Genetic changes of TGF-β pathway genes across cancer types. **(A)** SNVs of TGF-β pathway genes across cancer types. **(B)** Alterations in the expression of TGF-β pathway genes across cancer types. **(C)** Risk assessment of the effect of the TGF-β pathway genes on prognosis. KIRC: kidney renal clear cell carcinoma

significantly correlated with a low TGF-β score. In the KEGG-TGF-β low-score group, the expression of SIRT3 was significantly low. Previous studies have shown that low SIRT3 expression in RCC is associated with poor prognosis (Jeh et al., 2017), which supports our results. Transcriptome analysis of sirtuins and HDACs also indicated that the expression of these proteins in different TGF-β feature groups was different; therefore, selecting different targets in different TGF-β feature groups may be more effective for therapy.

Currently, targeted drug therapy is commonly used for local or distant metastatic RCC, and it is still worth discussing which targeted drug can benefit patients the most. The vasculature-rich nature of RCC has led to the approval of tyrosine kinase inhibitors, including sorafenib, sunitinib, pazopanib, and axitinib (Escudier et al., 2007; Motzer et al., 2007; Rini et al., 2011; Motzer et al., 2014), targeting the VEGF signaling axis as first- and second-line therapies for metastatic RCC in the United States and the European Union. Pazopanib as a first-line targeted agent is similar to sunitinib in improving progression-free survival (PFS) and overall survival (OS) (Motzer et al., 2013; Motzer et al., 2014). The choice between these two agents for patient treatment is a matter of debate, as with other agents. We attempted to address this issue using TGF-β scores in patients with renal cancer. We classified the samples according to KEGG-TGF-β scores and predicted the IC50 values of various targeted drugs. We observed that the sensitivities of the three groups to the targeted drugs were different. This suggests that choosing different targeted drugs according to the different patient characteristics can afford better efficacy or appropriately reduce the drug concentration to lessen the side effects of the drug. This provides a guide for a more detailed classification of patients according to their different characteristics, which will result in patients receiving more personalized treatment and ultimately improve the effectiveness of treatment.

Immunotherapy is a popular topic in the field of tumor therapy. TGF-β has systemic immunosuppressive effects and inhibits host immunosurveillance (Yang et al., 2010). Exploring the relationship between TGF-β and immunity will help us gain more insight into TGF-β-targeted therapy or immunotherapy. Our results show that the expression of TGF-β pathway genes is closely related to immune cell infiltration, and type-II-IFN-response and mast cells were most related to TGF-β scores. This discovery may be of great significance for the development of new or improved immunotherapy regimens. Two immune checkpoint inhibitors, PD-1 and CTLA-4, are the main drugs approved by the Food and Drug Administration for the treatment of advanced RCC (Xu et al., 2017; Rini et al., 2019). Recent studies have found that the combination therapy of blocking TGF-β and PD-1/PD-L1 has achieved relatively ideal efficacy (Lind et al., 2019; Wang et al., 2020b). It has also been found that selective inhibition of TGF-β1 produced by GARP-expressing Tregs can overcome resistance to PD-1/PD-L1 blocking in cancer (de Streel et al., 2020).

The TGF-β score was negatively correlated with the expression of CTLA4 and PDCD1, which indicated that the TGF-β low-score group had higher expression of these proteins, and blocking CTLA4 and PDCD1 immune checkpoints may have a better therapeutic effect in this group. However, there was no significant difference in the response of the TGF-β high-score

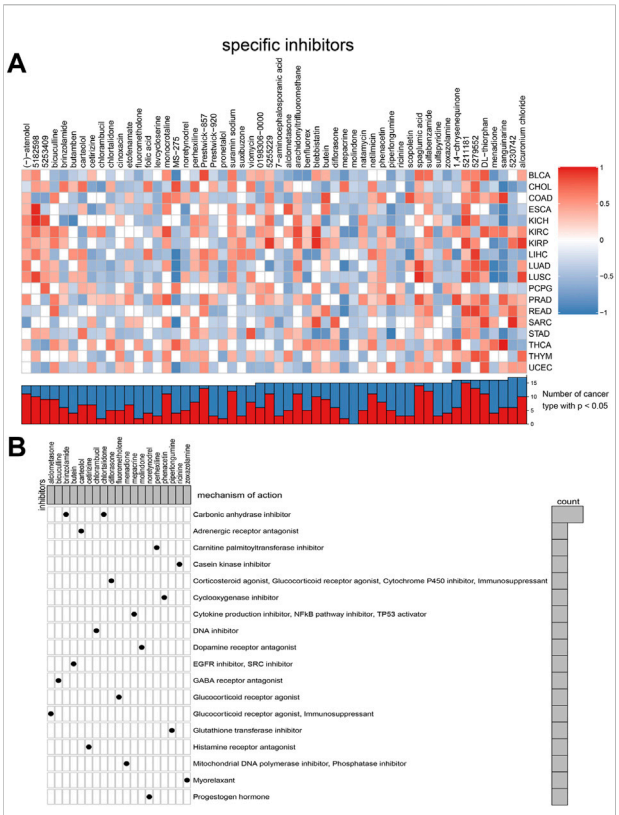


FIGURE 8
Connectivity map (CMAP) analysis identified potential compounds/inhibitors targeting TGF-β. **(A)** Potential inhibitors of TGF-β were predicted using CMAP. In the upper part of Figure 8A, the ordinate represents the type of tumor and the abscissa represents the name of the compound. Red indicates that the compound inhibits the TGF-β signaling pathway, whereas blue indicates that the compound promotes the TGF-β signaling pathway. The color depth represents the intensity of inhibition or promotion. In the lower half of Figure 8A, the red part represents the number of cancer types in which the compound inhibited the TGF-β signaling pathway. **(B)** Exploration of the mechanism of small molecular compounds. KIRC: kidney renal clear cell carcinoma.

and low-score groups to anti-CTLA4 and anti-PDCD1 therapy after Bonferroni correction. Further research is needed on the classification of TGF-β gene expression in RCC to guide immunity and on the close relation of TGF-β expression to immune cell infiltration and the expression of PD-1 and CTLA4.

IFN-γ, the only type-II interferon, is a key cytokine produced by activated T cells and natural killer (NK) and NK-T cells in the tumor microenvironment. IFN-γ signals play an important role in coordinating processes (Ayers et al., 2017) such as anti-cancer immunity, improving tumor immunogenicity, and causing anti-tumor effects through the host immune system (Castro et al., 2018). The main function of PD-1 is to weaken the response of effector T cells and prevent the escape of tumor cells from immune attack (Chen et al., 2019). However, IFN-γ signaling can ultimately induce feedback inhibition, compromising anti-tumor immunity. However, IFN-γ has another role. As part of the feedback loop, IFN-γ signals activate the PD-1 signal axis (Ayers et al., 2017). The main reason why the TGF-β score is positively correlated with IFN and negatively correlated with PD-1 may be that the main effect of IFN is an anti-tumor effect, which may also imply the importance of maintaining a relative balance of components in the tumor microenvironment. Some studies have shown that in metastatic RCC, the OS and PFS rates of patients with high mast cell density are significantly better than those of patients with low mast cell density (Yao et al., 2021). This may explain the positive correlation between mast cells and the TGF-β score, but the role of mast cells in tumors is diverse (Cherdantseva et al., 2017; Yao et al., 2021), and the specific mechanism requires further study.

We used Lasso regression to establish a prognostic model of RCC based on the TGF-β pathway genes. The results showed that the prognosis in the low-risk group was significantly better than that in the high-risk group. Multivariate Cox regression analysis showed that the TGF-β score was an independent risk factor for RCC. These results once again prove the importance of the TGF-β pathway in the prognosis of RCC.

TGF-β pathway genes also have a wide range of SNV and gene expression alterations across multiple cancer types and play a role in a variety of tumor types as prognostic molecules. Consistent with our study, a previous study performed an integration analysis of TGF-β superfamily genetic alterations in 9,125 tumor samples from 33 cancer types, elucidated the salient characteristics of TGF-β-related genes in a large group of different cancer types, and found high-frequency genetic alterations in the TGF-β superfamily across cancer types (Korkut et al., 2018). TGF-β pathway genes are risk factors in most tumors, but in ccRCC, most TGF-β pathway genes are protective factors, which makes the role of TGF-β pathway genes in KIRC different from that in other tumors. Therefore, the different characteristics of TGF-β pathway genes in RCC warrant further investigation.

This study has some limitations. Although the TGF-β score is closely related to immune cell infiltration and immune checkpoint expression, we found that the response of the TGF-β-activated and -inactivated clusters to immune checkpoint inhibitors was not statistically significant, which means that the TGF-β score cannot be used to guide immunotherapy in ccRCC patients. Moreover, our gene set did not include some downstream TGF-β signaling target genes, such as the EMT-related genes *CDH1*, *CDH2*, *SNAI1*, and *VIM*, which makes our investigation of pathway activity imperfect. Because we classified the samples mainly based on the TGF-β score, these genes were not within the TGF-β pathway given by KEGG, so we could not give TGF-β scores for these genes. As such, these genes were excluded.

In summary, compared with normal renal tissues, most genes of the TGF-β pathway are significantly differentially expressed in ccRCC and can serve as risk or protective factors that affect the

prognosis of patients with ccRCC. People with low TGF-β scores have a worse prognosis, and most genes of the TGF-β pathway are involved in the regulation of ccRCC as protective factors. Stratifying patients with RCC according to their TGF-β score is of great significance for evaluating the prognosis of patients and finding new targets.

Data availability statement

All the code used to generate the analysis results is available at the following address https://github.com/ljy11652/ljy11652.git.

Author contributions

GW and XC designed the study and analyzed the data; JL and QW analyzed the data and wrote the manuscript; YX performed the data analysis and interpreted the data. All the authors checked the manuscript.

Funding

This work was supported by the Scientific Research Fund of Liaoning Provincial Education Department (No. LZ2020071), the Dalian Young Stars of Science and Technology Fund Project (No. 2021RQ010), and the Liaoning Province Doctoral Research Startup Fund Program (No. 2021-BS-209).

Acknowledgments

The authors appreciate The Cancer Genome Atlas for providing the open data.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.953322/full#supplementary-material

References

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/JCI91190

Bao, J. M., Dang, Q., Lin, C. J., Lo, U. G., Feldkoren, B., Dang, A., et al. (2020). SPARC is a key mediator of TGF-β-induced renal cancer metastasis. *J. Cell. Physiol.* 236, 1926–1938. doi:10.1002/jcp.29975

Castro, F., Cardoso, A. P., Gonçalves, R. M., Serre, K., and Oliveira, M. J. (2018). Interferon-Gamma at the crossroads of tumor immune surveillance or evasion. *Front. Immunol.* 9, 847. doi:10.3389/fimmu.2018.00847

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools - an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009

Chen, J., Zaidi, S., Rao, S., Chen, J., Phan, L., Farci, P., et al. (2018). Analysis of Genomes and transcriptomes of hepatocellular carcinomas identifies mutations and gene expression changes in the transforming growth factor-β pathway. *GASTROENTEROLOGY* 154 (1), 195–210. doi:10.1053/j.gastro.2017.09.007

Chen, S., Crabill, G. A., Pritchard, T. S., McMiller, T. L., Wei, P., Pardoll, D. M., et al. (2019). Mechanisms regulating PD-L1 expression on tumor and immune cells. *J. Immunother. Cancer* 7 (1), 305. doi:10.1186/s40425-019-0770-2

Cherdantseva, T. M., Bobrov, I. P., Avdalyan, A. M., Klimachev, V. V., Kazartsev, A. V., Kryuchkova, N. G., et al. (2017). Mast cells in renal cancer: Clinical morphological correlations and prognosis. *Bull. Exp. Biol. Med.* 163 (6), 801–804. doi:10.1007/s10517-017-3907-7

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

Coppin, C., Kollmannsberger, C., Le, L., Porzsolt, F., and Wilt, T. J. (2011). Targeted therapy for advanced renal cell cancer (RCC): A cochrane systematic review of published randomised trials. *BJU Int.* 108 (10), 1556–1563. doi:10.1111/j.1464-410X.2011.10629.x

de Streel, G., Bertrand, C., Chalon, N., Liénart, S., Bricard, O., Lecomte, S., et al. (2020). Selective inhibition of TGF-β1 produced by GARP-expressing Tregs overcomes resistance to PD-1/PD-L1 blockade in cancer. *Nat. Commun.* 11 (1), 4545. doi:10.1038/s41467-020-17811-3

Du, G., Yan, X., Chen, Z., Zhang, R., Tuoheti, K., Bai, X., et al. (2020). Identification of transforming growth factor beta induced (TGFBI) as an immune-related prognostic factor in clear cell renal cell carcinoma (ccRCC). *Aging (Albany, NY.)* 12 (9), 8484–8505. doi:10.18632/aging.103153

Escudier, B., Eisen, T., Szczylik, C., Oudard, S., Negrier, S., Chevreau, C., et al. (2007). Sorafenib in advanced clear-cell renal cell carcinoma. *N. Engl. J. Med.* 356, 125–134. doi:10.1056/NEJMoa060655

Gao, S. L., Yin, R., Zhang, L. F., Wang, S. M., Chen, J. S., Wu, X. Y., et al. (2020). The oncogenic role of MUC12 in RCC progression depends on c-Jun/TGF-β signalling. *J. Cell. Mol. Med.* 24 (15), 8789–8802. doi:10.1111/jcmm.15515

Geeleher, P., Cox, N., Huang, R. S., and Barbour, J. D. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLOS ONE* 9 (9), e107468. doi:10.1371/journal.pone.0107468

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

Gong, J., Maia, M. C., Dizman, N., Govindarajan, A., and Pal, S. K. (2016). Metastasis in renal cell carcinoma: Biology and implications for therapy. *Asian J. Urol.* 3 (4), 286–292. doi:10.1016/j.ajur.2016.08.006

Hoshida, Y., Brunet, J. P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2007). Subclass mapping: Identifying common subtypes in independent disease data sets. *PLOS ONE* 2 (11), e1195. doi:10.1371/journal.pone.0001195

Jeh, S. U., Park, J. J., Lee, J. S., Kim, D. C., Do, J., Lee, S. W., et al. (2017). Differential expression of the sirtuin family in renal cell carcinoma: Aspects of carcinogenesis and prognostic significance. *Urol. Oncol.* 35 (12), e9–675. doi:10.1016/j.urolonc.2017.08.016

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Kominsky, S. L., Doucet, M., Brady, K., and Weber, K. L. (2007). TGF-B promotes the establishment of renal cell carcinoma bone metastasis. *J. Bone Min. Res.* 22 (1), 37–44. doi:10.1359/jbmr.061005

Korkut, A., Zaidi, S., Kanchi, R. S., Rao, S., Gough, N. R., Schultz, A., et al. (2018). A pan-cancer analysis reveals high-frequency genetic alterations in mediators of signaling by the TGF-β superfamily. *Cell. Syst.* 7 (4), 422–437. doi:10.1016/j.cels.2018.08.010

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *SCIENCE* 313 (5795), 1929–1935. doi:10.1126/science.1132939

Leibovich, B. C., Lohse, C. M., Crispen, P. L., Boorjian, S. A., Thompson, R. H., Blute, M. L., et al. (2010). Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. *J. Urol.* 183 (4), 1309–1315. doi:10.1016/j.juro.2009.12.035

Lind, H., Gameiro, S. R., Jochems, C., Donahue, R. N., Strauss, J., Gulley, J. L., et al. (2019). Dual targeting of TGF-β and PD-L1 via a bifunctional anti-PD-L1/TGF-βRII agent: Status of preclinical and clinical advances. *J. Immunother. Cancer* 8 (1), e000433. doi:10.1136/jitc-2019-000433

Liu, Y., and Shang, D. (2020). Transforming growth factor-β1 enhances proliferative and metastatic potential by up-regulating lymphoid enhancer-binding factor 1/integrin αMβ2 in human renal cell carcinoma. *Mol. Cell. Biochem.* 465 (1-2), 165–174. doi:10.1007/s11010-019-03676-8

Lu, X., Jiang, L., Zhang, L., Zhu, Y., Hu, W., Wang, J., et al. (2019). Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *NEOPLASIA* 21 (6), 591–601. doi:10.1016/j.neo.2019.04.003

Mallikarjuna, P., Raviprakash, T. S., Aripaka, K., Ljungberg, B., and Landstrom, M. (2019). Interactions between TGF-beta type I receptor and hypoxia-inducible factor-alpha mediates a synergistic crosstalk leading to poor prognosis for patients with clear cell renal cell carcinoma. *Cell. CYCLE* 18 (17), 2141–2156. doi:10.1080/15384101.2019.1642069

Motzer, R. J., Hutson, T. E., Cella, D., Reeves, J., Hawkins, R., Guo, J., et al. (2013). Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N. Engl. J. Med.* 369 (8), 722–731. doi:10.1056/NEJMoa1303989

Motzer, R. J., Hutson, T. E., McCann, L., Deen, K., and Choueiri, T. K. (2014). Overall survival in renal-cell carcinoma with pazopanib versus sunitinib. *N. Engl. J. Med.* 370 (18), 1769–1770. doi:10.1056/NEJMc1400731

Motzer, R. J., Thomas, E., Hutson, D. O., Tomczak, P., Michaelson, M. D., Bukowski, R. M., et al. (2007). Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N. Engl. J. Med.* 356, 115–124. doi:10.1056/NEJMoa065044

Rini, B. I., Battle, D., Figlin, R. A., George, D. J., Hammers, H., Hutson, T., et al. (2019). The society for immunotherapy of cancer consensus statement on immunotherapy for the treatment of advanced renal cell carcinoma (RCC). *J. Immunother. Cancer* 7 (1), 354. doi:10.1186/s40425-019-0813-8

Rini, B. I., Escudier, B., Tomczak, P., Kaprin, A., Szczylik, C., Hutson, T. E., et al. (2011). Comparative eff ectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): A randomised phase 3 trial. *Lancet* 378, 1931–1939. doi:10.1016/S0140-6736(11)61613-9

Siegel, R. L., Miller, K. D., and Cancer Statistics, J. A. (2018). Cancer statistics, 2018. *Ca. Cancer J. Clin.* 68 (1), 7–30. doi:10.3322/caac.21442

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

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *Ca. Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., et al. (2017). A next generation connectivity map: L1000 platform and the first 1, 000, 000 profiles. *Cell.* 171 (6), 1437–1452. doi:10.1016/j.cell.2017.10.049

Taguchi, L., Miyakuni, K., Morishita, Y., Morikawa, T., Fukayama, M., Miyazono, K., et al. (2019). c-Ski accelerates renal cancer progression by attenuating transforming growth factor β signaling. *Cancer Sci.* 110, 2063–2074. doi:10.1111/cas.14018

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* 19 (1A), A68–A77. doi:10.5114/wo.2014.47136

Wang, P., Chen, W., Ma, T., Lin, Z., Liu, C., Liu, Y., et al. (2020). lncRNA lnc-TSI inhibits metastasis of clear cell renal cell carcinoma by suppressing TGF-β-induced epithelial-mesenchymal transition. *Mol. Ther. Nucleic Acids* 22, 1–16. doi:10.1016/j.omtn.2020.08.003

Wang, Y., Gao, Z., Du, X., Chen, S., Zhang, W., Wang, J., et al. (2020). Co-inhibition of the TGF-β pathway and the PD-L1 checkpoint by pH-responsive clustered nanoparticles for pancreatic cancer microenvironment regulation and anti-tumor immunotherapy. *Biomater. Sci.* 8 (18), 5121–5132. doi:10.1039/d0bm00916d

Xu, J. X., Maher, V. E., Zhang, L., Tang, S., Sridhara, R., Ibrahim, A., et al. (2017). FDA approval summary: Nivolumab in advanced renal cell carcinoma after anti-angiogenic therapy and exploratory predictive biomarker analysis. *Oncologist* 22 (3), 311–317. doi:10.1634/theoncologist.2016-0476

Yang, L., Pang, Y., and Moses, H. L. (2010). TGF-B and immune cells: An important regulatory axis in the tumor microenvironment and progression. *Trends Immunol.* 31 (6), 220–227. doi:10.1016/j.it.2010.04.002

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2012). Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41 (D1), D955–D961. doi:10.1093/nar/gks1111

Yao, J., Xi, W., Chen, X., Xiong, Y., Zhu, Y., Wang, H., et al. (2021). Mast cell density in metastatic renal cell carcinoma: Association with prognosis and tumour-infiltrating lymphocytes. *Scand. J. Immunol.* 93 (4), e13006. doi:10.1111/sji.13006

Zhang, L., Zhao, Y., Dai, Y., Cheng, J., Gong, Z., Feng, Y., et al. (2018). Immune landscape of colorectal cancer tumor microenvironment from different primary tumor location. *Front. Immunol.* 9, 1578. doi:10.3389/fimmu.2018.01578

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 30 September 2022
DOI 10.3389/fgene.2022.970316

OPEN ACCESS

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

REVIEWED BY
Md. Nazim Uddin,
China Pharmaceutical University, China
Geoffroy Andrieux,
Universität Freiburg, Germany

*CORRESPONDENCE
Jun Shi,
shijun@ihcams.ac.cn
Yuan Zhou,
yuanzhou@ihcams.ac.cn

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

RECEIVED 15 June 2022
ACCEPTED 01 September 2022
PUBLISHED 30 September 2022

CITATION
Wang W, Liang Q, Zhao J, Pan H, Gao Z, Fang L, Zhou Y and Shi J (2022), Low expression of the metabolism-related gene SLC25A21 predicts unfavourable prognosis in patients with acute myeloid leukaemia.
Front. Genet. 13:970316.
doi: 10.3389/fgene.2022.970316

Low expression of the metabolism-related gene SLC25A21 predicts unfavourable prognosis in patients with acute myeloid leukaemia

Wenjun Wang, Qian Liang, Jingyu Zhao, Hong Pan, Zhen Gao, Liwei Fang, Yuan Zhou* and Jun Shi *

Regenerative Medicine Clinic, State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, China

Acute myeloid leukaemia (AML) is a heterogeneous disease associated with poor outcomes. To identify AML-specific genes with prognostic value, we analysed transcriptome and clinical information from The Cancer Genome Atlas (TCGA) database, Gene Expression Omnibus (GEO) datasets, and Genotype-Tissue Expression (GTEx) project. The metabolism-related gene, *SLC25A21* was found to be significantly downregulated in AML, and was associated with high white blood cell (WBC) counts, high pretrial blood (PB) and bone marrow (BM) blast abundance, *FLT3* mutation, *NPM1* mutation, and death events (all p value <0.05). We validated the expression of *SLC25A21* in our clinical cohort, and found that *SLC25A21* was downregulated in AML. Moreover, we identified low expression of *SLC25A21* as an independent prognostic factor by univariate Cox regression (hazard ratio [HR]: 0.550; 95% Confidence interval [CI]: 0.358–0.845; p value = 0.006) and multivariate Cox regression analysis (HR: 0.341; 95% CI: 0.209–0.557; p value <0.05). A survival prediction nomogram was established with a C-index of 0.735, which indicated reliable prognostic prediction. Subsequently, based on the median *SLC25A21* expression level, patients in the TCGA-LAML cohort were divided into low- and high-expression groups. Gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs highlighted growth factor binding, extracellular structure organization, cytokine–cytokine receptor interaction, etc. The results of gene set enrichment analysis (GSEA) indicated that the epithelial-mesenchymal transition, KRAS signalling, oxidative phosphorylation, and reactive oxygen species pathways were enriched. Through gene coexpression and protein–protein interaction (PPI) network analysis, we identified two hub genes, *EGFR* and *COL1A2*, which were linked to worse clinical outcomes. Furthermore, we found that lower *SLC25A21* expression was closely associated with a significant reduction in the levels of infiltrating immune cells, which might be associated with immune escape of AML cells. A similar trend was observed for the expression of checkpoint genes (*CTLA4*, *LAG3*, *TIGIT*, and *HAVCR2*). Finally, drug sensitivity testing suggested that the low-expression *SLC25A21* group is sensitive to

doxorubicin, mitomycin C, linifanib but resistant to JQ1, belinostat, and dasatinib. Hence, our study demonstrated that a low expression level of *SLC25A21* predicts an unfavourable prognosis in patients with AML.

KEYWORDS

SLC25A21, prognosis, bioinformatics, GEO, TCGA, immune checkpoint, drug sensitivity, acute myeloid leukaemia (AML)

Introduction

Acute myeloid leukaemia (AML) is a genetically and clinically heterogeneous disease characterized by clonal expansion, differentiation arrest, and evasion of apoptosis. Despite recent advances in chemotherapy, immunotherapy, and bone marrow transplantation, large numbers of AML patients still have a dismal prognosis, with a 5-years survival rate of only approximately 20% (Chen et al., 2019). The development of personalized biomarker-targeted therapies in AML has improved the efficacy of systemic therapies and prolonged patient survival to some extent. However, the lack of biomarkers hinders further improvements in accurate diagnosis and prediction of efficacy. Thus, it is extremely important to discover novel diagnostic and prognostic biomarkers for targeted therapy in AML.

In this research, various comprehensive bioinformatics and statistical methods were used to explore independent prognostic factors in AML. Differentially expressed gene (DEG) analysis, Kaplan-Meier analysis and Cox regression analysis helped us screen out Solute Carrier Family 25 Member 21 (*SLC25A21*) as an AML-specific prognostic marker. *SLC25A21*, also called *ODC*, is a metabolism-related gene located on chromosome 14q13.3, and it encodes a protein known as mitochondrial 2-oxodicarboxylate carrier (Fiermonte et al., 2001). The SLC25A21 protein not only facilitates the counterexchange of the oxodicarboxylates 2-oxoadipate and 2-oxoglutarate but also plays an essential role in the metabolism of several amino acids (Fiermonte et al., 1998; Kunji et al., 2020). Germline *SLC25A21* deficiency in humans causes the depletion of mitochondrial DNA and spinal muscular atrophy-like disease (Fiermonte et al., 1998; Boczonadi et al., 2018). Metabolic reprogramming is a hallmark of cancer, and targeting metabolic factors is an emerging therapeutic modality (Chen et al., 2020a; Bosc et al., 2020; Forte et al., 2020; Pei et al., 2020). Interestingly, a recent study showed that *SLC25A21* is a key tumor suppressor gene in bladder cancer (Wang et al., 2021). However, the potential role of *SLC25A21* in AML and whether it could serve as a novel target for metabolic therapy remain completely unknown.

Hence, we used GO and KEGG analyses, GSEA, PPI network construction, immune infiltration and immune checkpoint correlations, and drug sensitivity analysis to explore the underlying molecular pathological mechanisms of *SLC25A21* in AML. Based on the above results, we confirmed the prognostic value of *SLC25A21* and identified it as a potential therapeutic target for AML.

Material and methods

Data source

We included 804 samples from three independent cohorts in this study: the TCGA LAML cohort (RNA-seq, $n = 132$) (Cancer Genome Atlas Research et al., 2013), the GSE13159 microarray dataset ($n = 573$) (Haferlach et al., 2010) and the GSE12417 dataset (RNA-seq, $n = 163$) (Metzeler et al., 2008). The matrix of mRNA expression in normal samples ($n = 70$) was extracted from the GTEx project (Consortium, 2020). The RNA-seq and clinical information from the TCGA LAML and GTEx datasets were acquired using the UCSC XENA browser (https://xenabrowser.net/datapages/) (Vivian et al., 2017; Consortium, 2020; Goldman et al., 2020). The microarray dataset GSE13159 and RNA-seq dataset GSE12417 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). To maintain the comparability of data from different databases, TPM values from RNA-Seq were determined for intrasample comparison after log2 transformation. In our study, specimens with no survival data were excluded.

Gene expression profiling

To analyse the gene expression profiles of AML, 705 bone samples from the GSE13159 and LAML datasets were used. The GSE13159 dataset was collected from the Microarray Innovations in Leukaemia Study. The DEGs were predicted using the limma package in R, with an adjusted p value <0.05 and $|\log2FC| \geq 0.15$ (Ritchie et al., 2015). A list of 14 common differentially expressed AML-specific genes was obtained from the above databases by using the Venn online tool (https://bioinfogp.cnb.csic.es/tools/venny/).

Identification of overall survival-related genes

The LAML cohort was used to investigate the potential prognostic significance of the selected genes in AML patients.

OS-related genes with a p value <0.05 were selected using univariate Cox hazard regression analysis for further research. The external cohort GSE12417 ($n = 163$) was used to validate our results (Metzeler et al., 2008).

Human subjects and quantitative real-time PCR

Bone marrow samples were collected from 20 patients with AML diagnosis according to the 2016 WHO criteria. We also collected 10 bone marrow samples from healthy donors. The individuals in both cohorts were aged 18–70 years. All patients signed the informed consent form, and the study protocol was approved by the Ethics Committee of our hospital. The patient information collected is listed in Supplementary Table S3. Isolation of mononuclear cells was performed using standard Ficoll standard procedure. Total RNA was isolated with TRIzol reagent (Life Technologies) and then reverse transcribed to cDNA using the ImProm-II™ Reverse Transcription Kit (Promega, Madison, United States). For gene expression analysis, cDNA samples were mixed with SYBR reagent using a 7900 real-time PCR system (Applied Biosystems), and the data were normalized to GAPDH. The primer sequences are available in Supplementary Table S4.

Differentially expressed genes analysis

Based on the median *SLC25A21* expression level, patients in the LAML cohort were divided into two groups of low and high expression. A list of DEGs was obtained using DESeq2 with an adjusted p value < 0.05 and $|\log2FC| \geq 1$ (Love et al., 2014).

Gene ontology and kyoto encyclopedia of genes and genomes enrichment analysis of differentially expressed genes

By using the R package clusterProfiler, we carried out functional annotation analysis to investigate the underlying functions of DEGs in AML (Yu et al., 2012; Gene Ontology, 2021; Kanehisa et al., 2021). A Benjamin–Hochberg adjusted p value <0.05 was interpreted as statistically significant. Heatmap of clustered DEGs was generated using ClustVis software (Metsalu and Vilo, 2015).

Gene set enrichment analysis

GSEA was conducted by using the clusterProfiler package in R and hallmark signatures (h.all.v7.2. symbols.gmt) from MsigDB (Subramanian et al., 2005; Yu et al., 2012). Results were considered significant when |NES | >1, normalized p value <0.05[23].

Comprehensive protein–protein interaction analysis

Associations between *SLC25A21* and the expression of other genes were assessed using the LinkedOmics database (http://www.linkedomics.org/login.php). We derived the PPI network from the STRING database (https://string-db.org/) to estimate the interactional correlations of the DEGs (Szklarczyk et al., 2019). A confidence score >0.4 was considered significant. Hub proteins and key nodes in the constructed PPI network were identified using the Cytoscape plugin CytoHubba (Shannon et al., 2003). We investigated the association of *SLC25A21* expression with hub genes through correlation heatmaps by using the R package ggplot2.

Immune infiltrate analysis

By using Single Sample GSEA (ssGSEA) in the R package gsva, Spearman correlation coefficients were computed between the expression level of *SLC25A21* and ssGSEA-based immune cell infiltration levels (Bindea et al., 2013; Hanzelmann et al., 2013). The involved immune cells were of 22 immune cell subtypes, including B cells, monocytes, macrophages, neutrophils, NK cells, DC cells and all subtypes of T-cells. The relationships between *SLC25A21* expression and the expression of immune checkpoint molecules, including *PDCD1*, *CD274*, *CTLA-4*, *LAG-3*, *TIGIT*, and *HAVCR2*, were identified through correlation heatmaps by using the R package ggplot2.

Drug sensitivity prediction

The half-maximal inhibitory concentration (IC50), calculated using the pRRophetic package in R (Geeleher et al., 2014; Ding et al., 2021), was used for drug sensitivity prediction.

Survival and statistical analysis

All statistical analyses were performed with R, version 4.1.3 (https://www.r-project.org/). The Wilcoxon rank-sum test and Kruskal–Wallis test were used to detect differences among continuous variables. The correlation of clinical features with low and high *SLC25A21* expression were analysed with Pearson's correlation $\chi2$ test. For survival

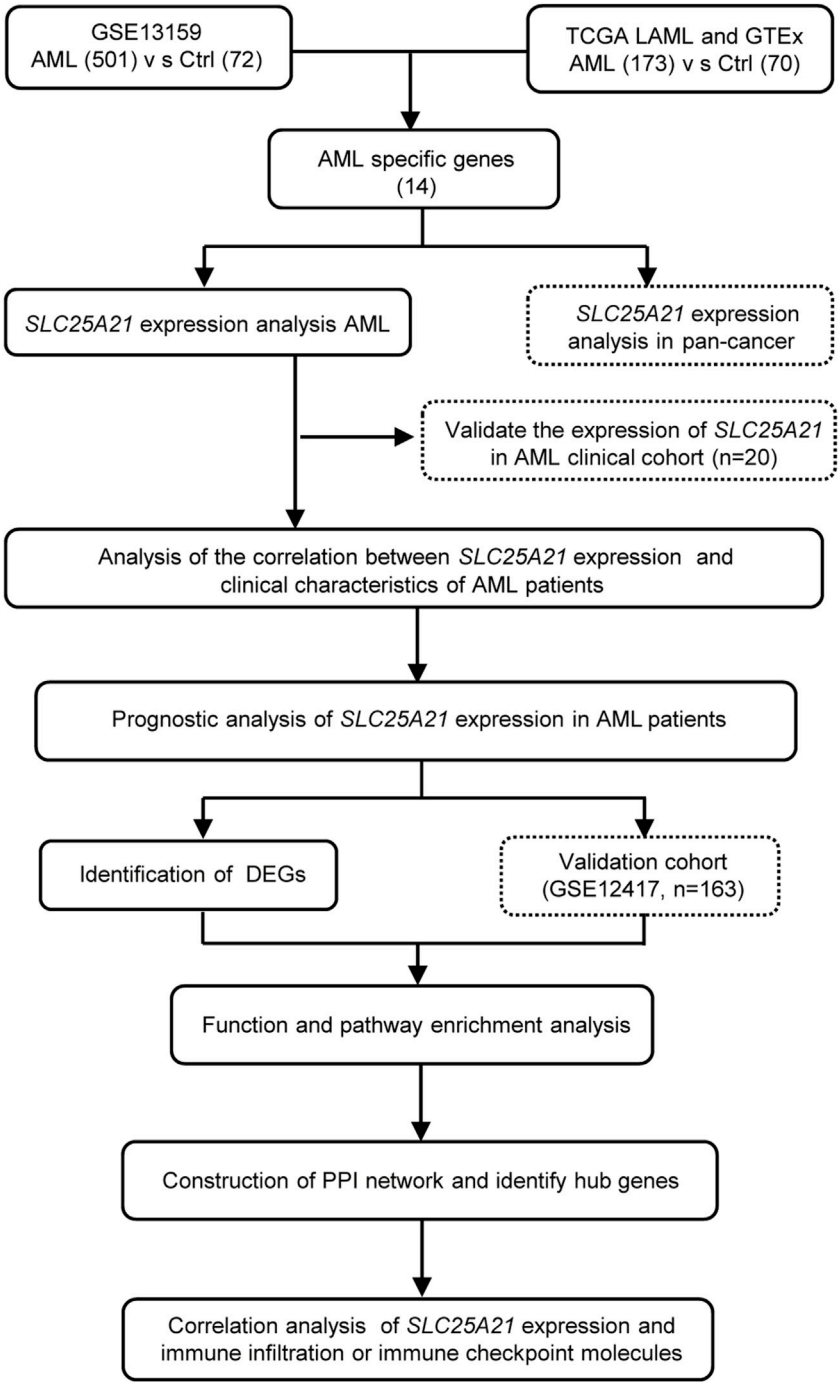


FIGURE 1
Flow chart of the study.

analysis, Cox proportional hazards analysis was conducted by using "survival" and "survminer" in R. Variables significant in Cox univariate analysis were selected for multivariate analysis. The Kaplan-Meier method was used for univariate analyses of OS. Receiver operating characteristic (ROC) curves and AUC values were generated by pROC in R to assess the diagnostic efficacy of *SLC25A21* for AML. All tests were two-sided, and a p value < 0.05 was considered to indicate statistical significance. A flow chart of the analyses is presented in Figure 1.

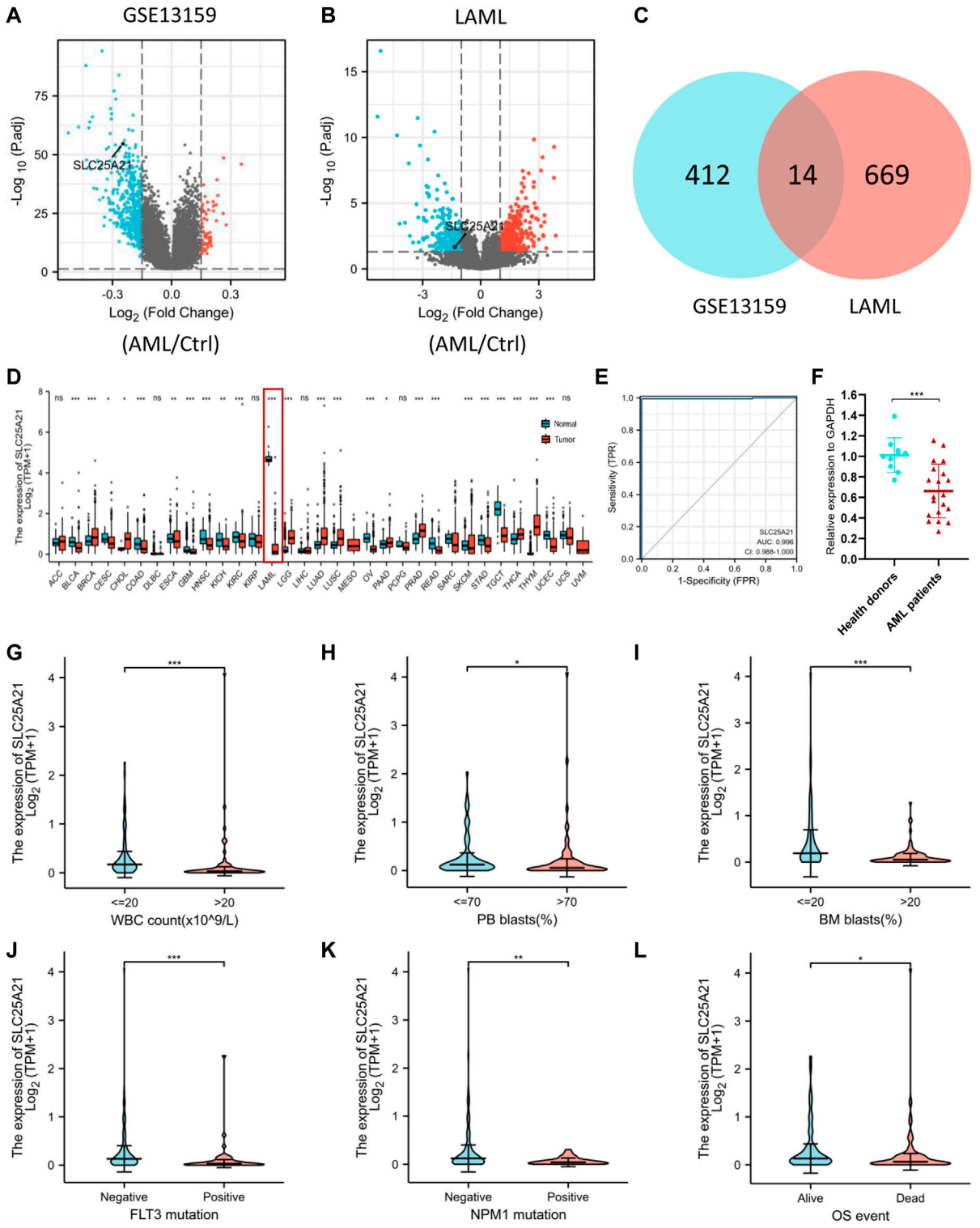


FIGURE 2
Identifying AML-specific genes and the association of SLC25A21 expression with clinical characteristics. **(A)** Volcano plot displaying DEGs between AML and control bone marrow samples in the GSE13159 dataset. Each point represents the average value of one transcript. **(B)** Volcano plot of DEGs between AML and normal samples in the TCGA-LAML and GTEx datasets. **(C)** Venn diagram of differential gene expression. Selected genes for further analysis based on the intersections of DEGs. **(D)** The level of SLC25A21 expression in different tumor from TCGA and GTEx database. **(E)** Receiver operating characteristic (ROC) analysis of SLC25A21 in AML. The analysis was performed with the TCGA-LAML and GTEx dataset. **(F)**
(*Continued*)

FIGURE 2 (Continued)
Differential expression of SLC25A21 between AML patients and healthy donors by qPCR analyses. The results were expressed as the fold change of AML patients relative to healthy donors. Clinical characteristics included **(G)** WBC count, **(H)** PB blasts abundance, **(I)** BM blasts abundance, **(J)** FLT3 mutation, **(K)** NPM1 mutation, **(L)** OS evens. Data are presented as the mean ± SD, and represent triplicate wells from one of two independent experiments. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Analysis between two groups of unpaired samples: Wilcoxon rank-sum test, analysis among multiple groups: Kruskal–Wallis test (ns $p \geq 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Results

Acute myeloid leukaemia-specific genes identified with screening datasets

We analysed the TCGA LAML database and the GEO dataset GSE13159 to understand the potential molecular changes in AML. For GSE13159, gene expression analysis was performed on bone marrow samples from 501 AML to 72 control samples (nonleukaemia and healthy donors). We identified 61 upregulated and 365 downregulated DEGs in the AML group ($|\log2FC| \geq 0.15$, adjusted p value < 0.05) by using the limma package in R. The volcano plots are shown in Figure 2A. By screening the TCGA LAML datasets ($|\log2FC| \geq 1$, adjusted p value < 0.05) with DESeq2, a total of 683 differentially expressed genes were obtained, which are shown in Figure 2B. Venn diagram software was used to obtain a common DEG list. A list of 14 intersectional genes was extracted, of which 11 were upregulated and 3 were downregulated in GSE13159 and TCGA LAML (Figure 2C), including *IL1R2*, *MMP8*, *FGF13*, *SLC25A21*, etc. In pan-cancer analysis, we determined the expression profiles of these genes in normal and malignant samples. Remarkably, *SLC25A21* was downregulated in multiple malignancies, especially in AML (Figure 2D). Furthermore, we performed receiver operating characteristic (ROC) curve analyses, and the area under the curve (AUC) was used to evaluate the discriminatory capacity. The calculated AUC value was 0.996 (95% confidence interval, CI = 0.988–1.0, Figure 2E), which means that *SLC25A21 has* excellent discrimination power to distinguish AML patients from normal controls. Finally, we validated the expression of *SLC25A21* in AML patient bone marrow samples collected in our centre. We compared the mRNA expression level of *SLC25A21* between AML patients ($n = 20$) and healthy donors ($n = 10$) by qPCR. *SLC25A21* was significantly downregulated in AML samples, with a p value of 0.0007 (Figure 2F). Therefore, *SLC25A21* could be a specific factor to distinguish AML from normal samples.

Low levels of SLC25A21 are associated with adverse clinical features in acute myeloid leukaemia

To investigate the clinical significance of *SLC25A21*, we analysed the TCGA LAML cohort, which includes 132 AML patients with clinical information. As shown in Figures 2F–K, low *SLC25A21* expression was associated with higher WBC counts (p value < 0.001, Figure 2G), higher PB blast abundance (p value < 0.05, Figure 2H), higher BM blast abundance (p value < 0.001, Figure 2I), FLT3 mutation (p value < 0.001, Figure 2J), NPM1 mutation (p value < 0.01, Figure 2K), and death evens (p value < 0.001, Figure 2L); however, no association was found with cytogenetic risk or French–American–British (FAB) classifications (Supplementary Figures 1A–B). In addition, similar trends were observed when patients were grouped by low or high *SLC25A21* expression; more details are shown in Table 1.

Low expression of *SLC25A21* predicts unfavourable prognosis in patients with acute myeloid leukaemia

We further investigated the prognostic value of *SLC25A21* in AML. First, patients in the low *SLC25A21* expression group had shorter OS than those in the high *SLC25A21* expression group (p value = 0.006, Figure 3A), which indicated that a low *SLC25A21* expression level was associated with an unfavourable prognosis in patients with AML. Furthermore, we validated our results in an independent external validation cohort GSE12417 (p value = 0.027, Figure 3B).

In addition, univariate and multivariate logistic regression analyses were performed to determine whether low expression of *SLC25A21* was an independent prognostic factor for AML. Univariate Cox regression analysis showed that low levels of *SLC25A21* expression were associated with poor OS (hazard ratio, [HR]: 0.55; 95% confidence interval [CI]: 0.358–0.845; p value = 0.006). Meanwhile, increasing age and unfavourable cytogenetics were also risk factors associated with poor outcomes. Then, all variables significant in univariate Cox regression analysis (p value <0.05) were included in multivariate Cox regression analysis. Subsequently, age, unfavourable cytogenetics and low levels of *SLC25A21* expression (HR: 1.733; 95% CI: 1.079–2.781; p value = 0.023) were identified as independent prognostic factors for OS. The forest plots present the Cox regression results in Figures 3C,D (more details are provided in Supplementary Table S1).

Moreover, a nomogram including the prediction model was established based on multivariable logistic regression

TABLE 1 Clinical characteristics of AML patients with differential SLC25A21 expression levels.

Characteristics	Total(N)	HR (95%CI) univariate analysis	*p* value univariate analysis	HR (95%CI) multivariate analysis	*p* value multivariate analysis
Gender	140				
Female	63	References			
Male	77	1.030 (0.674–1.572)	0.892		
Age	140				
≤60	79	References			
>60	61	3.333 (2.164–5.134)	<0.001	3.903 (2.425–6.281)	<0.001
WBC count (x109/L)	139				
≤20	75	References			
>20	64	1.161 (0.760–1.772)	0.490		
PB blasts (%)	140				
≤70	66	References			
>70	74	1.230 (0.806–1.878)	0.338		
BM blasts (%)	140				
≤20	59	References			
>20	81	1.165 (0.758–1.790)	0.486		
Cytogenetic risk	138				
Favorable	31	References			
Intermediate	76	2.957 (1.498–5.836)	0.002	2.268 (1.135–4.533)	0.020
Poor	31	4.157 (1.944–8.893)	<0.001	3.540 (1.607–7.800)	0.002
FLT3 mutation	136				
Negative	97	References			
Positive	39	1.271 (0.801–2.016)	0.309		
NPM1 mutation	139				
Negative	106	References			
Positive	33	1.137 (0.706–1.832)	0.596		
RAS mutation	139				
Negative	131	References			
Positive	8	0.643 (0.235–1.760)	0.390		
IDH1 R132 mutation	138				
Negative	126	References			
Positive	12	0.588 (0.238–1.452)	0.249		
IDH1 R140 mutation	138				
Negative	127	References			
Positive	11	1.131 (0.565–2.264)	0.727		
IDH1 R172 mutation	138				
Negative	136	References			
Positive	2	0.610 (0.085–4.385)	0.623		
SLC25A21	140				
Low	66	References			
High	74	0.550 (0.358–0.845)	0.006	0.341 (0.209–0.557)	<0.001

analysis. The established nomogram was well calibrated and had good discriminative power, with a concordance index (C-index) of 0.735 for OS prediction (Figure 3E). Furthermore, we utilized calibration curves and decision curve analysis (DCA) to report the clinical net benefit of our model. The calibration curve at 1, 3, or 5 years still showed high consistency between the predicted survival probability and actual OS proportions (Figure 3F). In addition, the decision curve analysis for the individualized prediction nomogram is presented in Figures 3G–I. In

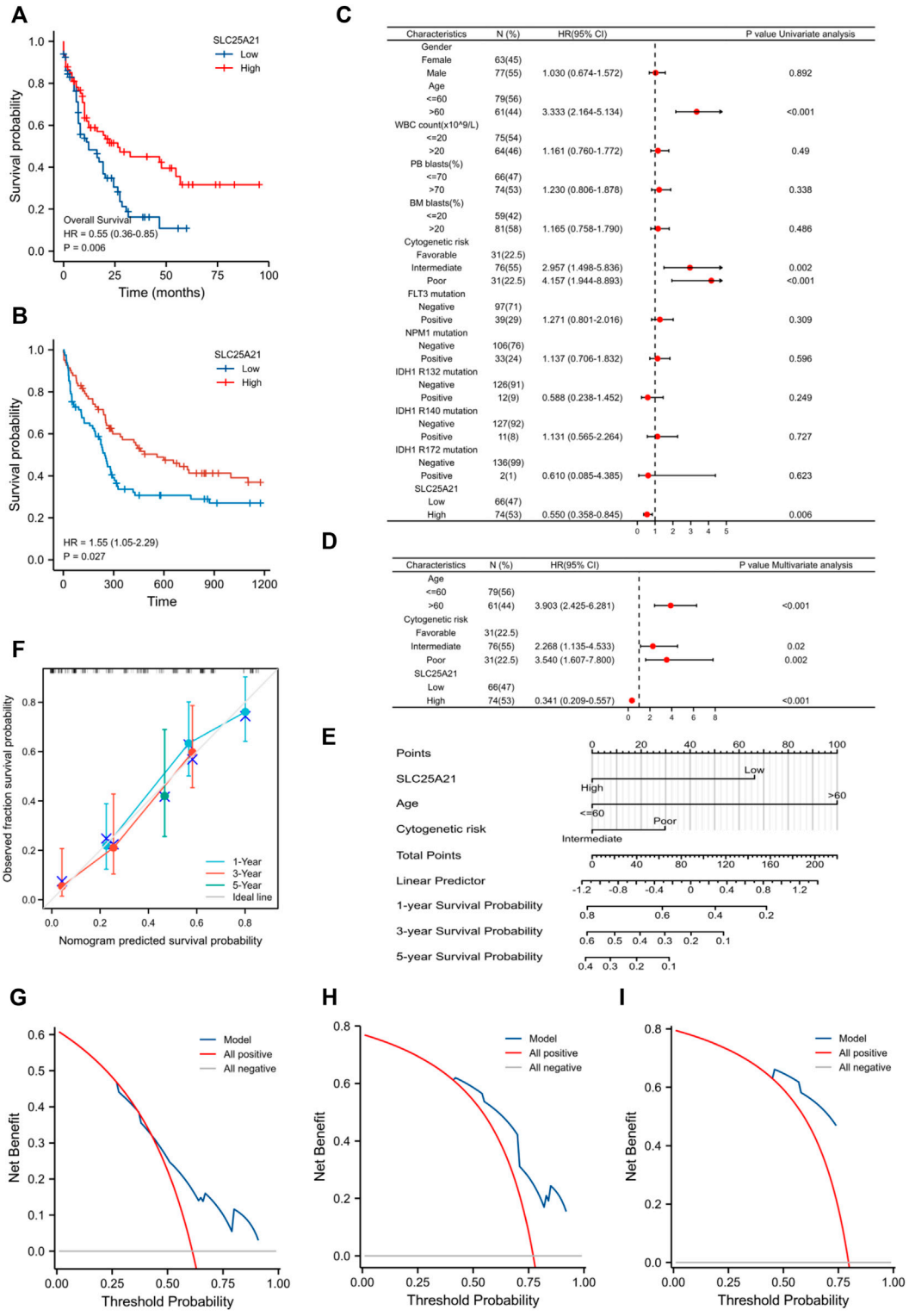


FIGURE 3
The prognostic value of SLC25A21 in AML. **(A)** Kaplan-Meier curve analysis of overall survival (OS) between the high- and low-SLC25A21 expression groups in the TCGA-LAML dataset. **(B)** OS analysis of SLC25A21 in the independent validation cohort GSE12417. **(C)** Univariate analyses of OS shown as a by forest plot. **(D)** Multivariate analyses of OS shown as a forest plot. **(E)** A nomogram integrating SLC25A21 and other prognostic factors for AML (mut: mutation, wt, wild type; Int, Intermediate; Fav, Favourable). **(F)** The calibration curve of the nomogram. The DCA curves of the nomogram at 1 year **(G)**, 3 years **(H)**, and 5 years **(I)**.

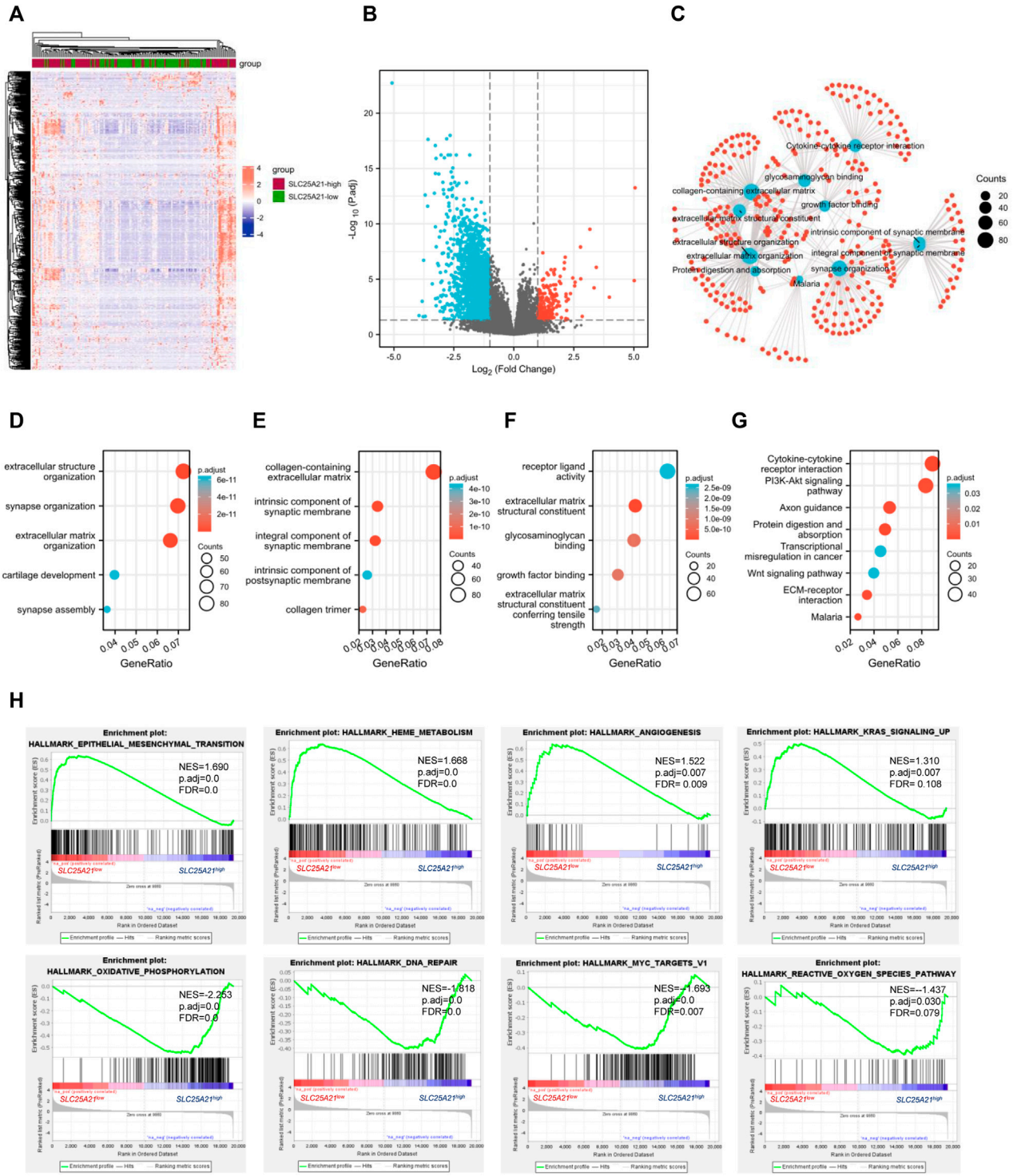


FIGURE 4
DEGs and functional enrichment of the high- and low- SLC25A21 expression groups in AML. **(A)** Heatmap of SLC25A21-related DEGs. **(B)** Volcano plot of SLC25A21-related DEGs. **(B,C)** Interactive analyses of GO and KEGG, including biological processes **(D)**, cellular components **(E)**, and molecular functions **(F)**, and KEGG pathways of SLC25A21-related DEGs. **(G)** Each red plot in the graph represents a specific gene included in the gene-set. Each blue plot represents the enriched gene-sets. The size of the blue plot represents the number of gene read counts in the gene-sets. **(H)** GSEA of SLC25A21-related DEGs.

TABLE 2 Seventeen items from gene set enrichment analysis. (A) Gene sets enriched in phenotype SLC25A21 low. (B) Gene sets enriched in phenotypeSLC25A21 high.

	GS	SIZE	ES	NES	NOM p-value	FDR q-value	FWER p-value	RANK AT MAX	LEADING EDGE
1	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	200	0.64	1.69	0	0	0	2921	tags = 45%, list = 15%, signal = 52%
2	HALLMARK_HEME_METABOLISM	197	0.64	1.67	0	0	0	3583	tags = 38%, list = 18%, signal = 46%
3	HALLMARK_ANGIOGENESIS	36	0.64	1.52	0.007	0.01	0.037	2740	tags = 53%, list = 14%, signal = 61%
4	HALLMARK_HEDGEHOG_SIGNALING	36	0.62	1.46	0.02	0.026	0.13	3664	tags = 50%, list = 19%, signal = 62%
5	HALLMARK_ESTROGEN_RESPONSE_EARLY	199	0.55	1.45	0	0.025	0.148	3940	tags = 38%, list = 20%, signal = 47%
6	HALLMARK_APICAL_SURFACE	44	0.6	1.44	0.009	0.023	0.172	3050	tags = 34%, list = 16%, signal = 40%
7	HALLMARK_MYOGENESIS	199	0.52	1.37	0.002	0.056	0.404	2994	tags = 30%, list = 15%, signal = 35%
8	HALLMARK_UV_RESPONSE_DN	144	0.51	1.32	0.012	0.099	0.651	4417	tags = 35%, list = 23%, signal = 45%
9	HALLMARK_KRAS_SIGNALING_UP	199	0.5	1.31	0.007	0.108	0.716	3932	tags = 36%, list = 20%, signal = 45%
11	HALLMARK_SPERMATOGENESIS	134	0.49	1.27	0.037	0.145	0.874	4396	tags = 34%, list = 23%, signal = 43%
12	HALLMARK_ESTROGEN_RESPONSE_LATE	198	0.48	1.26	0.025	0.16	0.917	4280	tags = 36%, list = 22%, signal = 46%
13	HALLMARK_KRAS_SIGNALING_DN	199	0.48	1.25	0.026	0.161	0.934	3498	tags = 33%, list = 18%, signal = 39%

	GS	SIZE	ES	NES	NOM p-value	FDR q-value	FWER p-value	RANK AT MAX	LEADING EDGE
1	HALLMARK_DNA_REPAIR	148	−0.4018086	−1.818951	0	0	0	7204	tags = 64%, list = 37%, signal = 101%
2	HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	49	−0.3934737	−1.4372096	0.030303031	0.07993332	0.025	4629	tags = 49%, list = 24%, signal = 64%
3	HALLMARK_CHOLESTEROL_HOMEOSTASIS	74	−0.35663486	−1.3603169	0.03846154	0.11938669	0.047	3913	tags = 43%, list = 20%, signal = 54%
4	HALLMARK_INTERFERON_GAMMA_RESPONSE	200	−0.25014699	−1.24814	0	0.19125329	0.092	3746	tags = 37%, list = 19%, signal = 45%
5	HALLMARK_FATTY_ACID_METABOLISM	156	−0.2812873	−1.2210402	0	0.19619812	0.109	2894	tags = 26%, list = 15%, signal = 30%

summary, the nomogram model we established had good predictive accuracy for AML patient survival.

Biological function enrichment of the *SLC25A21* gene in acute myeloid leukaemia

Next, we aimed to further investigate the underlying mechanisms and functional pathways of *SLC25A21* in AML. We identified DEGs between the low- and high- *SLC25A21* expression groups. The final list of DEGs included 1,270 genes, with 128 genes upregulated and 1,142 genes downregulated ($|log2FC| \geq 1$, adjusted p value <0.05). The heatmap and the volcano map are shown in Figures 4A,B.

To elucidate the potential biological function of *SLC25A21* in AML, we performed enrichment analyses. The top 15 GO enrichment items (Figures 4D–F) and top 5 KEGG pathways are shown in Figure 4G. The main enriched GO terms of the DEGs were extracellular structure organization, synapse organization, extracellular matrix organization, collagen-containing extracellular matrix, integral component of synaptic membrane, receptor ligand activity, extracellular matrix structural constituent, glycosaminoglycan binding, growth factor binding, etc.

We found that the enriched pathways included cytokine–cytokine receptor interaction, PI3K-Akt signalling pathway, focal adhesion, proteoglycans in cancer, transcriptional misregulation in cancer, Wnt signalling pathway, and TGF-beta signalling pathway (Supplementary Table S2). Furthermore, interaction analysis was carried out with the results of GO and KEGG

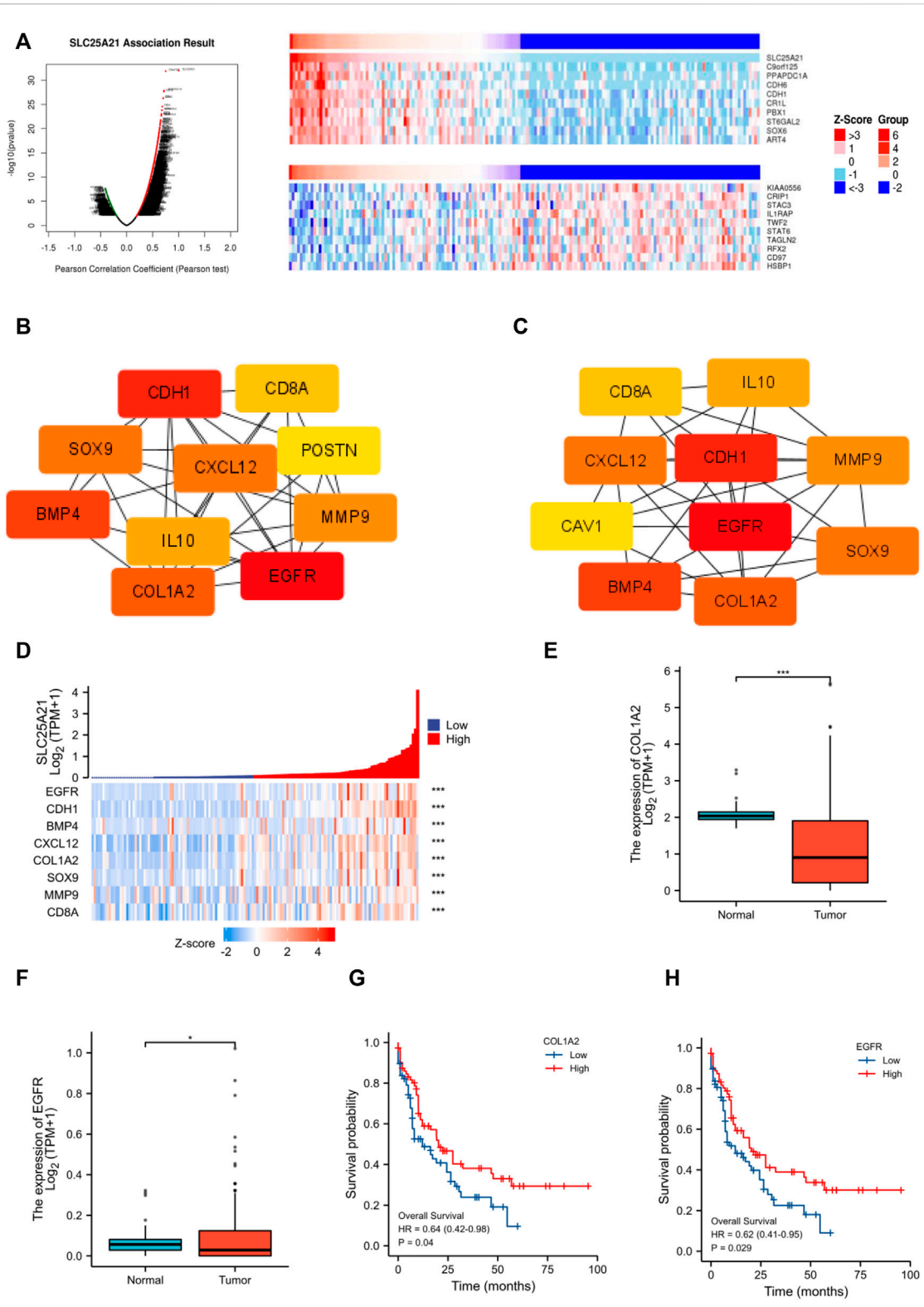


FIGURE 5
PPI network construction and clinical significance of hub genes. **(A)** Coexpression analysis of SLC25A21 in the TCGA-LAML dataset. The top 10 positively/negatively correlated genes are displayed. **(B–C)** The top 15 hub genes were selected on the basis of **(B)** MNC and **(C)** degree. **(D)** The association of SLC25A21 with eight hub genes (EGFR, CDH1, BMP4, CXCL12, COL1A2, SOX9, MMP9, and CD8A). **(E)** Expression levels of COL1A2 in AML patients (n = 132) and normal participants (n = 70). **(F)** Expression levels of EGFR in AML patients (n = 132) and normal participants (n = 70). **(G)** The difference in OS between patients with high and low COL1A2 expression levels shown by Kaplan-Meier curves. **(H)** The difference in OS between patients with high and low EGFR expression levels shown by Kaplan-Meier curves. (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

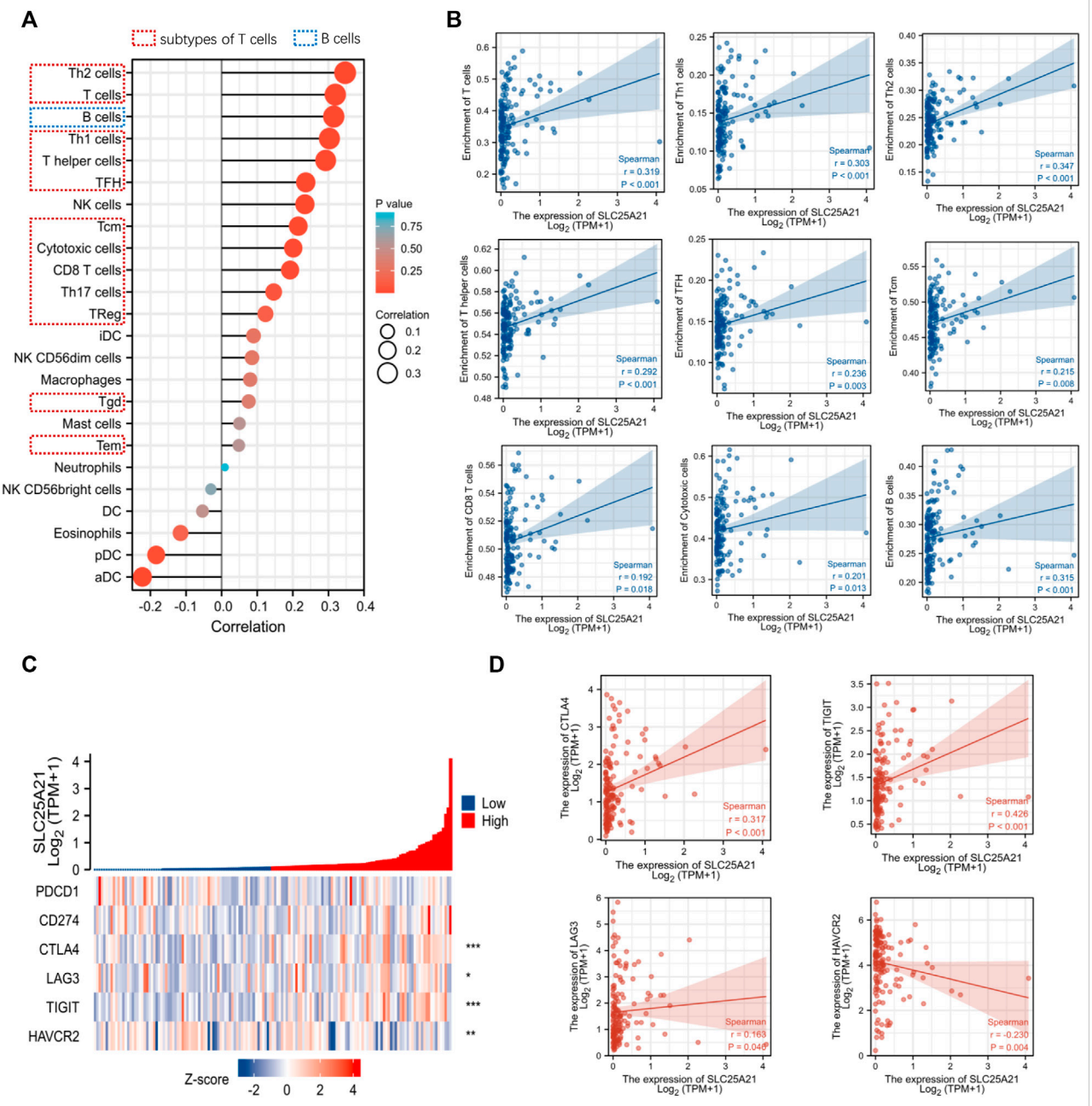


FIGURE 6
Correlation analysis between the level of SLC25A21 expression and immune cell infiltration or immune checkpoint molecules. **(A)** The relative contents of 24 kinds of immune cells in AML. **(B)** Spearman's correlations were used to quantify the correlation of SLC25A21 expression with the number of infiltrating level B cells and subtypes of T-cells. **(C)** The association of SLC25A21 with five immune checkpoint molecules (PDCD1, CD274, CTLA-4, LAG-3, TIGIT, and HAVCR2). **(D)** Spearman's correlation was used to quantify the correlation of SLC25A21 expression with immune checkpoint molecules (r is Spearman's correlation coefficient).

analysis to explore interrelationships. The number of enriched genes was ranked from most to least common: collagen-containing extracellular matrix, extracellular structure organization, extracellular matrix organization, cytokine–cytokine receptor interaction, extracellular matrix structural constituent, growth factor binding, protein digestion and absorption, malaria (Figure 4C).

Finally, we utilized GSEA to assess key regulatory pathways for *SLC25A21* expression. We found 17 significant pathways associated with *SLC25A21* (Table 2), of which the major affected pathways

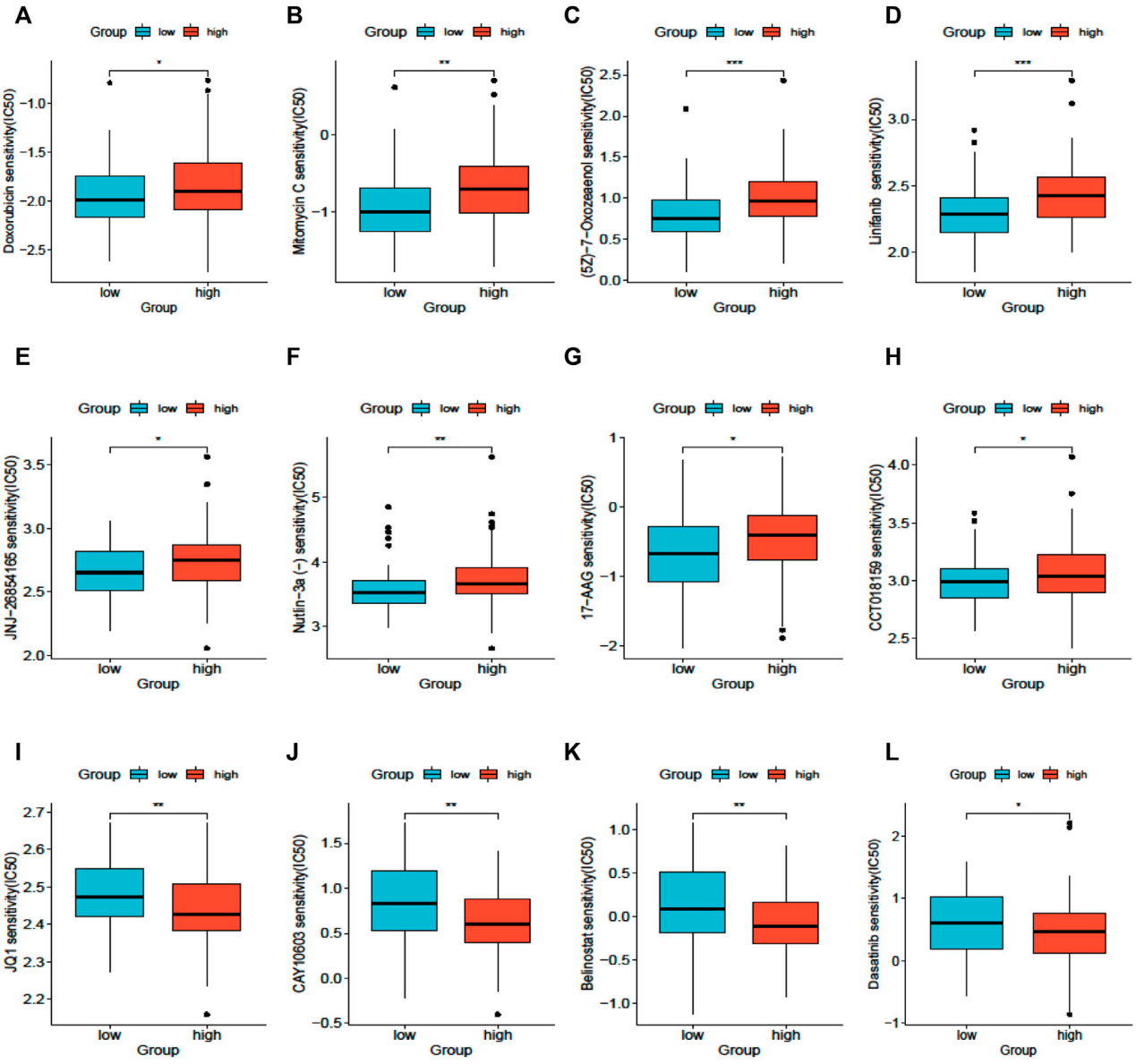


FIGURE 7
Drug sensitivity analysis based on SLC25A21. **(A)**, Doxorubicin. **(B)**, Mitomycin **(C)**, 7-Oxozeaenol. **(D)**, Linifanib. **(E)**, JNJ-26854,165. **(F)**, Niutlin-3a. **(G)**, 17-AAG. **(H)**, CCT018159. **(I)**, JQ1. **(J)**, CAY10603. **(K)**, Belinostat. **(L)**, Dasatinib.

included epithelial-mesenchymal transition, heme metabolism, angiogenesis, KRAS signalling, oxidative phosphorylation, DNA repair, MYC targets and reactive oxygen species (Figure 4H).

Identification of hub genes associated with *SLC25A21* expression

Next, we constructed and analysed the PPI network and coexpression modules. As indicated in Figure 5A, most genes in AML were positively correlated with the expression of *SLC25A21*. A DEG-related PPI network was constructed to determine hub genes. The top 10 hub genes were identified by the MNC and Degree methods by using the cytoHubba plug-in of Cytoscape (Figures 5B,C). Furthermore, we observed eight shared hub genes (*EGFR*, *CDH1*, *CXCL12*, *CD8A*, *MMP9*, *SOX9*, *BMP4*, and *COL1A2*) between the above two gene lists. In addition, we detected the associations between *SLC25A21* and hub genes. The results showed that *SLC25A21* had significant correlations with *EGFR* (p value <0.001, correlation coefficient:

0.569), *CDH1* (p value <0.001, correlation coefficient: 0.709), *CXCL12* (p value <0.001, correlation coefficient: 0.590), *CD8A* (p value <0.001, correlation coefficient: 0.441), *MMP9* (p value <0.001, correlation coefficient: 0.295), *SOX9* (p value <0.001, correlation coefficient: 0.438), *BMP4* (p value <0.001, correlation coefficient: 0.383), and *COL1A2* (p value <0.001, correlation coefficient: 0.538) (Figure 5D). Finally, we examined the relationship between the levels of hub genes and prognosis and found that only *EGFR* and *COL1A2* were positively correlated with *SLC25A21* and linked to poor clinical outcomes in patients with AML (Figures 5E–H).

Correlation analysis of *SLC25A21* and immune cells or immune checkpoint molecules

Tumor infiltrating lymphocytes affect the survival of patients with various cancers. Therefore, 24 kinds of infiltrating immune cells were evaluated to describe the association between the levels of *SLC25A21* expression and immune infiltration in AML. The results showed that the expression level of *SLC25A21* had an obvious positive correlation with the numbers of infiltrating B cells, T-cells, Th1 cells, Th2 cells, T helper cells, Tfh cells, CD8 T-cells, cytotoxic cells and Tcm cells (Figure 6A). The details of the quantitative analysis with Spearman's correlation coefficient are shown in Figure 6B.

Furthermore, we clarified the relationship between *SLC25A21* and immune checkpoint (*PDCD1*, *CD274*, *CTLA4*, *LAG-3*, *TIGIT*, and *HAVCR2*) expression. In our study, *SLC25A21* was significantly correlated with *CTLA4*, *LAG3*, *TIGIT*, *CD274*, and *TIGIT*. Details of the correlation analysis are shown in Figures 6C,D.

Drug sensitivity analysis

The results of drug sensitivity analysis for the high- and low-*SLC25A21* groups showed that the *SLC25A21* low expression group may be more sensitive to cell cycle inhibitors (doxorubicin and mitomycin C), vascular endothelial growth factor receptor (VEGFR) tyrosine kinase inhibitors (linifanib and 7-oxozeaenol), p53 activators (JNJ-26854165 and Nutlin-3a), and heat shock protein 90 (HsP90) inhibitors (CCT018159 and 17-AAG) but resistant to histone deacetylase (HDAC) inhibitors (JQ1, CAY10603, and belinostat) and tyrosine kinase inhibitors (dasatinib). These results indicated that *SLC25A21* has a significant correlation with chemotherapy and targeted therapy regimens for AML (Figures 7A–L).

Discussion

AML is a highly heterogeneous disease with various cytogenetic and genetic alterations. Genetic abnormalities are not only the pathogenic basis of AML, but they have important treatment and prognostic implications. In this study, we screened transcriptome data for AML in public databases to discover novel molecular biomarkers with a potential impact on prognosis and/or therapeutic response. We identified DEGs between AML patients and healthy donors in two independent cohorts. Therefore, a list of 14 AML-specific genes was obtained, including *IL1R2*, *MMP8*, *FGF13*, *SLC25A21*, etc.

(Le Sommer et al., 2018; Nobrega-Pereira et al., 2018) *SLC25A21* SLC25A21 is a 2-oxoglutarate transporter embedded in the mitochondrial inner membrane and, in some cases, organelle membranes. The expression of human *SLC25A21* has a wide distribution with very little variation between tissues. A recent study revealed that *SLC25A21* suppresses cell growth and plays a pathogenic role on bladder cancer (Wang et al., 2021). Recent studies have shown that metabolic molecules are dysregulated in AML cells and play key roles in leukaemogenesis, contributing to chemoresistance and disease relapse (Le Sommer et al., 2018; Nobrega-Pereira et al., 2018). Targeting cell metabolism is now considered a viable therapeutic strategy for AML. Therefore, we focused on the metabolism-related gene *SLC25A21* for further studies.

Given the above, we first explored the association of *SLC25A21* gene expression levels with main clinical features in TCGA-LAML cohorts of AML patients. *SLC25A21* We found that *SLC25A21* was significantly downregulated in AML patients. As expected, a low level of *SLC25A21* was associated with higher WBC counts, higher BM and PB blast abundance and poor prognosis. Thus, we speculated that abnormally low expression of *SLC25A21* plays an unfavourable role in promoting AML cell proliferation and survival while preventing leukaemic cell differentiation.

Hence, we explored the possible molecular mechanism underlying this association by using a bioinformatics approach. The enriched GO terms and KEGG pathways were mainly involved in growth factor binding, collagen-containing extracellular matrix, extracellular structure organization, the PI3K-Akt signalling pathway and the Wnt signalling pathway. Concurrently, GSEA showed enrichment of epithelial-to-mesenchymal transition, the KRAS signalling pathway, oxidative phosphorylation, DNA repair and the reactive oxygen species (ROS) pathway.

Mitochondria are the primary intracellular source of ROS and play important roles in aerobic metabolism and oxidative phosphorylation. Thus, dysregulation of mitochondrial metabolism is closely related to the development and progression of haematopoietic malignancies (Basak and Banerjee, 2015; Porporato et al., 2018). As a carrier embedded in mitochondria, overexpression of *SCL25A21* resulted in efflux of α-KG from mitochondria, leading to upregulation of ROS accumulation, which in turn induced mitochondrial apoptosis (Wang et al., 2021). Moreover, increased ROS levels drive a cycle of genomic instability. Leading to DNA double-strand breaks (DSBs) and altered DNA repair. The accumulation of intracellular ROS can

promote tumour proliferation, but excessive accumulation of ROS can lead to cell apoptosis (Trotta et al., 2017; Aggarwal et al., 2019). Recent studies have revealed that the majority of functionally defined leukaemia stem cells (LSCs) are functionally characterized by relatively low levels of ROS. Meanwhile, several EMT-related genes conferring properties of "stemness" were strongly associated with shorter OS in AML patients (Stavropoulou et al., 2016; Carmichael et al., 2020; Almotiri et al., 2021). More importantly, the PI3K and KRAS signalling pathways play important roles in the proliferation and differentiation of haematopoietic cells (Crespo and Leon, 2000; Martelli et al., 2009; Nepstad et al., 2018). Thus, we speculated that the pathological mechanism of *SLC25A21* may be related to these signalling pathways.

Furthermore, through a series of rigorous screens, two hub genes (*EGFR* and *COL1A2*) that could accurately predict the prognosis of AML were found. It has been reported that dysregulation of *EGFR* can lead to the development of malignancy (Cheng et al., 2011; Singh et al., 2016). *EGFR* repairs the DNA of HSCs by activating DNA-dependent protein kinase catalytic subunit (DNA-PKcs), leading to the regeneration of normal haematopoietic cells. Experimental studies have shown that deletion of *EGFR* in progenitor cells results in reduced DNA-PKcs activity, thus reducing the ability of the cells to regenerate normal HSCs (Fang et al., 2020). In AML, we speculate that EGFR may play the same role to inhibit normal haematopoiesis, resulting in a shorter survival time for patients with low EGFR expression. *COL1A2* has also been implicated in gastric cancer, colorectal cancer, prostate cancer, pancreatic cancer (Yu et al., 2018; Wu et al., 2019; Nie et al., 2020; Liu et al., 2022), and so on. Moreover, *COL1A2* has been identified as a hub gene in *FLT3*-mutated AML (Chen et al., 2020b). Although *SLC25A21, EGFR*, and *COL1A2* are linked to tumour-associated signalling pathways, the precise mechanisms of this synergy remain unclear. Further in-depth studies are needed to address this issue in more detail.

Additionally, metabolic molecular abnormalities may facilitate AML cell escape and immune detection and severely reduce the efficacy of immunotherapy (Mougiakakos, 2019). Several studies have shown that the level of immune infiltration and immune evasion mechanisms of AML cells determine their immune evasion ability (Rosenthal et al., 2019). Therefore, we investigated the relationship between *SLC25A21* expression and immune infiltration levels in AML patients and found that with downregulation of *SLC25A21*, the infiltration levels of various T-cells and B cells were greatly decreased. Next, we also showed that *SLC25A21* expression had a positive correlation with some immune checkpoint genes (CTLA-4, LAG3, TIGIT, and HAVCR2), which serve as activation markers of T-cells and affect antitumor immunity. We speculate that this may be because the number of infiltrating immune cells is significantly reduced, resulting in decreased expression of molecular markers of T-cell activation. These results suggest that *SLC25A21* may lead to immune escape in AML. This observation may provide a framework to guide further investigation of *SLC25A21* in clinical and basic science research.

Last, the ultimate objective of our research is to provide clinicians with guidelines to choose the appropriate therapeutic regimens for each AML patient. With the development of next-generation sequencing technology, several genetic aberrations have been found to contribute to drug resistance in AML (Gollner et al., 2017; Hou et al., 2017; Nechiporuk et al., 2019). In this study, we analysed the correlation between *SLC25A21* and drug resistance in AML. Patients with low *SLC25A21* expression levels were sensitive to doxorubicin, mitomycin, lapatinib, midostaurin, sorafenib, linifanib, Nutlin-3a, 17-AAG, 5-fluorouracil, 7-oxozeaenol, JNJ-26854165, CCT018159, bleomycin, and FH535 but resistant to JQ1, CUDC-101, dasatinib, and GNF-2. These results indicate that downregulation of *SLC25A21* may promote sensitivity to doxorubicin, the cornerstone regimen for AML. These results suggest that while *SLC25A21* affects prognosis in AML, patients with low expression of *SLC25A21* may still benefit from traditional chemotherapy regimens.

However, our study has several limitations. First, we explored the mutational frequency of *SLC25A21* in 6 independent AML studies ($n = 2,177$) and found a frequency of approximately 0.1%–0.2% (Supplementary Figure S2). In addition, we observed enrichment of transcriptional regulation pathways in GO analysis. Therefore, the upstream transcriptional regulatory mechanism of *SLC25A21* remains to be uncovered. Second, we evaluated the diagnostic value and drug sensitivity of *SLC25A21* by using public resources. However, we have not yet tested some new clinically emerging targeted drugs, such as venetoclax, due to the limitations of the training database. Last, all associations between *SCL25A21* and AML-associated immune molecules lack functional validation and detection of the potential cellular and molecular mechanisms. Future studies will build on these points with a view toward providing new options for precision medicine approaches and improving the treatment of AML patients.

Conclusion

Taken together, our preliminary findings showed that low expression levels of the metabolism-related gene *SLC25A21* had an unfavourable effect on the overall survival of AML patients and may be correlated with immune escape. A low level of *SLC25A21* could be an independent predictor of poor prognostic for AML patients. This discovery could promote the development of novel targeted drugs and provide therapeutic options for personalized therapy.

Data availability statement

The datasets presented in this study can be found in the GEO,TCGA and GETx repositories: https://www.ncbi.nlm.nih.gov/geo/; https://portal.gdc.cancer.gov/projects/TCGA-LAML;

https://xenabrowser.net/datapages/. The accession numbers can be found in the article.

Ethics statement

The studies involving human participants were reviewed and approved by the ethics committee of Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences by (KT2020016-EC-2). The patients/participants provided their written informed consent to participate in this study.

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

Author contributions

WW, QL, YZ, and JS conceived the study. WW and QL collected and analyzed the data, wrote the paper. WW, QL, JZ, HP, ZG, and LF contributed to the data collection and analysis. YZ and JS designed the research and give an approval of the final manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.970316/full#supplementary-material

SUPPLEMENTARY FIGURE S1
(A), SLC25A21 expression was not associated with cytogenetics in AML. **(B)**, SLC25A21 expression was not associated with French–American–British (FAB) classifications.

SUPPLEMENTARY FIGURE S2
(A), Mutation frequency of *SLC25A21* in leukaemia.

SUPPLEMENTARY TABLE S1
Univariate and multivariate Cox regression analysis.

SUPPLEMENTARY TABLE S2
GO enrichment and KEGG pathway analysis.

SUPPLEMENTARY TABLE S3
Patient sample information.

SUPPLEMENTARY TABLE S4
List of primer sequences.

References

Aggarwal, V., Tuli, H. S., Varol, A., Thakral, F., Yerer, M. B., Sak, K., et al. (2019). Role of reactive oxygen species in cancer progression: Molecular mechanisms and recent advancements. *Biomolecules* 9, E735. doi:10.3390/biom9110735

Almotiri, A., Alzahrani, H., Menendez-Gonzalez, J. B., Abdelfattah, A., Alotaibi, B., Saleh, L., et al. (2021). Zeb1 modulates hematopoietic stem cell fates required for suppressing acute myeloid leukemia. *J. Clin. Invest.* 131, 129115. doi:10.1172/JCI129115

Basak, N. P., and Banerjee, S. (2015). Mitochondrial dependency in progression of acute myeloid leukemia. *Mitochondrion* 21, 41–48. doi:10.1016/j.mito.2015.01.006

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Boczonadi, V., King, M. S., Smith, A. C., Olahova, M., Bansagi, B., Roos, A., et al. (2018). Mitochondrial oxodicarboxylate carrier deficiency is associated with mitochondrial DNA depletion and spinal muscular atrophy-like disease. *Genet. Med.* 20, 1224–1235. doi:10.1038/gim.2017.251

Bosc, C., Broin, N., Fanjul, M., Saland, E., Farge, T., Courdy, C., et al. (2020). Autophagy regulates fatty acid availability for oxidative phosphorylation through mitochondria-endoplasmic reticulum contact sites. *Nat. Commun.* 11, 4056. doi:10.1038/s41467-020-17882-2

Cancer Genome Atlas Research, N., Ley, T. J., Miller, C., Ding, L., Raphael, B. J., Mungall, A. J., et al. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 368, 2059–2074. doi:10.1056/NEJMoa1301689

Carmichael, C. L., Wang, J., Nguyen, T., Kolawole, O., Benyoucef, A., De Maziere, C., et al. (2020). The EMT modulator SNAI1 contributes to AML pathogenesis via its interaction with LSD1. *Blood* 136, 957–973. doi:10.1182/blood.2019002548

Chen, C. C., Li, B., Millman, S. E., Chen, C., Li, X., Morris, J. P., et al. (2020). Vitamin B6 addiction in acute myeloid leukemia. *Cancer Cell.* 37, 71–84. doi:10.1016/j.ccell.2019.12.002

Chen, K. T. J., Gilabert-Oriol, R., Bally, M. B., and Leung, A. W. Y. (2019). Recent treatment advances and the role of nanotechnology, combination products, and immunotherapy in changing the therapeutic landscape of acute myeloid leukemia. *Pharm. Res.* 36, 125. doi:10.1007/s11095-019-2654-z

Chen, S., Chen, Y., Zhu, Z., Tan, H., Lu, J., Qin, P., et al. (2020). Identification of the key genes and microRNAs in adult acute myeloid leukemia with FLT3 mutation by bioinformatics analysis. *Int. J. Med. Sci.* 17, 1269–1280. doi:10.7150/ijms.46441

Cheng, L., Zhang, S., Alexander, R., Yao, Y., MacLennan, G. T., Pan, C. x., et al. (2011). The landscape of EGFR pathways and personalized management of non-small-cell lung cancer. *Future Oncol.* 7, 519–541. doi:10.2217/fon.11.25

Consortium, G. T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. doi:10.1126/science.aaz1776

Crespo, P., and Leon, J. (2000). Ras proteins in the control of the cell cycle and cell differentiation. *Cell. Mol. Life Sci.* 57, 1613–1636. doi:10.1007/pl00000645

Ding, C., Shan, Z., Li, M., Chen, H., Li, X., and Jin, Z. (2021). Characterization of the fatty acid metabolism in colorectal cancer to guide clinical therapy. *Mol. Ther. Oncolytics* 20, 532–544. doi:10.1016/j.omto.2021.02.010

Fang, T., Zhang, Y., Chang, V. Y., Roos, M., Termini, C. M., Signaevskaia, L., et al. (2020). Epidermal growth factor receptor-dependent DNA repair promotes murine

and human hematopoietic regeneration. *Blood* 136, 441–454. doi:10.1182/blood.2020005895

Fiermonte, G., Dolce, V., PaLmieri, L., VenturaM.Runswick, M. J., PalmieriF., et al. (2001). Identification of the human mitochondrial oxodicarboxylate carrier. Bacterial expression, reconstitution, functional characterization, tissue distribution, and chromosomal location. *J. Biol. Chem.* 276, 8225–8230. doi:10.1074/jbc.M009607200

Fiermonte, G., PaLmieri, L., Dolce, V., Lasorsa, F. M., PalmieriF.Runswick, M. J., et al. (1998). The sequence, bacterial expression, and functional reconstitution of the rat mitochondrial dicarboxylate transporter cloned via distant homologs in yeast and *Caenorhabditis elegans*. *J. Biol. Chem.* 273, 24754–24759. doi:10.1074/jbc.273.38.24754

Forte, D., Garcia-Fernandez, M., Sanchez-Aguilera, A., Stavropoulou, V., Fielding, C., Martin-Perez, D., et al. (2020). Bone marrow mesenchymal stem cells support acute myeloid leukemia bioenergetics and enhance antioxidant defense and escape from chemotherapy. *Cell. Metab.* 32, 829–843. doi:10.1016/j.cmet.2020.09.001

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

Gene Ontology, C. (2021). The gene ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. doi:10.1093/nar/gkaa1113

Goldman, M. J., Craft, B., Hastie, M., Repecka, K., McDade, F., Kamath, A., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* 38, 675–678. doi:10.1038/s41587-020-0546-8

Gollner, S., Oellerich, T., Agrawal-Singh, S., Schenk, T., Klein, H. U., Rohde, C., et al. (2017). Loss of the histone methyltransferase EZH2 induces resistance to multiple drugs in acute myeloid leukemia. *Nat. Med.* 23, 69–78. doi:10.1038/nm.4247

Haferlach, T., Kohlmann, A., Wieczorek, L., Basso, G., Kronnie, G. T., Bene, M. C., et al. (2010). Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: Report from the international microarray Innovations in leukemia study group. *J. Clin. Oncol.* 28, 2529–2537. doi:10.1200/JCO.2009.23.4732

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

Hou, P., Wu, C., Wang, Y., Qi, R., Bhavanasi, D., Zuo, Z., et al. (2017). A genome-wide CRISPR screen identifies genes critical for resistance to FLT3 inhibitor AC220. *Cancer Res.* 77, 4402–4413. doi:10.1158/0008-5472.CAN-16-1627

Kanehisa, M., Sato, Y., and Kawashima, M. (2021). KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci.* 31, 47–53. doi:10.1002/pro.4172

Kunji, E. R. S., King, M. S., Ruprecht, J. J., and Thangaratnarajah, C. (2020). The SLC25 carrier family: Important transport proteins in mitochondrial physiology and pathology. *Physiol. (Bethesda)* 35, 302–327. doi:10.1152/physiol.00009.2020

Le Sommer, S., Morrice, N., Pesaresi, M., Thompson, D., Vickers, M. A., Murray, G. I., et al. (2018). Deficiency in protein tyrosine phosphatase PTP1B shortens lifespan and leads to development of acute leukemia. *Cancer Res.* 78, 75–87. doi:10.1158/0008-5472.CAN-17-0946

Liu, S., He, B., and Li, H. (2022). Bisphenol S promotes the progression of prostate cancer by regulating the expression of COL1A1 and COL1A2. *Toxicology* 472, 153178. doi:10.1016/j.tox.2022.153178

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Martelli, A. M., Evangelisti, C., Chiarini, F., Grimaldi, C., Manzoli, L., and McCubrey, J. A. (2009). Targeting the PI3K/AKT/mTOR signaling network in acute myelogenous leukemia. *Expert Opin. Investig. Drugs* 18, 1333–1349. doi:10.1517/14728220903136775

Metsalu, T., and Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43, W566–W570. doi:10.1093/nar/gkv468

Metzeler, K. H., Hummel, M., Bloomfield, C. D., Spiekermann, K., Braess, J., Sauerland, M. C., et al. (2008). An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* 112, 4193–4201. doi:10.1182/blood-2008-02-134411

Mougiakakos, D. (2019). The induction of a permissive environment to promote T cell immune evasion in acute myeloid leukemia: The metabolic perspective. *Front. Oncol.* 9, 1166. doi:10.3389/fonc.2019.01166

Nechiporuk, T., Kurtz, S. E., Nikolova, O., Liu, T., Jones, C. L., D'Alessandro, A., et al. (2019). The TP53 apoptotic network is a primary mediator of resistance to BCL2 inhibition in AML cells. *Cancer Discov.* 9, 910–925. doi:10.1158/2159-8290.CD-19-0125

Nepstad, I., Reikvam, H., Brenner, A. K., Bruserud, O., and Hatfield, K. J. (2018). Resistance to the antiproliferative *in vitro* effect of PI3K-Akt-mTOR inhibition in primary human acute myeloid leukemia cells is associated with altered cell metabolism. *Int. J. Mol. Sci.* 19, E382. doi:10.3390/ijms19020382

Nie, K., Shi, L., Wen, Y., Pan, J., Li, P., Zheng, Z., et al. (2020). Identification of hub genes correlated with the pathogenesis and prognosis of gastric cancer via bioinformatics methods. *Minerva Med.* 111, 213–225. doi:10.23736/S0026-4806.19.06166-4

Nobrega-Pereira, S., Caiado, F., Carvalho, T., Matias, I., Graca, G., Goncalves, L. G., et al. (2018). VEGFR2-Mediated reprogramming of mitochondrial metabolism regulates the sensitivity of acute myeloid leukemia to chemotherapy. *Cancer Res.* 78, 731–741. doi:10.1158/0008-5472.CAN-17-1166

Pei, S., Pollyea, D. A., Gustafson, A., Stevens, B. M., Minhajuddin, M., Fu, R., et al. (2020). Monocytic subclones confer resistance to venetoclax-based therapy in patients with acute myeloid leukemia. *Cancer Discov.* 10, 536–551. doi:10.1158/2159-8290.CD-19-0710

Porporato, P. E., Filigheddu, N., Pedro, J. M. B., Kroemer, G., and Galluzzi, L. (2018). Mitochondrial metabolism and cancer. *Cell. Res.* 28, 265–280. doi:10.1038/cr.2017.155

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

Rosenthal, R., Cadieux, E. L., Salgado, R., Bakir, M. A., Moore, D. A., Hiley, C. T., et al. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature* 567, 479–485. doi:10.1038/s41586-019-1032-7

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

Singh, D., Attri, B. K., Gill, R. K., and Bariwal, J. (2016). Review on EGFR inhibitors: Critical updates. *Mini Rev. Med. Chem.* 16, 1134–1166. doi:10.2174/1389557516666160321114917

Stavropoulou, V., Kaspar, S., Brault, L., Sanders, M. A., Juge, S., Morettini, S., et al. (2016). MLL-AF9 expression in hematopoietic stem cells drives a highly invasive AML expressing EMT-related genes linked to poor outcome. *Cancer Cell.* 30, 43–58. doi:10.1016/j.ccell.2016.05.011

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

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

Trotta, A. P., Gelles, J. D., Serasinghe, M. N., Loi, P., Arbiser, J. L., and Chipuk, J. E. (2017). Disruption of mitochondrial electron transport chain function potentiates the pro-apoptotic effects of MAPK inhibition. *J. Biol. Chem.* 292, 11727–11739. doi:10.1074/jbc.M117.786442

Vivian, J., Rao, A. A., Nothaft, F. A., Ketchum, C., Armstrong, J., Novak, A., et al. (2017). Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* 35, 314–316. doi:10.1038/nbt.3772

Wang, Y., Gao, J., Hu, S., Zeng, W., Yang, H., Chen, H., et al. (2021). SLC25A21 suppresses cell growth in bladder cancer via an oxidative stress-mediated mechanism. *Front. Oncol.* 11, 682710. doi:10.3389/fonc.2021.682710

Wu, J., Liu, J., Wei, X., Yu, Q., Niu, X., Tang, S., et al. (2019). A feature-based analysis identifies COL1A2 as a regulator in pancreatic cancer. *J. Enzyme Inhib. Med. Chem.* 34, 420–428. doi:10.1080/14756366.2018.1484734

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

Yu, Y., Liu, D., Liu, Z., Li, S., Ge, Y., Sun, W., et al. (2018). The inhibitory effects of COL1A2 on colorectal cancer cell proliferation, migration, and invasion. *J. Cancer* 9, 2953–2962. doi:10.7150/jca.25542

frontiers | Frontiers in Genetics

TYPE Systematic Review
PUBLISHED 31 October 2022
DOI 10.3389/fgene.2022.976673

OPEN ACCESS

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

REVIEWED BY
Salvador F. Aliño,
University of Valencia, Spain
Claudia Banescu,
University of Medicine, Pharmacy, Sciences and Technology of TârguMureş, Romania

*CORRESPONDENCE
Xu-Liang Shen,
shenxlcyp@sohu.com
Xiao-Feng He,
393120823@qq.com

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

RECEIVED 23 June 2022
ACCEPTED 13 October 2022
PUBLISHED 31 October 2022

CITATION
Zhao Y, Wang D, Zhang C-Y, Liu Y-J, Wang X-H, Shi M-Y, Wang W, Shen X-L and He X-F (2022), Individual and combined effects of the GSTM1, GSTT1, and GSTP1 polymorphisms on leukemia risk: An updated meta-analysis. *Front. Genet.* 13:976673. doi: 10.3389/fgene.2022.976673

Individual and combined effects of the *GSTM1*, *GSTT1*, and *GSTP1* polymorphisms on leukemia risk: An updated meta-analysis

Yan Zhao[1], Di Wang[1], Cheng-Yu Zhang[1], Yan-Ju Liu[1], Xiao-Hui Wang[1], Meng-Ying Shi[1], Wei Wang[2], Xu-Liang Shen[3]* and Xiao-Feng He[4,5]*

[1]Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi, China, [2]Beijing Zhendong Guangming Pharmaceutical Research Institute, Beijing, China, [3]Department of Hematology, Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi, China, [4]Institute of Evidence-Based Medicine, Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi, China, [5]Department of Epidemiology, School of Public Health to Southern Medical University, Guangzhou, China

Background: Several meta-analyses have analyzed the association of GSTM1 present/null, GSTT1 present/null, and GSTP1 Ile105Val polymorphisms with leukemia risk. However, the results of these meta-analyses have been conflicting. Moreover, they did not evaluate the combined effects of the three aforementioned gene polymorphisms. Furthermore, they did not appraise the credibility of the positive results. Finally, many new studies have been published. Therefore, an updated meta-analysis was conducted.

Objectives: To further explore the relationship of the three aforementioned gene polymorphisms with leukemia risk.

Methods: The crude odds ratios (ORs) and 95% confidence intervals (CIs) were applied to evaluate the association of the individual and combined effects of the three aforementioned genes. Moreover, the false-positive report probability (FPRP) and Bayesian false discovery probability (BFDP) were applied to verify the credibility of these statistically significant associations.

Results: Overall, the individual GSTM1, GSTT1, and GSTP1 Ile105Val polymorphisms added leukemia risk. On combining GSTM1 and GSTT1, GSTM1 and GSTP1, and GSTT1 and GSTP1 polymorphisms, positive results were also observed. However, no significant association was observed between the combined effects of these three polymorphisms with leukemia risk in the overall analysis. Moreover, when only selecting Hardy–Weinberg

Abbreviations: GSTs, glutathione S-transferases; GSTM1, glutathione S-transferase M1; GSTT1, glutathione S-transferase T1; GSTP1, glutathione S-transferase P1; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; CML, chronic myeloid leukemia; CLL, chronic lymphoblastic leukemia; HC, healthy controls; NBDC, Nonblood disease controls; BFDP, Bayesian false discovery probability; FPRP, false-positive report probability; HWE, Hardy–Weinberg equilibrium; ORs, odds ratios; CIs, confidence intervals.

equilibrium (HWE) and medium- and high-quality studies, we came to similar results. However, when the FPRP and BFDP values were applied to evaluate the credibility of positive results, the significant association was only observed for the GSTT1 null genotype with leukemia risk in Asians (BFDP = 0.367, FPRP = 0.009).

Conclusion: This study strongly suggests a significant increase in the risk of leukemia in Asians for the GSTT1 null genotype.

KEYWORDS

glutathione S-transferases, *GSTM1*, *GSTT1*, *GSTP1*, leukemia

Introduction

Leukemia is a cancer of hematology, characterized by abnormal hematopoietic function and malignant cloning of white blood cells. Leukemia includes acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), chronic myeloid leukemia (CML), and chronic lymphoblastic leukemia (CLL) (Ouerhani et al., 2011). Over the past few decades, we have made giant progress in the early diagnosis of diseases and treatment, yet the number of new cases of leukemia are still increasing, and the death cases also continue to increase. Therefore, leukemia has become one huge threat to human health (Ferlay et al., 2015). As we all know, leukemia is deemed to be a complex disease, which is determined by hereditary and environmental factors (Arruda et al., 2001; Krajinovic et al., 2001). Although previous studies showed that chemicals, ionizing radiation, and viral infections were the potential pathogenic factors of leukemia (Maia Rda and Wünsch Filho, 2013; Schüz and Erdmann, 2016), there were great individual differences in disease susceptibility when these patients were exposed to the aforementioned carcinogenic agents. Therefore, research studies on hereditary factors that affect leukemia may improve our further understanding of the pathogenesis of leukemia; in addition, they might provide new evidence for the treatment of leukemia.

Glutathione S-transferase (GST) is a kind of phase II enzyme which includes M1, P1, and T1; the main functions of the three aforementioned genes were the metabolism of xenobiotics, reactive oxygen species, and carcinogens for detoxification and metabolism (Strange et al., 2001). A partial gene deletion of *GSTM1* and *GSTT1* (null genotypes) can result in the complete absence of *GSTM1* and *GSTT1* enzyme activities; the former is located on chromosome 1 (1p13.3) and the latter is situated at chromosome 22 (22q11.2) (Pearson et al., 1993; Webb et al., 1996; Strange and Fryer, 1999). *GSTP1* gene polymorphism is a single-nucleotide polymorphism, whose polymorphism lies in exon 5 codon 105, when substitution of A with G leads to change in isoleucine (IIe) to valine (Val), thereby giving rise to decreased enzymatic activity (Harries et al., 1997; Ryberg et al., 1997). Previous research studies have indicated that the complete deletion of *GSTM1*, *GSTT1,* or *GSTP1* polymorphisms can bring about diminished gene expression and enzymatic activity (Strange et al., 1998; Strange et al., 2001; Hollman et al., 2016). The *GSTM1* and *GSTT1* showed a high degree of polymorphism, one of the polymorphisms being the entire deletion of the gene that results in the lapse of enzymatic activity (Alves et al., 2002).

Several meta-analyses analyzed the association of *GSTM1* present/null, *GSTT1* present/null, and *GSTP1* IIe105Val polymorphisms with leukemia risk. However, results of these meta-analyses were conflicting. Moreover, they did not evaluate the combined effects of the three aforementioned gene polymorphisms. Furthermore, they did not appraise the credibility of the positive results. Finally, many new studies have been published. Therefore, an updated meta-analysis was conducted.

Materials and methods

Search strategy

Five databases including PubMed, Embase, Web of Science, CNKI, and WanFang were applied to search the literature (deadline, 26 May 2022). The following retrieval strategy was employed: (glutathione S-transferase M1 OR GSTM1 OR glutathione S-transferase T1 OR GSTT1 OR glutathione S-transferase P1 OR GSTP1) AND (polymorphism OR genotype OR mutation OR variant OR allele) AND (leukemia OR leukaemia). Furthermore, if necessary, we contacted the corresponding authors by e-mail.

Inclusion and exclusion criteria

The studies that met the following criteria were included: 1) case-control or cohort study, 2) genotype data or odds ratio (OR) with 95% confidence interval (CI) provided, and 3) investigation of the association of the three aforementioned gene polymorphisms with the risk of leukemia. Studies such as overlapping data, case reports, editorials, reviews, letters, and meta-analyses were excluded.

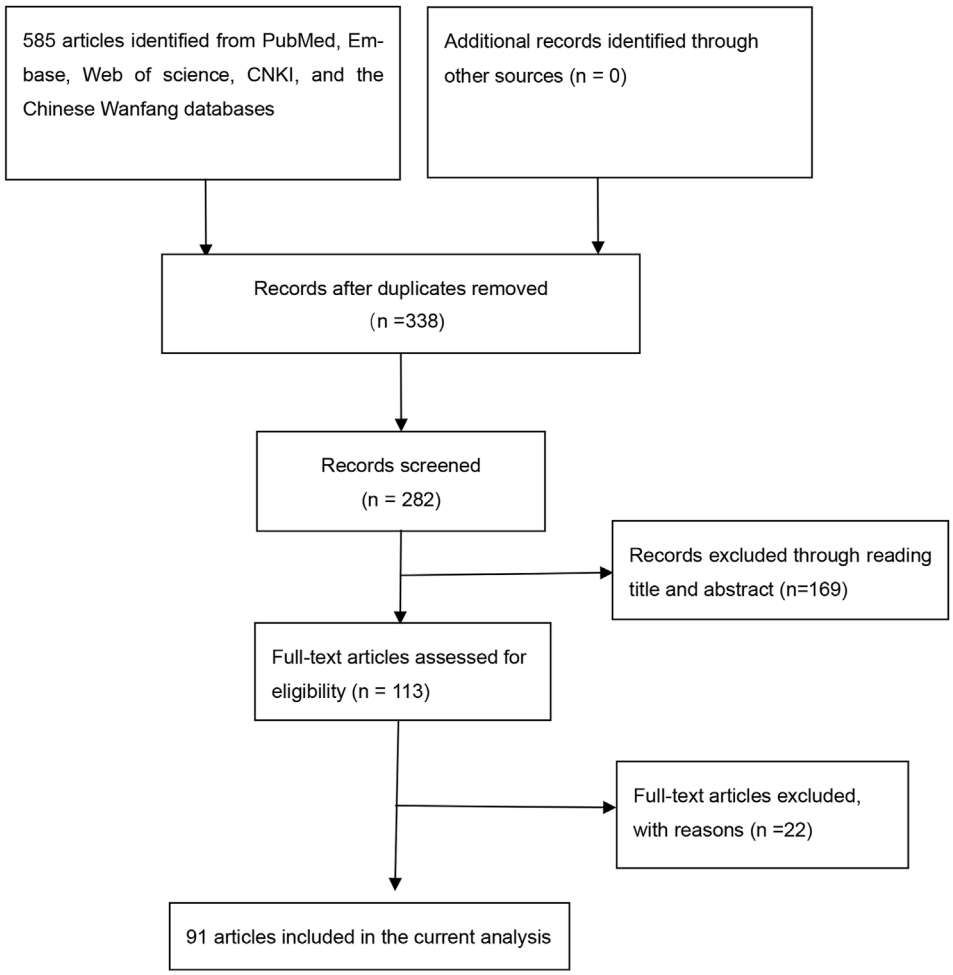


FIGURE 1
Flow diagram for the literature search.

Data extraction and quality assessment

Information was extracted and checked by two researchers from all selected studies. Any disagreement was solved through discussion. Extracted information in shown in Supplementary Tables S1–S3. Quality assessment was conducted by two authors independently (Supplementary Table S4). For *GSTM1* and *GSTT1* null genotypes, we considered studies that scored ≥10 as high quality; for *GSTP1* IIe105Val, studies scoring ≥12 were deemed as high quality.

Statistical analysis

We used crude ORs and 95% CIs to estimate the associations between GST (M1, T1, and P1 IIe105Val) polymorphisms and leukemia risk. The Q statistic and I^2 value were carried out to evaluate heterogeneity (Higgins et al., 2003). Only a random-effect model was used because the pooled results were same when $I^2 = 0\%$ using random-effect and fixed-effect models (Der Simonian and Laird, 2015). We performed ORs with the corresponding 95% CIs following the genetic models. In GSTM1 and GSTT1 null genotypes, we used null vs. present model to calculate the pooled ORs with their 95% CIs. In GSTP1 IIe105Val, five genetic models were used (Val/Val vs. IIe/IIe, IIe/Val vs. IIe/IIe, Val/Val vs. IIe/IIe + IIe/Val, Val/Val + IIe/Val vs. IIe/IIe, and Val vs. IIe). In the combination of GSTM1 present/null and GSTT1 present/null, we applied the following six genetic models: model 1: M1 present/T1 null vs. M1 present/T1 present, model 2: M1 null/T1 present vs. M1 present/T1 present, model 3: M1 null/T1 null vs. M1 present/T1 present, model 4: All one risk genotypes vs. M1 present/T1 present, model 5: All risk genotypes vs. M1 present/T1 present, and model 6: M1 null/T1 null vs. M1 present/T1 present + M1 present/T1 null + M1 null/T1 present in the

TABLE 1 Meta-analysis of the association of GSTM1 polymorphism with risk of leukemia.

Variable	n	Cases/Controls	Test of association	Test of heterogeneity		Model
			OR (95%CI)	P_h	I^2 (%)	
Overall	98	13477/22523	1.28 (1.17–1.40)	<0.001	68.3	Random-effect
Ethnicity						
Indian	14	1600/2465	1.25 (0.89–1.77)	<0.001	84.6	Random-effect
Asian	24	3265/6028	1.50 (1.29–1.73)	0.002	51.2	Random-effect
Caucasian	47	7466/11124	1.17 (1.07–1.28)	<0.001	46.0	Random-effect
African	6	662/886	1.99 (1.30–3.94)	0.006	69.0	Random-effect
Age group						
Adults	37	5811/9440	1.26 (1.11–1.43)	<0.001	65.6	Random-effect
Children	31	4377/7321	1.42 (1.23–1.64)	<0.001	64.4	Random-effect
Adults and Children	25	2688/5205	1.10 (0.89–1.37)	<0.001	76.6	Random-effect
Type of control						
HC	65	7,442/11989	1.29 (1.15–1.44)	<0.001	66.6	Random-effect
NBDC	32	5978/10282	1.29 (1.13–1.48)	<0.001	71.9	Random-effect
Matching						
Yes	23	3819/5389	1.36 (1.12–1.65)	<0.001	77.7	Random-effect
No	75	9658/17134	1.25 (1.14–1.38)	<0.001	63.7	Random-effect
Type of leukemia						
AML	33	5530/10043	1.20 (1.04–1.38)	<0.001	71.1	Random-effect
ALL	41	5082/7895	1.44 (1.25–1.65)	<0.001	66.8	Random-effect
CML	20	2079/3426	1.17 (0.93–1.46)	<0.001	71.0	Random-effect
Sensitivity analysis						
Quality score≥10						
Overall	54	9420/15146	1.18 (1.07–1.30)	<0.001	65.6	Random-effect
Ethnicity						
Indian	10	1133/1690	1.04 (0.71–1.52)	<0.001	81.6	Random-effect
Asian	11	2323/4122	1.17 (1.05–1.31)	0.870	0.0	Random-effect
Caucasian	25	5293/7774	1.17 (1.06–1.30)	0.008	45.5	Random-effect
African	5	628/683	2.01 (1.23–3.30)	0.003	75.1	Random-effect
Age group						
Adults	28	5011/7,863	1.31 (1.15–1.50)	<0.001	63.2	Random-effect
Children	10	2282/3652	1.21 (1.06–1.39)	0.196	27.0	Random-effect
Adults and Children	14	1892/3455	0.90 (0.71–1.14)	<0.001	74.0	Random-effect
Type of control						
HC	36	5433/7693	1.21 (1.05–1.39)	<0.001	69.4	Random-effect
NBDC	18	4114/7581	1.14 (0.99–1.30)	0.002	56.8	Random-effect
Matching						
Yes	21	3525/4887	1.26 (1.05–1.52)	<0.001	73.5	Random-effect
No	33	6023/10387	1.13 (1.02–1.27)	<0.001	57.7	Random-effect
Type of leukemia						
AML	21	4598/8072	1.12 (0.97–1.28)	<0.001	63.7	Random-effect
ALL	16	2626/3754	1.22 (1.01–1.46)	<0.001	63.7	Random-effect
CML	14	1551/2417	1.23 (0.92–1.65)	<0.001	76.3	Random-effect

HC, healthy control; NBDC, nonblood disease control; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; and CML, chronic myeloid leukemia.

TABLE 2 Meta-analysis of the association of GSTT1 polymorphism with -risk of leukemia.

Variable	n	Cases/Controls	Test of association	Test of heterogeneity		Model
			OR (95%CI)	P_h	I^2 (%)	
Overall	89	12357/20636	1.46 (1.32–1.60)	<0.001	62.5	Random-effect
Ethnicity						
Indian	14	1600/2465	1.74 (1.27–2.38)	<0.001	71.9	Random-effect
Asian	24	3265/6028	1.30 (1.16–1.46)	0.140	24.2	Random-effect
Caucasian	38	6346/9237	1.37 (1.17–1.59)	<0.001	65.0	Random-effect
African	6	662/886	2.08 (1.32–3.26)	0.011	66.5	Random-effect
Age group						
Adults	37	5811/9440	1.55 (1.32–1.82)	<0.001	69.6	Random-effect
Children	27	3521/6123	1.24 (1.09–1.43)	0.028	37.2	Random-effect
Adults and Children	20	2424/4516	1.59 (1.27–1.99)	<0.001	67.1	Random-effect
Type of control						
HC	57	6522/10286	1.45 (1.28–1.66)	<0.001	63.7	Random-effect
NBDC	31	5778/10105	1.46 (1.26–1.69)	<0.001	62.7	Random-effect
Matching						
Yes	23	8272/14543	1.80 (1.44–2.24)	<0.001	74.8	Random-effect
No	66	4085/6093	1.35 (1.22–1.49)	<0.001	51.7	Random-effect
Type of leukemia						
AML	30	4851/9092	1.41 (1.19–1.66)	<0.001	67.7	Random-effect
ALL	37	4665/7,215	1.33 (1.16–1.53)	<0.001	53.0	Random-effect
CML	19	2068/3298	1.88 (1.47–2.41)	<0.001	64.5	Random-effect
Sensitivity analysis						
Quality score≥10						
Overall	52	8710/14300	1.52 (1.34–1.72)	<0.001	66.9	Random-effect
Ethnicity						
Indian	11	1225/1840	1.53 (1.08–2.17)	<0.001	69.6	Random-effect
Asian	11	2323/4122	1.15 (1.01–1.31)	0.239	21.5	Random-effect
Caucasian	22	4363/6650	1.64 (1.37–1.96)	<0.001	64.5	Random-effect
African	5	628/683	2.12 (1.26–3.58)	0.007	71.9	Random-effect
Age group						
Adults	28	5011/7,863	1.58 (1.33–1.89)	<0.001	71.3	Random-effect
Children	8	1552/2705	1.35 (1.00–1.82)	0.005	65.1	Random-effect
Adults and Children	14	1784/3428	1.45 (1.14–1.83)	0.001	61.5	Random-effect
Type of control						
HC	34	4704/6746	1.56 (1.31–1.86)	<0.001	71.0	Random-effect
NBDC	18	4006/7,554	1.45 (1.23–1.72)	0.002	56.7	Random-effect
Matching						
Yes	21	3525/4887	1.73 (1.37–2.17)	<0.001	73.6	Random-effect
No	31	5185/9413	1.41 (1.23–1.62)	<0.001	59.2	Random-effect
Type of leukemia						
AML	19	3937/7,249	1.35 (1.12–1.63)	<0.001	68.3	Random-effect
ALL	16	2449/3603	1.49 (1.19–1.88)	<0.001	64.2	Random-effect
CML	14	1551/2417	1.93 (1.44–2.59)	<0.001	64.9	Random-effect

HC, healthy control; NBDC, nonblood disease control; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; and CML, chronic myeloid leukemia.

TABLE 3 Meta-analysis of the association of GSTP1 polymorphism with risk of leukemia.

Variable	n (Cases/Controls)	Val/Val vs. Ile/Ile		Ile/Val vs. Ile/Ile		Val/Val vs. Ile/Ile + Ile/Val		Val/Val + Ile/Val vs. Ile/Ile		Val vs. Ile	
		OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)
Overall	34(5391/8729)	1.77 (1.40–2.24)	0.000/59.8	1.24 (1.08–1.43)	0.000/67.7	1.59 (1.29–1.95)	0.100/50.9	1.32 (1.15–1.53)	0.000/72.6	1.31 (1.16–1.47)	0.000/75.0
Ethnicity											
Indian	10(1392/2094)	3.01 (1.60–5.66)	0.000/76.8	1.28 (1.08–1.53)	0.167/30.3	2.65 (1.47–4.79)	0.000/74.8	1.45 (1.17–1.80)	0.013/57.2	1.47 (1.19–1.80)	0.000/72.1
Asian	10(1895/3338)	1.27 (0.98–1.66)	0.381/6.5	1.25 (0.91–1.72)	0.000/78.8	1.22(0.96–1.55)	0.799/0.0	1.30 (0.96–1.76)	0.000/80.1	1.26 (1.00–1.60)	0.000/78.1
Caucasian	12(1791/2976)	1.49 (1.10–2.01)	0.073/40.2	1.28 (0.98–1.68)	0.000/73.8	1.31 (1.04–1.65)	0.294/15.3	1.32 (1.02–1.72)	0.000/75.6	1.28 (1.05–1.55)	0.000/74.0
Age group											
Adults	14(1392/2094)	1.39 (1.06–1.82)	0.102/34.1	1.17 (0.95–1.43)	0.000/67.3	1.27 (1.01–1.61)	0.233/20.2	1.20 (0.99–1.46)	0.000/68.1	1.18 (1.02–1.37)	0.000/64.6
Children	8(1392/2094)	1.68 (1.10–2.58)	0.115/39.6	1.14 (0.88–1.46)	0.038/52.8	1.60 (1.11–2.32)	0.223/25.8	1.23 (0.94–1.60)	0.012/61.3	1.26 (1.00–1.58)	0.004/66.3
Adults and Children	9(1392/2094)	3.25 (1.61–6.53)	0.000/76.8	1.64 (1.16–2.31)	0.000/73.7	2.65 (1.41–5.02)	0.000/72.9	1.82 (1.29–2.57)	0.000/77.3	1.72 (1.29–2.30)	0.000/80.4
Type of control											
HC	21(2699/3569)	2.38 (1.66–3.41)	0.000/61.2	1.27 (1.05–1.54)	0.000/65.2	2.12 (1.53–2.94)	0.001/55.3	1.39 (1.15–1.69)	0.000/69.6	1.40 (1.19–1.63)	0.000/71.5
NBDC	13 (2692/5160)	1.19 (0.99–1.44)	0.395/5.1	1.19 (0.96–1.48)	0.000/71.0	1.14 (0.96–1.36)	0.836/0.0	1.22 (0.99–1.50)	0.000/73.9	1.17 (1.00–1.38)	0.000/72.9
Matching											
Yes	14 (2510/3287)	1.37 (1.07–1.76)	0.203/23.1	1.07 (0.95–1.20)	0.665/0.0	1.30 (1.03–1.64)	0.244/19.2	1.12 (1.00–1.24)	0.594/0.0	1.13 (1.03–1.24)	0.384/6.1
No	20(2881/5442)	2.13 (1.49–3.06)	0.000/69.0	1.36 (1.08–1.71)	0.000/78.4	1.86 (1.36–2.54)	0.000/60.6	1.47 (1.17–1.86)	0.000/81.7	1.44 (1.19–1.74)	0.000/82.8
Type of leukemia											
AML	13 (2225/4667)	1.57 (1.10–2.24)	0.000/66.4	1.37 (1.02–1.84)	0.000/83.1	1.37 (1.01–1.84)	0.008/55.1	1.42 (1.06–1.89)	0.000/84.8	1.34 (1.07–1.68)	0.000/85.1
ALL	12 (1540/2445)	1.90 (1.28–2.81)	0.018/52.1	1.15 (0.96–1.38)	0.103/36.0	1.77 (1.25–2.53)	0.051/44.0	1.26 (1.03–1.53)	0.020/51.4	1.29 (1.08–1.53)	0.003/61.0
CML	6 (926/810)	1.29 (1.08–1.53)	0.009/67.3	1.13 (0.84–1.53)	0.068/51.3	2.13 (1.08–4.24)	0.028/60.1	1.23 (0.86–1.76)	0.007/68.7	1.27 (0.92–1.74)	0.001/75.3
Sensitivity analysis											
HWE											
Overall	24 (3781/6111)	1.58 (1.27–1.95)	0.118/26.3	1.18 (1.02–1.37)	0.000/59.0	1.45 (1.21–1.74)	0.361/7.3	1.25 (1.07–1.45)	0.000/64.1	1.24 (1.10–1.40)	0.000/64.4
Ethnicity											
Indian	7 (842/1350)	1.83 (1.11–3.03)	0.033/56.3	1.24 (1.01–1.51)	0.333/12.7	1.67 (1.05–2.64)	0.055/51.3	1.34 (1.06–1.69)	0.110/42.1	1.33 (1.06–1.66)	0.017/61.2
Asian	6 (1376/2603)	1.14 (0.79–1.64)	0.728/0.0	1.04 (0.83–1.30)	0.129/41.4	1.15 (0.80–1.64)	0.799/0.0	1.06 (0.85–1.32)	0.113/43.9	1.06 (0.89–1.26)	0.165/36.3
Caucasian	9 (1250/1837)	1.70 (1.23–2.34)	0.257/21.0	1.30 (0.94–1.80)	0.000/75.8	1.50 (1.14–1.95)	0.524/0.0	1.39 (1.01–1.90)	0.000/76.4	1.36 (1.08–1.70)	0.000/72.8
Age group											
Adults	10 (1835/3530)	1.39 (1.07–1.81)	0.493/0.0	1.20 (0.93–1.56)	0.000/72.1	1.31 (1.02–1.69)	0.717/0.0	1.24 (0.97–1.60)	0.000/72.8	1.22 (1.01–1.46)	0.001/67.7
Children	8 (1124/1633)	1.68 (1.10–2.58)	0.115/39.6	1.14 (0.88–1.46)	0.038/52.8	1.60 (1.11–2.32)	0.223/25.8	1.23 (0.94–1.60)	0.012/61.3	1.26 (1.00–1.58)	0.004/66.3
Adults and Children	5 (622/848)	1.81 (0.95–3.44)	0.038/60.6	1.31 (0.99–1.72)	0.195/33.9	1.59 (0.91–2.79)	0.084/51.3	1.39 (1.01–1.92)	0.054/57.0	1.34 (1.00–1.80)	0.013/68.2
Type of control											

(Continued on following page)

10.3389/fgene.2022.976673

TABLE 3 (*Continued*) Meta-analysis of the association of GSTP1 polymorphism with risk of leukemia.

Variable	n (Cases/Controls)	Val/Val vs. Ile/Ile		Ile/Val vs. Ile/Ile		Val/Val vs. Ile/Ile + Ile/Val		Val/Val + Ile/Val vs. Ile/Ile		Val vs. Ile	
		OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)
HC	17 (2099/2775)	1.86 (1.38–2.50)	0.083/34.2	1.21 (0.98–1.50)	0.000/65.8	1.71 (1.33–2.21)	0.266/16.0	1.31 (1.05–1.62)	0.000/68.9	1.31 (1.11–1.55)	0.000/68.1
NBDC	7 (1682/3336)	1.21 (0.93–1.59)	0.930/0.0	1.08 (0.93–1.26)	0.289/18.4	1.16 (0.89–1.50)	0.985/0.0	1.11 (0.95–1.29)	0.223/27.0	1.09 (0.97–1.22)	0.296/17.6
Matching											
Yes	9 (1546/1696)	1.51 (1.11–2.05)	0.787/0.0	1.12 (0.96–1.30)	0.929/0.0	1.43 (1.07–1.93)	0.742/0.0	1.17 (1.01–1.35)	0.908/0.0	1.18 (1.05–1.32)	0.796/0.0
No	15 (2235/4415)	1.66 (1.21–2.27)	0.023/47.1	1.24 (0.98–1.56)	0.000/73.6	1.50 (1.15–1.94)	0.142/28.8	1.31 (1.03–1.66)	0.000/76.9	1.29 (1.07–1.55)	0.000/76.6
Type of leukemia											
AML	7 (1141/2762)	1.18 (0.86–1.62)	0.424/0.0	1.17 (0.80–1.69)	0.000/80.0	1.13 (0.83–1.54)	0.717/0.0	1.18 (0.83–1.69)	0.000/80.1	1.14 (0.89–1.48)	0.000/75.5
ALL	10 (1345/2047)	1.60 (1.15–2.22)	0.197/26.8	1.10 (0.91–1.31)	0.180/28.8	1.53 (1.14–2.06)	0.280/17.8	1.18 (0.97–1.43)	0.075/42.4	1.21 (1.03–1.43)	0.025/52.6
Quality score≥12											
Overall	18 (3430/5975)	1.62 (1.25–2.11)	0.018/45.7	1.16 (0.99–1.36)	0.001/58.6	1.49 (1.18–1.89)	0.067/35.6	1.23 (1.05–1.44)	0.000/64.0	1.23 (1.09–1.40)	0.000/65.5
Ethnicity											
Caucasian	6 (1064/2009)	1.54 (1.09–2.17)	0.220/28.7	1.41 (0.98–2.04)	0.001/76.9	1.28 (1.01–1.64)	0.489/0.0	1.48 (1.04–2.10)	0.001/77.2	1.38 (1.08–1.77)	0.002/73.1
Indian	7 (1010/1448)	2.15 (1.22–3.76)	0.013/62.6	1.17(0.95–1.43)	0.231/25.9	1.97 (1.16–3.35)	0.019/60.3	1.29 (1.03–1.63)	0.079/47.1	1.32 (1.07–1.64)	0.016/61.6
Age group											
Adults	11 (1392/2094)	1.42 (1.07–1.88)	0.134/33.1	1.19 (0.97–1.46)	0.001/67.4	1.29 (1.02–1.64)	0.279/17.3	1.24 (1.01–1.51)	0.000/68.9	1.21 (1.04–1.41)	0.001/65.5
Type of control											
HC	13 (1856/2372)	1.94 (1.37–2.76)	0.044/44.2	1.21 (0.97–1.50)	0.002/62.1	1.77 (1.29–2.44)	0.088/36.9	1.30 (1.05–1.62)	0.001/64.8	1.31 (1.10–1.55)	0.001/64.0
NBDC	5 (1574/3603)	1.14 (0.89–1.46)	0.735/0.0	1.05 (0.88–1.27)	0.152/40.4	1.12 (0.88–1.42)	0.901/0.0	1.08 (0.90–1.30)	0.112/46.6	1.07 (0.93–1.23)	0.150/40.8
Matching											
Yes	11 (2226/2966)	1.45 (1.11–1.90)	0.283/16.9	1.06 (0.94–1.20)	0.942/0.0	1.41 (1.08–1.83)	0.263/18.9	1.12 (0.99–1.25)	0.919/0.0	1.14 (1.04–1.25)	0.762/0.0
No	7 (1204/3009)	1.81 (1.07–3.06)	0.006/67.1	1.34 (0.90–2.00)	0.000/83.3	1.58 (1.02–2.46)	0.034/55.9	1.43 (0.96–2.15)	0.000/85.6	1.37 (1.00–1.89)	0.000/85.5
Type of leukemia											
AML	7 (1461/3684)	1.11 (0.88–1.41)	0.470/0.0	1.21 (0.88–1.65)	0.000/78.4	1.08 (0.86–1.35)	0.880/0.0	1.21 (0.89–1.64)	0.000/79.2	1.14 (0.92–1.43)	0.000/75.5
CML	5 (855/743)	3.17 (1.89–5.32)	0.308/16.7	1.18 (0.85–1.64)	0.052/57.5	2.80 (1.79–4.39)	0.489/0.0	1.35 (0.94–1.94)	0.014/68.0	1.41 (1.05–1.89)	0.013/68.3
HWE and Quality score≥12											
Overall	16 (2750/4705)	1.63 (1.24–2.13)	0.081/35.2	1.20 (1.00–1.44)	0.001/61.8	1.49 (1.18–1.88)	0.223/20.2	1.27 (1.06–1.53)	0.000/66.8	1.26 (1.08–1.46)	0.000/67.7
Ethnicity											
Indian	6 (750/1200)	1.91 (1.07–3.40)	0.020/62.6	1.23 (0.97–1.55)	0.236/26.5	1.74 (1.03–2.96)	0.037/57.8	1.34 (1.02–1.76)	0.066/51.7	1.34 (1.04–1.74)	0.009/67.4
Caucasian	5 (644/987)	1.87 (1.28–2.74)	0.649/0.0	1.55 (1.02–2.34)	0.007/71.9	1.59 (1.11–2.30)	0.734/0.0	1.63 (1.12–2.37)	0.012/68.7	1.50 (1.17–1.91)	0.057/56.4
Age group											
Adults	9 (1735/3430)	1.38 (1.05–1.82)	0.408/3.2	1.27 (0.97–1.66)	0.000/72.2	1.28 (0.98–1.66)	0.705/0.0	1.30 (1.00–1.70)	0.000/73.9	1.24 (1.02–1.52)	0.001/70.4

(Continued on following page)

10.3389/fgene.2022.976673

TABLE 3 (*Continued*) Meta-analysis of the association of GSTP1 polymorphism with risk of leukemia.

Variable	n (Cases/Controls)	Val/Val vs. Ile/Ile		Ile/Val vs. Ile/Ile		Val/Val vs. Ile/Ile + Ile/Val		Val/Val + Ile/Val vs. Ile/Ile		Val vs. Ile	
		OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)
Type of control											
HC	12 (1596/2124)	1.83 (1.29–2.58)	0.072/40.3	1.23 (0.97–1.57)	0.002/63.5	1.66 (1.23–2.25)	0.169/28.2	1.33 (1.05–1.68)	0.001/66.7	1.32 (1.09–1.59)	0.001/66.8
Matching											
Yes	9 (1546/1696)	1.51 (1.11–2.05)	0.787/0.0	1.12 (0.96–1.30)	0.929/0.0	1.43 (1.07–1.93)	0.742/0.0	1.17 (1.01–1.35)	0.908/0.0	1.18 (1.05–1.32)	0.796/0.0
No	7 (1204/3009)	1.81 (1.07–3.06)	0.006/67.1	1.34 (0.90–2.00)	0.000/83.3	1.58 (1.02–2.46)	0.034/55.9	1.43 (0.96–2.15)	0.000/85.6	1.37 (1.00–1.89)	0.000/85.5
Type of leukemia											
AML	6 (1041/2662)	1.16 (0.82–1.64)	0.355/9.5	1.27 (0.84–1.90)	0.000/81.5	1.08 (0.78–1.48)	0.792/0.0	1.26 (0.85–1.88)	0.000/82.4	1.18 (0.88–1.57)	0.000/79.3

HC, healthy control; NBDC, nonblood disease control; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; and CML, chronic myeloid leukemia.

analysis of the data. The combination of GSTM1 present/null and GSTP1 IIe105Val was also used for the six genetic models, model 1: M1 null/P1 IIe/IIe vs. M1 present/P1 IIe/IIe, model 2: M1 present/P1 Val* vs. M1 present/P1 IIe/IIe, model 3: (M1 null/P1 IIe/IIe + M1 present/P1 Val*) vs. M1 present/P1 IIe/IIe, model 4: M1 null/P1 Val* vs. M1 present/P1 IIe/IIe, model 5: All risk genotypes vs. M1 present/P1 IIe/IIe, and model 6: M1 null/P1 Val* vs. (M1 present/P1 IIe/IIe + M1 null/P1 IIe/IIe + M1 Present/P1 Val*). There were six genetic models used in the combination of GSTT1 present/null and GSTP1 IIe105Val: model 1: T1 null/P1 IIe/IIe vs. T1 present/P1 IIe/IIe, model 2: T1 present/P1 Val* vs. T1 present/P1 IIe/IIe, model 3: = (T1 null/P1 IIe/IIe + T1 present/P1 Val*) vs. T1 present/P1 IIe/IIe, model 4: T1 null/P1 Val* vs. T1 present/P1 IIe/IIe, model 5: All risk genotypes vs. T1 present/P1 IIe/IIe, and model 6: T1 null/P1 Val* vs. (T1 present/P1 IIe/IIe + T1 null/P1 IIe/IIe + T1 Present/P1 Val*). In the combination of GSTM1 present/null, GSTT1 present/null, and GSTP1 IIe105Val, the following genetic models were employed: model 1: M1 null/T1 present/P1 IIe/IIe vs. M1 present/T1 present/P1 IIe/IIe, model 2: M1 present/T1 null/P1 IIe/IIe vs. M1 present/T1 present/P1 IIe/IIe, model 3: M1 present/T1 present/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe, model 4: all one high-risk genotype vs. M1 present/T1 present/P1 IIe/IIe, model 5: M1 null/T1 null/P1 IIe/IIe vs. M1 present/T1 present/P1 IIe/IIe, model 6: M1 null/T1 present/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe, model 7: M1 present/T1 null/P1 Val1 vs. M1 present/T1 present/P1 IIe/IIe, model 8: all two high-risk genotype vs. M1 present/T1 present/P1 IIe/IIe, model 9: M1 null/T1 null/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe, and model 10: M1 null/T1 null/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe + all one high-risk genotype + all two high-risk genotypes. Moreover, a metaregression analysis was used to explore sources of heterogeneity (Baker et al., 2009). Sensitivity analysis was conducted by excluding low-quality and Hardy–Weinberg disequilibrium (HWD) in control studies. The Hardy–Weinberg equilibrium (HWE) was checked using Chi-square goodness-of-fit test, which was deemed as HWE in controls if $p \geq 0.05$. Begg's funnel plot (Begg and Mazumdar, 1994) and Egger's test (Egger et al., 1997) were carried out to verify publication bias. Furthermore, we applied the FPRP (Wacholder et al., 2004), BFDP (Wakefield, 2007), and Venice criteria (Ioannidis et al., 2008) to appraise the credibility of statistically significant associations. All statistical analyses were performed using Stata 12.0 software in the current study.

Results

Search results and study characteristics

Overall, 91 articles (Supplemental References 1–91) were eligible (Figure 1), and Supplementary Tables S1–S3 show the characteristics and scores of each study. Multiple eligible studies were included in one article. Therefore, there were 98 eligible studies (13,477 leukemia cases and 22,523 controls, Table 1) on the *GSTM1* present/null

polymorphism, 89 eligible studies (12,357 leukemia cases and 20,636 controls, Table 2) on the *GSTT1* present/null polymorphism, 34 studies (5,391 leukemia cases and 8,729 controls, Table 3) on the GSTP1 IIe105Val polymorphism, 25 studies (3,522 leukemia cases and 4,974 controls, Table 4) belonging to the combined effects of the *GSTM1* and *GSTT1* polymorphisms, six studies (737 leukemia cases and 995 controls, Table 5) describing the combined *GSTM1* and *GSTP1* effects, five studies (645 leukemia cases and 845 controls, Table 6) on the combined *GSTT1* and *GSTP1* effects, and seven studies (1,036 leukemia cases and 1,418 controls, Table 7) belonging to the combined effects of the three aforementioned polymorphisms with leukemia risk.

Quantitative synthesis

The *GSTM1* null genotype significantly added leukemia risk in the overall analysis (OR = 1.28, 95% CI: 1.17–1.40, Table 1 and Figure 2) of Asians (OR = 1.50, 95% CI: 1.29–1.73), Caucasians (OR = 1.17, 95% CI: 1.07–1.28), and Africans (OR = 1.99, 95% CI: 1.30–3.94). However, it showed that the GSTM1 null genotype did not affect leukemia risk in Indians (OR = 1.25, 95% CI: 0.89–1.77). Moreover, similar association was also found in other subgroup analyses, such as in adult leukemia, child leukemia, AML, ALL, and so on (Table 1).

The GSTT1 null genotype added leukemia risk in the overall population (OR = 1.46, 95% CI: 1.32–1.60, Table 2 and Figure 3). Moreover, an increased risk of leukemia was also found in Indians (OR = 1.74, 95% CI: 1.27–2.38), Asians (OR = 1.30, 95% CI: 1.16–1.46), Caucasians (OR = 1.37, 95% CI: 1.17–1.59), and Africans (OR = 2.08, 95% CI: 1.32–3.26) (Table 2; Figure 3). Similarly, the significantly increased risk of leukemia was also observed in adult leukemia, child leukemia, AML, ALL, and CML, and so on (Table 2).

The GSTP1 IIe105Val polymorphism yielded a significantly increased leukemia risk in overall population (Val/Val vs. IIe/IIe: OR = 1.77, 95% CI = 1.40–2.24; IIe/Val vs. IIe/IIe: OR = 1.24, 95% CI = 1.08–1.43; Val/Val vs. IIe/IIe + IIe/Val: OR = 1.59, 95% CI = 1.29–1.95; Val/Val + IIe/Val vs. IIe/IIe: OR = 1.32, 95% CI = 1.15–1.53; and Val vs. IIe: OR = 1.31, 95% CI = 1.16–1.47, Table 3 and Figure 4). Moreover, the GSTP1 IIe105Val polymorphism was associated with increased leukemia risk in Indians (Val/Val vs. IIe/IIe: OR = 3.01, 95% CI = 1.60–5.66; IIe/Val vs. IIe/IIe: OR = 1.28, 95% CI = 1.08–1.53; Val/Val vs. IIe/IIe + IIe/Val: OR = 2.65, 95% CI = 1.47–4.79; Val/Val + IIe/Val vs. IIe/IIe: OR = 1.45, 95% CI = 1.17–1.80; and Val vs. IIe: OR = 1.47, 95% CI = 1.19–1.80) and in Caucasians (Val/Val vs. IIe/IIe: OR = 1.49, 95% CI = 1.10–2.01; Val/Val vs. IIe/IIe + IIe/Val: OR = 1.31, 95% CI = 1.04–1.65; Val/Val + IIe/Val vs. IIe/IIe: OR = 1.32, 95% CI = 1.02–1.72; and Val vs. IIe: OR = 1.28, 95% CI = 1.05–1.55). Similarly, the significantly increased risk of leukemia was also observed in adult leukemia, child leukemia, AML, ALL, CML, etc. (Table 3).

Combined *GSTM1* and *GSTT1* null genotypes were found to significantly increase leukemia risk in the overall analysis (M1 present/T1 null vs. M1 present/T1 present: OR = 1.66, 95% CI = 1.37–2.00; M1 null/T1 null vs. M1 present/T1 present: OR = 2.44, 95% CI = 1.86–3.21; all one risk genotypes vs. M1 present/T1 present: OR = 1.29, 95% CI = 1.11–1.50; all risk genotypes vs. M1 present/T1 present: OR = 1.44, 95% CI = 1.25–1.66; and M1 null/T1 null vs. M1 present/T1 present + M1 present/T1 null + M1 null/T1 present: OR = 2.16, 95% CI = 1.65–2.81; Table 4 and Figure 5). Moreover, there was a significantly increased leukemia risk in Indians (M1 present/T1 null vs. M1 present/T1 present: OR = 1.92, 95% CI = 1.18–3.12; M1 null/T1 null vs. M1 present/T1 present: OR = 3.16, 95% CI = 1.90–5.25; M1 null/T1 null vs. M1 present/T1 present + M1 present/T1 null + M1 null/T1 present: OR = 2.83, 95% CI = 1.73–4.64), Asians (M1 present/T1 null vs. M1 present/T1 present: OR = 1.43, 95% CI = 1.04–1.97; M1 null/T1 null vs. M1 present/T1 present: OR = 2.47, 95% CI = 1.55–3.95; all one risk genotypes vs. M1 present/T1 present: OR = 1.35, 95% CI = 1.02–1.80; all risk genotypes vs. M1 present/T1 present: OR = 1.57, 95% CI = 1.20–2.05; M1 null/T1 null vs. M1 present/T1 present + M1 present/T1 null + M1 null/T1 present: OR = 2.05, 95% CI = 1.40–3.00), and Caucasians (M1 present/T1 null vs. M1 present/T1 present: OR = 1.65, 95% CI = 1.14–2.39; M1 null/T1 null vs. M1 present/T1 present: OR = 1.98, 95% CI = 1.16–3.37; all one risk genotypes vs. M1 present/T1 present: OR = 1.30, 95% CI = 1.05–1.60; all risk genotypes vs. M1 present/T1 present: OR = 1.37, 95% CI = 1.17–1.61). Similar results were found in adult leukemia, AML, ALL, CML, and so on (Table 4).

An increased risk of leukemia was yielded on the combined GSTM1 and GSTP1 polymorphisms (M1 null/P1 Val* vs. M1 present/P1 IIe/IIe: OR = 1.95, 95% CI = 1.35–2.80; M1 null/P1 Val* vs. M1 present/P1 IIe/IIe + M1 null/P1 IIe/IIe + M1 Present/P1 Val*: OR = 1.95, 95% CI = 1.37–2.77; Table 5 and Figure 6) in overall analysis. Moreover, increased leukemia risk was also demonstrated in Indians (M1 null/P1 Val* vs. M1 present/P1 IIe/IIe: OR = 1.72, 95% CI = 1.10–2.70, M1 null/P1 Val* vs. M1 present/P1 IIe/IIe + M1 null/P1 IIe/IIe + M1 Present/P1 Val*: OR = 1.65, 95% CI = 1.14–2.40). Furthermore, a similar connection was also found in ALL, CML, and so on (Table 5).

On combining GSTT1 and GSTP1 polymorphisms, there was a strong connection with leukemia risk in the overall analysis ((T1 null/P1 IIe/IIe + T1 present/P1 Val*) vs. T1 present/P1 IIe/IIe: OR = 1.50, 95% CI = 1.04–2.15; T1 null/P1 Val* vs. T1 present/P1 IIe/IIe: OR = 4.24, 95% CI = 2.49–7.24; all risk genotypes vs. T1 present/P1 IIe/IIe: OR = 1.70, 95% CI = 1.30–2.22; and T1 null/P1 Val* vs. (T1 present/P1 IIe/IIe + T1 null/P1 IIe/IIe + T1 Present/P1 Val*): OR = 3.31, 95% CI = 1.85–5.91) and increased risk of leukemia among Indians

TABLE 4 Meta-analysis of the combined effects of GSTM1 present/null and GSTT1 present/null on leukemia risk.

Variable	N (Case/ Control)	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
		OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)
Overall	25 (3522/4974)	1.66 (1.37–2.00)	0.077/30.3	1.11 (0.93–1.33)	0.000/60.4	2.44 (1.86–3.21)	0.002/51.2	1.29 (1.11–1.50)	0.001/52.2	1.44 (1.25–1.66)	0.002/ 51.5	2.16 (1.65–2.81)	0.000/55.4
Ethnicity													
Indian	5 (555/829)	1.92 (1.18–3.12)	0.075/52.9	0.87 (0.54–1.40)	0.017/66.6	3.16 (1.90–5.25)	0.519/0.0	1.18 (0.76–1.85)	0.006/72.4	1.32 (0.83–2.10)	0.002/ 76.1	2.83 (1.73–4.64)	0.759/0.0
Asian	5 (1000/1148)	1.43 (1.04–1.97)	0.274/22.1	1.34 (0.99–1.81)	0.146/41.4	2.47 (1.55–3.95)	0.051/57.5	1.35 (1.02–1.80)	0.120/45.3	1.57 (1.20–2.05)	0.129/ 44.0	2.05 (1.40–3.00)	0.090/50.3
Caucasian	10 (1506/1916)	1.65 (1.14–2.39)	0.087/40.6	1.15 (0.92–1.43)	0.101/38.6	1.98 (1.16–3.37)	0.005/61.5	1.30 (1.05–1.60)	0.082/41.3	1.37 (1.17–1.61)	0.317/ 13.7	1.71 (0.94–3.09)	0.000/71.8
Age group													
Adults	15 (2424/2884)	1.44 (1.18–1.76)	0.600/0.0	1.27(1.04–1.54)	0.013/50.7	2.51 (1.71–3.68)	0.001/60.0	1.34 (1.15–1.57)	0.087/35.3	1.50 (1.29–1.74)	0.104/ 33.0	2.26 (1.53–3.33)	0.000/65.8
Adults and children	5 (488/1112)	1.63 (0.87–3.07)	0.014/68.0	0.75 (0.50–1.12)	0.122/45.0	2.05 (0.96–4.37)	0.063/55.1	0.99 (0.66–1.49)	0.044/59.2	1.10 (0.71–1.71)	0.016/ 67,2	1.94 (1.04–4.36)	0.131/43.6
Type of control													
HC	15(1693/2058)	1.73 (1.31–2.30)	0.060/39.2	1.02 (0.76–1.38)	0.000/69.8	2.59 (1.71–3.93)	0.012/51.1	1.27 (1.00–1.62)	0.000/63.9	1.45 (1.16–1.80)	0.001/ 60.3	2.33 (1.52–3.58)	0.002/59.7
NBDC	9 (1772/2671)	1.60 (1.22–2.10)	0.215/25.7	1.29 (1.11–1.50)	0.466/0.0	2.31 (1.56–3.43)	0.021/55.7	1.36 (1.18–1.57)	0.421/1.6	1.49 (1.25–1.78)	0.147/ 33.9	1.86 (1.33–2.61)	0.059/46.7
Matching													
Yes	11 (1958/2382)	1.60 (1.29–1.99)	0.493/0.0	1.13 (0.85–1.50)	0.000/70.1	2.57 (1.61–4.12)	0.002/64.2	1.31 (1.09–1.58)	0.075/41.0	1.46 (1.24–1.73)	0.129/ 33.8	2.33 (1.44–3.76)	0.000/69.6
No	14 (1564/2592)	1.67 (1.24–2.27)	0.023/48.0	1.09 (0.86–1.40)	0.013/51.6	2.38 (1.70–3.33)	0.076/37.6	1.28 (1.00–1.63)	0.002/60.7	1.43 (1.13–1.80)	0.001/ 62.1	2.07 (1.52–2.81)	0.084/36.5
Type of leukemia													
AML	6 (1176/1859)	1.47 (0.96–2.26)	0.084/48.5	1.24 (0.97–1.58)	0.202/31.2	2.15 (1.35–3.43)	0.049/55.1	1.29 (0.99–1.69)	0.088/47.8	1.41 (1.09–1.82)	0.095/ 46.7	1.85 (1.22–2.80)	0.069/51.1
ALL	7 (670/1060)	2.15 (1.43–3.23)	0.125/39.9	1.19 (0.77–1.85)	0.008/65.2	2.79 (1.47–5.30)	0.052/52.0	1.52 (1.13–2.05)	0.094/44.5	1.66 (1.25–2.20)	0.106/ 42.7	2.23 (1.20–4.14)	0.036/55.4
CML	11 (1234/1613)	1.54 (1.18–2.01)	0.375/7.2	1.01 (0.72–1.42)	0.000/69.7	2.58 (1.57–4.24)	0.024/51.4	1.19 (0.92–1.56)	0.004/61.4	1.37 (1.06–1.77)	0.003/ 62.5	2.41 (1.45–4.00)	0.012/55.8
Sensitivity analysis													
Quality score													
≥10													
Overall	21 (3105/4266)		0.132/26.3	1.12 (0.92–1.37)	0.000/61.6		0.001/55.3		0.007/48.3				0.001/57.8

(Continued on following page)

10.3389/fgene.2022.976673

TABLE 4 (*Continued*) Meta-analysis of the combined effects of GSTM1 present/null and GSTT1 present/null on leukemia risk.

Variable	N (Case/ Control)	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
		OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)	OR (95%CI)	P_h/ I^2(%)
		1.56 (1.28–1.91)				2.41 (1.76–3.29)		1.27 (1.09–1.48)		1.42 (1.22–1.65)	0.003/ 51.7	2.17 (1.61–2.94)	
Ethnicity													
Indian	5 (555/829)	1.92 (1.18–3.12)	0.075/52.9	0.87 (0.54–1.40)	0.017/66.6	3.16 (1.90–5.25)	0.519/0.0	1.18 (0.76–1.85)	0.006/72.4	1.32 (0.83–2.10)	0.002/ 76.1	2.83 (1.73–4.64)	0.759/0.0
Caucasian	8 (1121/1352)	1.38 (0.96–1.98)	0.260/21.4	1.21 (0.95–1.53)	0.100/41.7	1.80 (0.96–3.37)	0.008/63.6	1.28 (1.04–1.57)	0.160/33.6	1.34 (1.13–1.58)	0.341/ 11.5	1.63 (0.83–3.21)	0.001/71.5
Age group													
Adults	14 (2317/2754)	1.43 (1.16–1.76)	0.527/0.0	1.31 (1.08–1.60)	0.020/48.8	2.40 (1.61–3.58)	0.002/59.9	1.37 (1.16–1.61)	0.089/35.8	1.51 (1.29–1.77)	0.078/ 37.4	2.13 (1.43–3.18)	0.000/64.7
Adults and children	5 (488/1112)	1.63 (0.87–3.07)	0.014/68.0	0.75 (0.50–1.12)	0.122/45.0	2.05 (0.96–4.37)	0.063/55.1	0.99 (0.66–1.49)	0.044/59.2	1.10 (0.71–1.71)	0.016/ 67,2	1.94 (1.04–4.36)	0.131/43.6
Type of Control													
HC	13 (1539/1826)	1.63 (1.21–2.18)	0.103/34.9	1.06 (0.76–1.47)	0.000/73.5	2.56 (1.60–4.10)	0.009/55.0	1.25 (0.97–1.61)	0.001/64.6	1.41 (1.11–1.80)	0.001/ 63.6	2.39 (1.50–3.80)	0.004/58.7
NBDC	8 (1566/2440)	1.47 (1.12–1.94)	0.306/15.7	1.24 (1.06–1.45)	0.693/0.0	2.17 (1.43–3.29)	0.030/54.9	1.30 (1.12–1.50)	0.724/0.0	1.41 (1.19–1.66)	0.294/ 17.2	1.85 (1.27–2.69)	0.045/51.3
Matching													
Yes	11 (1958/2382)	1.60 (1.29–1.99)	0.493/0.0	1.13 (0.85–1.50)	0.000/70.1	2.57 (1.61–4.12)	0.002/64.2	1.31 (1.09–1.58)	0.078/41.0	1.46 (1.24–1.73)	0.129/ 33.8	2.33 (1.44–3.76)	0.000/69.6
No	10 (1147/1884)	1.51 (1.04–2.19)	0.044/48.0	1.11 (0.84–1.47)	0.032/50.8	2.26 (1.45–3.53)	0.051/46.7	1.22 (0.93–1.61)	0.012/57.2	1.37 (1.03–1.82)	0.002/ 65.1	2.04 (1.40–2.98)	0.107/37.7
Type of leukemia													
ALL	6 (623/958)	1.92 (1.28–2.86)	0.213/29.6	1.22 (0.76–1.96)	0.004/70.7	3.10 (1.48–6.49)	0.036/58.1	1.43 (1.07–1.91)	0.144/39.3	1.59 (1.18–2.14)	0.100/ 45.8	2.66 (1.38–5.15)	0.053/54.1
CML	10 (1127/1483)	1.55 (1.15–2.09)	0.292/16.4	1.04 (0.73–1.51)	0.000/71.8	2.39 (1.37–4.16)	0.020/54.3	1.22 (0.91–1.64)	0.003/64.6	1.38 (1.03–1.83)	0.002/ 66.2	2.21 (1.26–3.87)	0.012/57.7

Model 1, M1 present/T1 null vs. M1 present/T1 present; Model 2, M1 null/T1 present vs. M1 present/T1 present; Model 3, M1 null/T1 null vs. M1 present/T1 present; Model 4, all one risk genotypes vs. M1 present/T1 present; Model 5, all risk genotypes vs. M1 present/T1 present; Model 6, M1 null/T1 null vs. M1 present/T1 present + M1 present/T1 null + M1 null/T1 present; HC, healthy control; NBDC, nonblood disease control; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; and CML, chronic myeloid leukemia.

10.3389/fgene.2022.976673

TABLE 5 Meta-analysis of the combined effects of GSTM1 present/null and GSTP1 Ile105Val on leukemia risk.

Variable	Sample size	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
		OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)
Overall	6 (737/995)	0.83 (0.55–1.26)	0.038/57.5	1.16 (0.74–1.84)	0.017/63.9	1.02 (0.74–1.39)	0.063/52.2	1.95 (1.35–2.80)	0.272/21.5	1.19 (0.90–1.58)	0.100/45.9	1.95 (1.37–2.77)	0.208/30.4
Ethnicity													
Indian	4 (492/750)	0.75 (0.39–1.45)	0.015/71.4	1.26 (0.74–2.13)	0.021/69.2	1.05 (0.65–1.68)	0.018/70.2	1.72 (1.10–2.70)	0.211/33.5	1.18 (0.77–1.79)	0.030/66.4	1.65 (1.14–2.40)	0.292/19.6
Type of control													
HC	5 (645/845)	0.74 (0.49–1.12)	0.081/51.9	1.14 (0.65–2.02)	0.008/71.1	0.97 (0.68–1.38)	0.052/57.5	1.82 (1.21–2.74)	0.249/25.9	1.13 (0.83–1.54)	0.097/49.1	1.88 (1.23–2.89)	0.143/41.8
Matching													
Yes	3 (395/395)	0.72 (0.37–1.41)	0.033/70.6	0.82 (0.43–1.57)	0.147/47.8	0.78 (0.51–1.21)	0.142/48.7	1.89 (0.90–3.96)	0.113/54.1	0.99 (0.63–1.56)	0.087/59.0	2.20 (1.25–3.89)	0.204/37.1
No	3 (342/600)	0.97 (0.53–1.76)	0.123/52.3	1.53 (0.91–2.56)	0.097/57.2	1.31 (0.97–1.77)	0.581/0.0	2.07 (1.34–3.20)	0.413/0.0	1.44 (1.08–1.92)	0.771/0.0	1.76 (1.05–2.96)	0.178/42.1
Type of leukemia													
ALL	3 (342/600)	0.83 (0.34–2.03)	0.008/79.5	0.98 (0.70–1.38)	0.403/0.0	0.91 (0.53–1.57)	0.038/69.3	1.86 (1.01–3.43)	0.125/51.9	1.08 (0.63–1.84)	0.028/72.0	1.92 (1.30–2.83)	0.498/0.0
CML	3 (395/395)	0.86 (0.60–1.24)	0.377/0.0	1.34 (0.51–3.48)	0.015/76.1	1.17 (0.83–1.63)	0.306/15.6	2.08 (1.27–3.40)	0.363/1.4	1.34 (1.00–1.92)	0.705/0.0	2.00 (0.90–4.46)	0.055/65.4
Sensitivity analysis													
HWE and Quality score													
≥10	6 (737/995)	0.83 (0.55–1.26)	0.038/57.5	1.16 (0.74–1.84)	0.017/63.9	1.02 (0.74–1.39)	0.063/52.2	1.95 (1.35–2.80)	0.272/21.5	1.19 (0.90–1.58)	0.100/45.9	1.95 (1.37–2.77)	0.208/30.4

Model 1, M1 null/P1 Ile/Ile vs. M1 present/P1 Ile/Ile; Model 2, M1 present/P1 Val* vs. M1 present/P1 Ile/Ile; Model 3, (M1 null/P1 Ile/Ile + M1 present/P1 Val*) vs. M1 present/P1 Ile/Ile; Model 4 = M1 null/P1 Val* vs. M1 present/P1 Ile/Ile; Model 5, All risk genotypes vs. M1 present/P1 Ile/Ile; Model 6, M1 null/P1 Val* vs. (M1 present/P1 Ile/Ile + M1 null/P1 Ile/Ile + M1 Present/P1 Val*); HC, healthy control; NBDC, nonblood disease controls; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; and CML, chronic myeloid leukemia.

10.3389/fgene.2022.976673

TABLE 6 Meta-analysis of the combined effects of GSTT1 present/null and GSTP1 Ile105Val on leukemia risk.

Variable	Sample size	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
		OR (95%CI)	P_h/I^2	OR (95%CI)	P_h/I^2	OR (95%CI)	P_h/I^2	OR (95%CI)	P_h/I^2	OR (95%CI)	P_h/I^2	OR (95%CI)	P_h/I^2
Overall	5 (645/845)	1.56 (0.76–3.19)	0.009/70.6	1.49 (0.97–2.28)	0.032/62.2	1.50 (1.04–2.15)	0.041/59.8	4.24 (2.49–7.24)	0.596/0.0	1.70 (1.30–2.22)	0.207/32.2	3.31 (1.85–5.91)	0.320/14.8
Ethnicity													
Indian	3 (400/600)	1.90 (0.99–3.66)	0.086/59.3	1.45 (0.72–2.92)	0.006/80.4	1.65 (1.05–2.59)	0.072/61.9	4.39 (2.51–7.68)	0.741/0.0	1.91 (1.45–2.50)	0.365/0.8	3.39 (1.94–5.94)	0.338/7.8
Type of control													
HC	5 (645/845)	1.56 (0.76–3.19)	0.009/70.6	1.49 (0.97–2.28)	0.032/62.2	1.50 (1.04–2.15)	0.041/59.8	4.24 (2.49–7.24)	0.596/0.0	1.70 (1.30–2.22)	0.207/32.2	3.31 (1.85–5.91)	0.320/14.8
Matching													
Yes	3 (395/395)	1.44 (0.48–4.35)	0.032/70.8	1.16 (0.65–2.08)	0.082/60.0	1.18 (0.76–1.83)	0.135/50.1	4.61 (1.64–12.97)	0.301/16.8	1.40 (1.04–1.89)	0.368/0.0	4.15 (0.78–7.37)	0.278/21.9
Type of leukemia													
CML	3 (395/395)	0.88 (0.41–1.88)	0.218/34.3	1.91 (1.35–2.68)	0.441/0.0	1.49 (0.89–2.51)	0.059/64.6	3.29 (1.37–7.89)	0.361/1.9	1.61 (1.05–2.47)	0.133/50.4	2.40 (1.21–14.26)	0.231/31.8
Sensitivity analysis													
HWE and Quality score≥10													
Overall	5 (645/845)	1.56 (0.76–3.19)	0.009/70.6	1.49 (0.97–2.28)	0.032/62.2	1.50 (1.04–2.15)	0.041/59.8	4.24 (2.49–7.24)	0.596/0.0	1.70(1.30–2.22)	0.207/32.2	3.31 (1.85–5.91)	0.320/14.8

Model 1, T1 null/P1 IIe/IIe vs. T1 present/P1 IIe/IIe; Model 2, T1 present/P1 Val* vs. T1 present/P1 IIe/IIe; Model 3, (T1 null/P1 IIe/IIe + T1 present/P1 Val*) vs. T1 present/P1 IIe/IIe; Model 4, T1 null/P1 Val* vs. T1 present/P1 IIe/IIe; Model 5, all risk genotypes vs. T1 present/P1 IIe/IIe; Model 6, T1 null/P1 Val* vs. (T1 present/P1 IIe/IIe + T1 null/P1 IIe/IIe + T1 Present/P1 Val*); HB, hospital-based studies; PB, population-based studies; HC, healthy control; NBDC, nonblood disease controls; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; and CML, chronic myeloid leukemia.

10.3389/fgene.2022.976673

TABLE 7 Meta-analysis of the combined effects of GSTM1 present/null, GSTT1 present/null, and GSTP1 Ile105Val on leukemia risk.

Variable	Sample size	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
		OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)	OR (95%CI)	P_h/I^2(%)
Overall	7 (1036/1418)	0.93 (0.72–1.21)	0.945/0.0	1.38 (0.92–2.07)	0.261/22.0	1.12 (0.86–1.47)	0.723/0.0	1.07 (0.87–1.33)	0.486/0.0	0.81 (0.23–2.92)	0.000/90.1	1.18 (0.78–1.79)	0.092/44.9	0.95 (0.57–1.61)	0.148/36.8	1.09 (0.71–1.68)	0.006/66.5	2.04 (0.89–4.70)	0.007/66.2	1.87 (0.97–3.62)	0.038/55.1
Sensitivity analysis																					
Quality score																					
>8	7 (1036/1418)	ACT	0.945/0.0	1.38 (0.92–2.07)	0.261/22.0	1.12 (0.86–1.47)	0.723/0.0	1.07 (0.87–1.33)	0.486/0.0	0.81 (0.23–2.92)	0.000/90.1	1.18 (0.78–1.79)	0.092/44.9	0.95 (0.57–1.61)	0.148/36.8	1.09 (0.71–1.68)	0.006/66.5	2.04 (0.89–4.70)	0.007/66.2	1.87 (0.97–3.62)	0.038/55.1
HWE																					
Yes	6 (603/705)	0.90 (0.68–1.18)	0.987/0.0	1.38 (0.83–2.31)	0.179/34.4	1.11 (0.83–1.48)	0.604/0.0	1.05 (0.83–1.32)	0.402/2.2	0.69 (0.16–2.94)	0.000/90.0	1.12 (0.70–1.79)	0.076/49.9	0.81 (0.45–1.44)	0.225/28.0	1.00 (0.62–1.62)	0.009/67.4	1.85 (0.69–4.96)	0.009/67.3	1.79 (0.79–4.08)	0.032/59.1

Model 1 = M1 null/T1 present/P1 IIe/IIe vs. M1 present/T1 present/P1 IIe/IIe, Model 2 = M1 present/T1 null/P1 IIe/IIe vs. M1 present/T1 present/P1 IIe/IIe, Model 3 = M1 present/T1 present/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe, Model 4 = all one high-risk genotype vs. M1 present/T1 present/P1 IIe/IIe, Model 5 = M1 null/T1 null/P1 IIe/IIe vs. M1 present/T1 present/P1 IIe/IIe, Model 6 = M1 null/T1 present/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe, Model 7 = M1 present/T1 null/P1 Val1 vs. M1 present/T1 present/P1 IIe/IIe, Model 8 = all two high-risk genotype vs. M1 present/T1 present/P1 IIe/IIe, Model 9 = M1 null/T1 null/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe, and Model 10 = M1 null/T1 null/P1 Val 1 vs. M1 present/T1 present/P1 IIe/IIe + all one high-risk genotype + all two high-risk genotypes.

((T1 null/P1 IIe/IIe + T1 present/P1 Val*) vs. T1 present/P1 IIe/IIe: OR = 1.65, 95% CI = 1.05–2.59; T1 null/P1 Val* vs. T1 present/P1 IIe/IIe: OR = 4.39, 95% CI = 2.51–7.68; all risk genotypes vs. T1 present/P1 IIe/IIe: OR = 1.91, 95% CI = 1.45–2.50; T1 null/P1 Val* vs. (T1 present/P1 IIe/IIe + T1 null/P1 IIe/IIe + T1 Present/P1 Val*): OR = 3.39, 95% CI = 1.94–5.94; Table 6 and Figure 7).

No significantly increased leukemia risk was observed in the three combined polymorphisms in the overall populations (Table 7; Figure 8).

Heterogeneity and sensitivity analyses

The metaregression analysis showed that race (p = 0.000) and quality score (p = 0.038) were sources of heterogeneity for the *GSTM1* null genotype. For *GSTP1* IIe105Val polymorphism, in Val/Val vs. IIe/IIe + IIe/Val, type of controls (p = 0.002), matching studies (p = 0.023), and HWE (p = 0.005) were the heterogeneity sources. Similar results were observed in Val/Val vs. lle/lle + lle/Val where type of controls (p = 0.001), matching studies (p = 0.037), and HWE (p = 0.007) were the sources of heterogeneity. For the combined *GSTM1* and *GSTT1* polymorphisms, the sample size (model 1: p = 0.015) was the source of heterogeneity (Table 8). Three methods were performed to appraise the sensitivity analysis, and all results did not change (Tables 1–7), indicating that the present study was stable.

Publication bias

Publication bias was found for the *GSTM1* null genotype (p = 0.003, Figure 9), GSTT1 null genotype (p = 0.041, Figure 10), and *GSTP1* IIe105Val (Val/Val vs. IIe/IIe: p = 0.001, IIe/Val vs. IIe/IIe: p = 0.030, Val/Val vs. IIe/IIe + IIe/Val: p = 0.020, Val/Val + IIe/Val vs. IIe/IIe: p = 0.022, Val vs. IIe: p = 0.033, Figure 11). Then, we used nonparametric "trim and fill" to adjust publication bias, and the results did not change (data not shown).

Credibility of the positive results

The "reliable results" was defined as the positive results that met the following criteria (Theodoratou et al., 2012). First, these positive results were observed in at least two of the genetic models (exclude individual *GSTM1* and *GSTT1* polymorphisms with the risk of leukemia), second, FPRP <0.2 and BFDP <0.8, third, I^2 < 50%, and fourth, statistical power >80%. Table 9 lists the credibility of the present meta-analysis on the individual and the composite effects of *GSTM1*, *GSTT1,* and *GSTP1* IIe105Val polymorphisms with the risk of leukemia. Only the GSTT1 null genotype with leukemia risk in Asians was considered as "positive" results (OR = 1.30, 95% CI = 1.16–1.46, I^2 = 24.2%, statistical power = 0.992, FPRP = 0.009, and

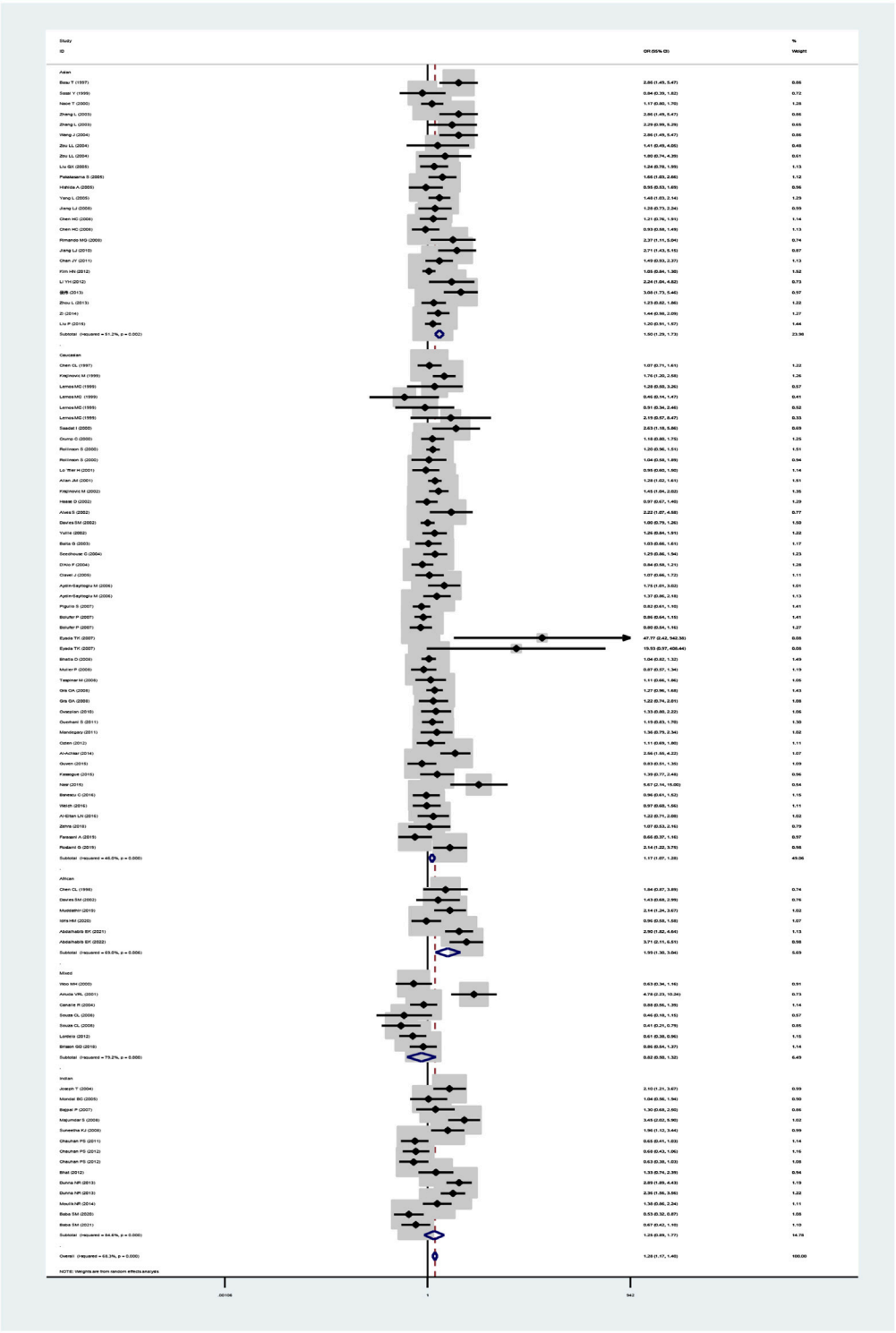


FIGURE 2
Forest plot for the association between GSTM1 polymorphism and leukemia risk in ethnicity subgroup analysis.

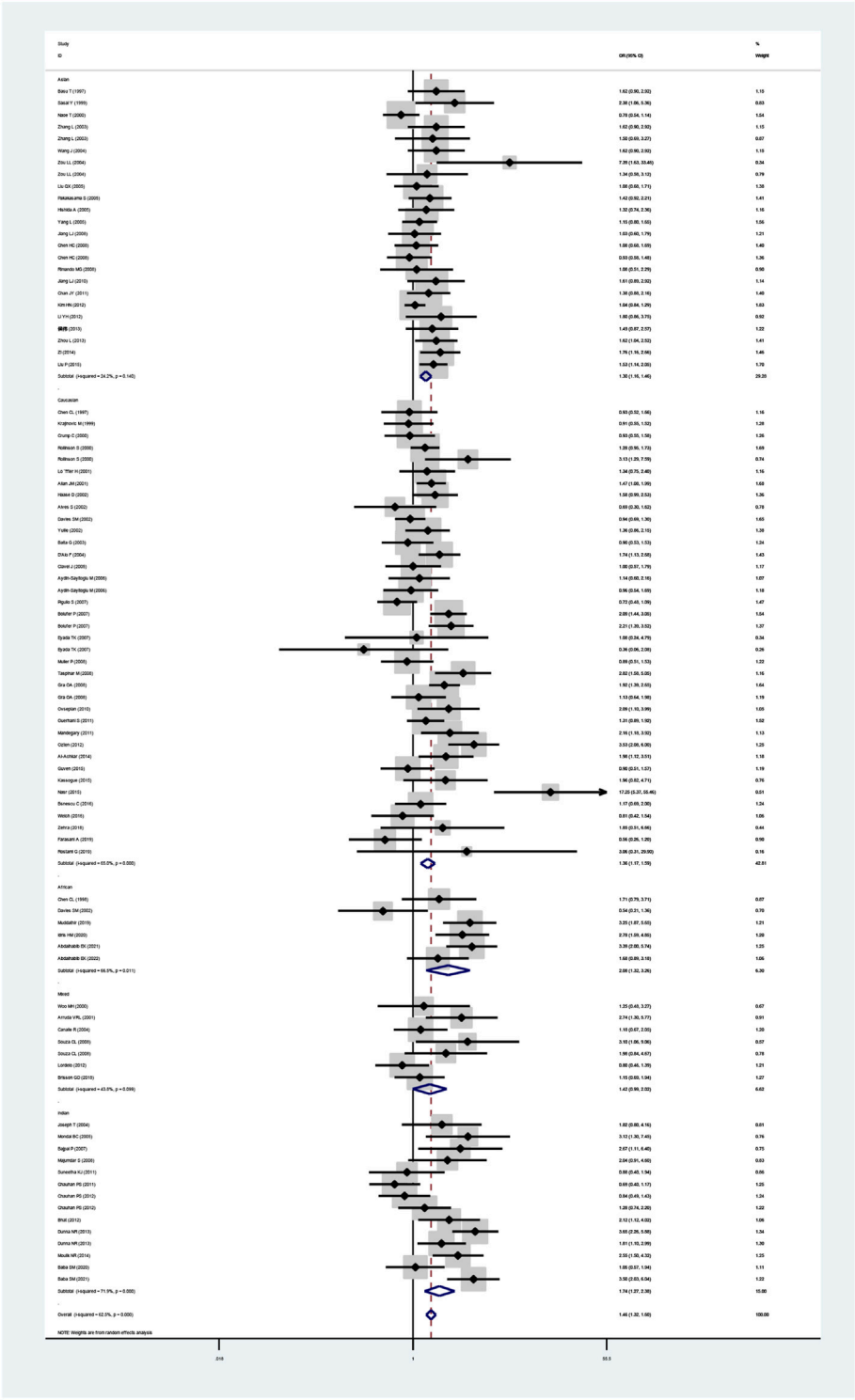


FIGURE 3
Forest plot for the association between GSTT1 polymorphism and leukemia risk in ethnicity subgroup analysis.

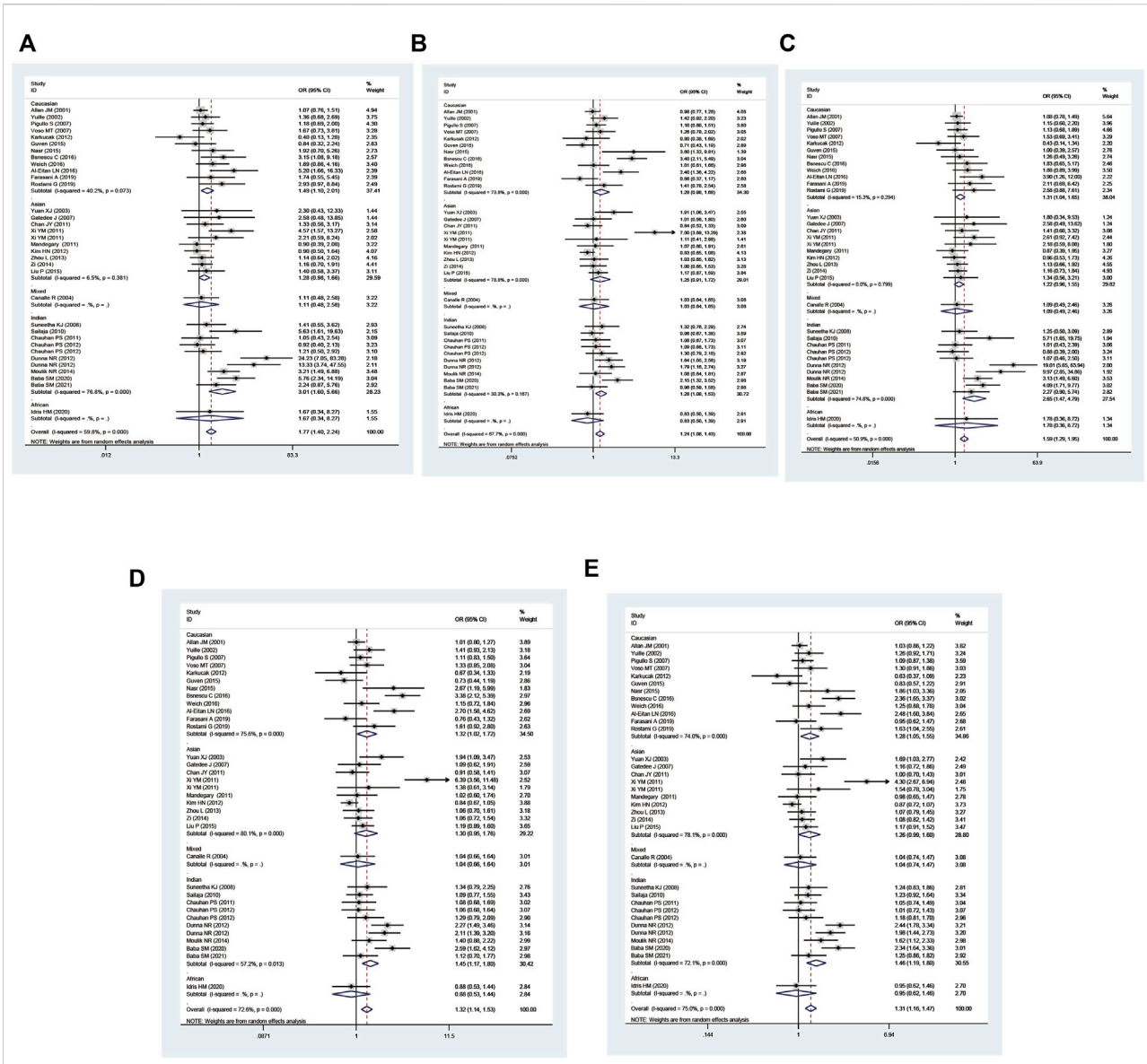


FIGURE 4
Forest plot for the association between GSTP1 polymorphism and leukemia risk in ethnicity subgroup analysis [**(A)**: Val/Val vs. Ile/Ile; **(B)** Ile/Val vs. Ile/Ile; **(C)** Val/Val vs. Ile/Ile + Ile/Val; **(D)** Val/Val + Ile/Val vs. Ile/Ile; and **(E)** Val vs. Ile].

BFDP = 0.367). All other important connections were regarded as less-credible results, also shown in Table 9.

Discussion

Leukemia is characterized by abnormal hematopoietic function and malignant cloning of white blood cells (Ouerhani et al., 2011). Gene polymorphisms play a significant role in the development of leukemia, and GST null has been studied by many scholars. Studies demonstrated that complete deletion of *GSTM1*, *GSTT1,* or *GSTP1* polymorphisms brought about diminished gene expression and enzymatic activity (Strange et al., 1998; Strange et al., 2001; Hollman et al., 2016). Thus, it is significant to study the connection between GST polymorphisms and leukemia risk. Many studies have analyzed the roles of M1, T1, and P1 polymorphisms in leukemia risk. Regrettably, no reliable testimony has been obtained to show whether there is an association between them. This may be due to heterogeneities such as ethnicity, small sample size, matching, type of leukemia, etc. Therefore, an updated meta-analysis was generated to explore these issues. At this point, totally 91 articles were finally selected to provide proof

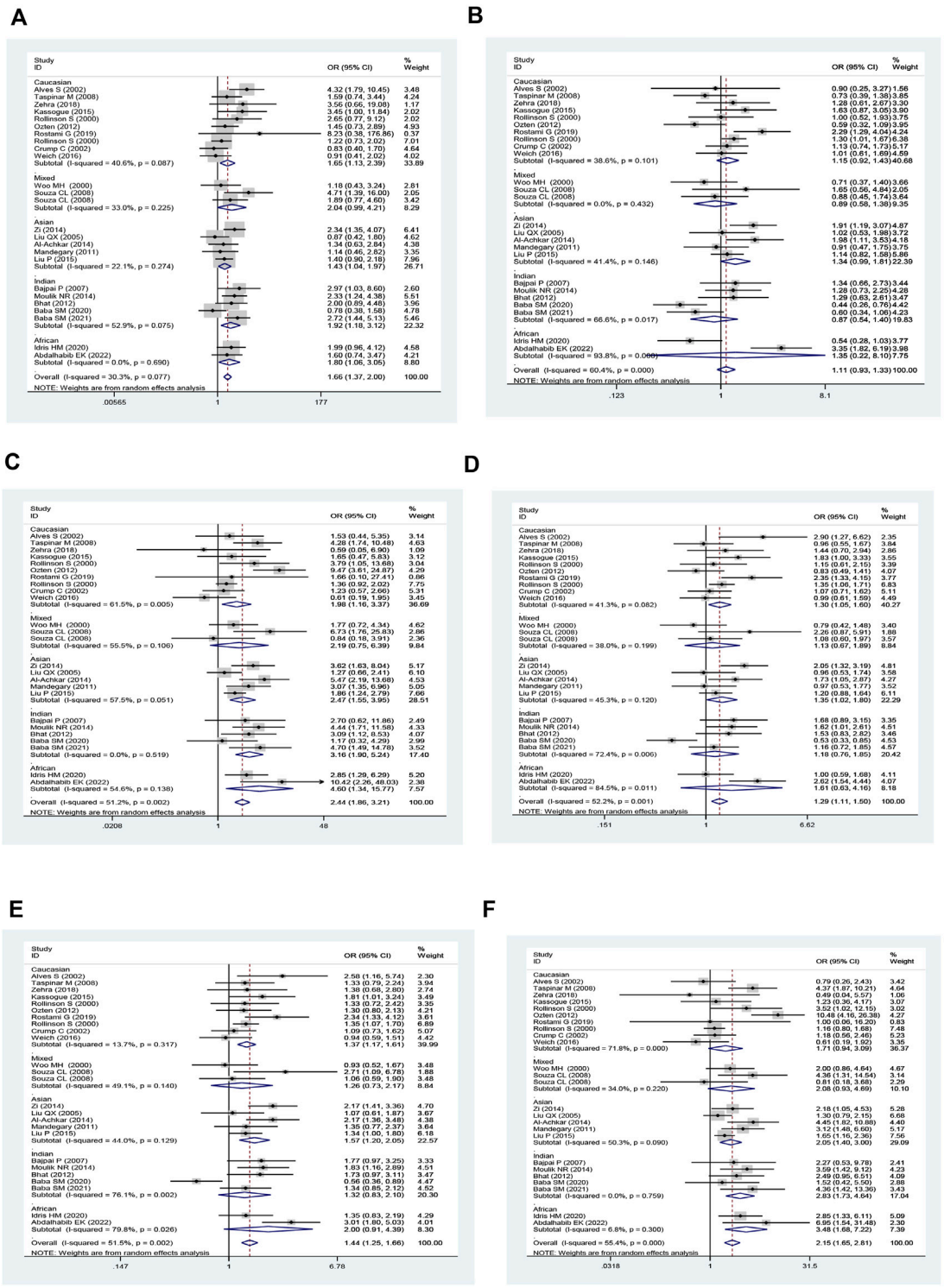


FIGURE 5
Forest plot of the association between combined effects of GSTM1 present/null and GSTT1 present/null polymorphisms and leukemia risk in ethnicity subgroup analysis [**(A)**: Model 1; **(B)** Model 2; **(C)** Model 3; **(D)** Model 4; **(E)** Model 5; and **(F)** Model 6].

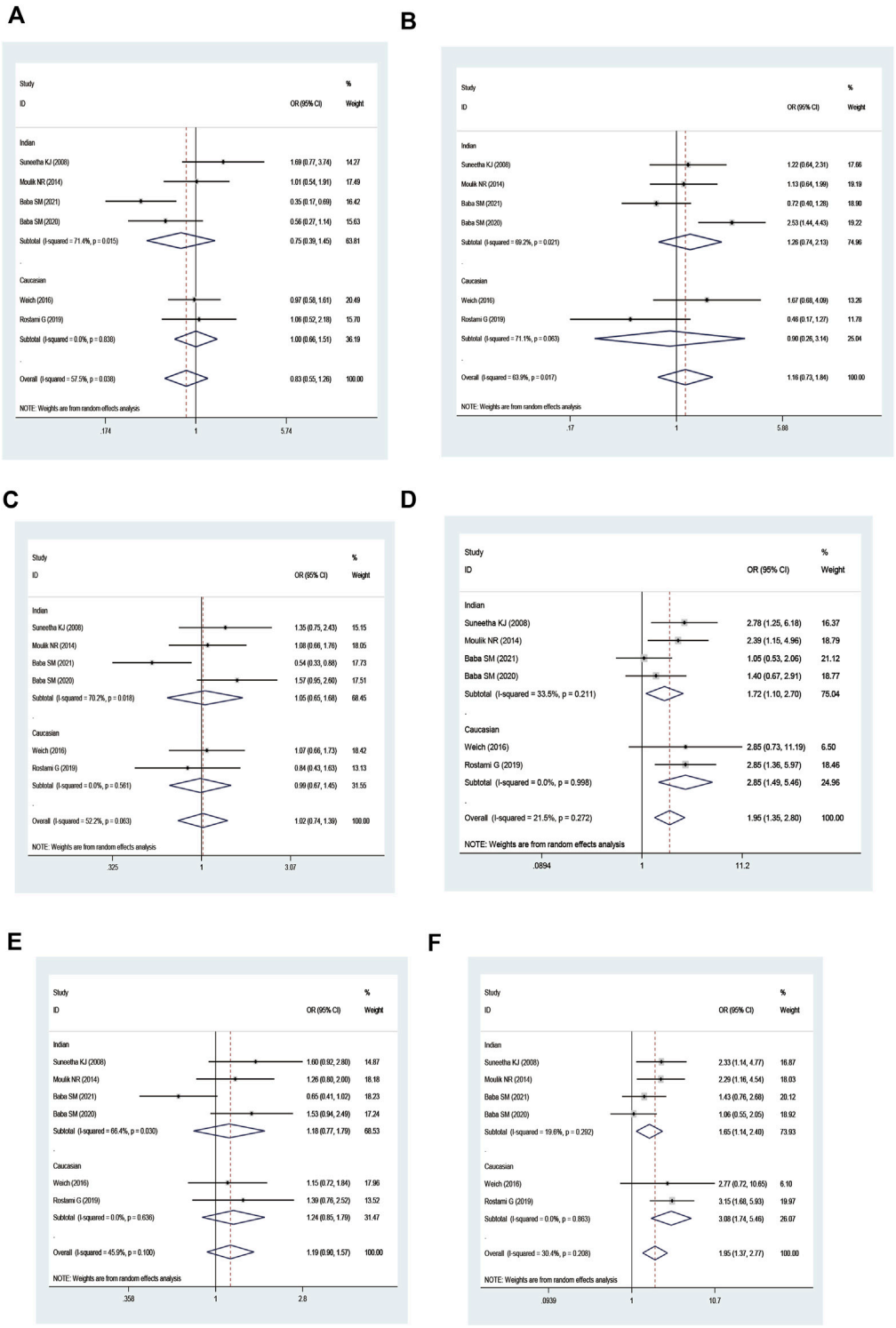


FIGURE 6
Forest plot of the association between combined effects of GSTM1 present/null and GSTP1 11e105Val polymorphisms and leukemia risk in ethnicity subgroup analysis [**(A)**: Model 1; **(B)** Model 2; **(C)** Model 3; **(D)** Model 4; **(E)** Model 5; and **(F)** Model 6].

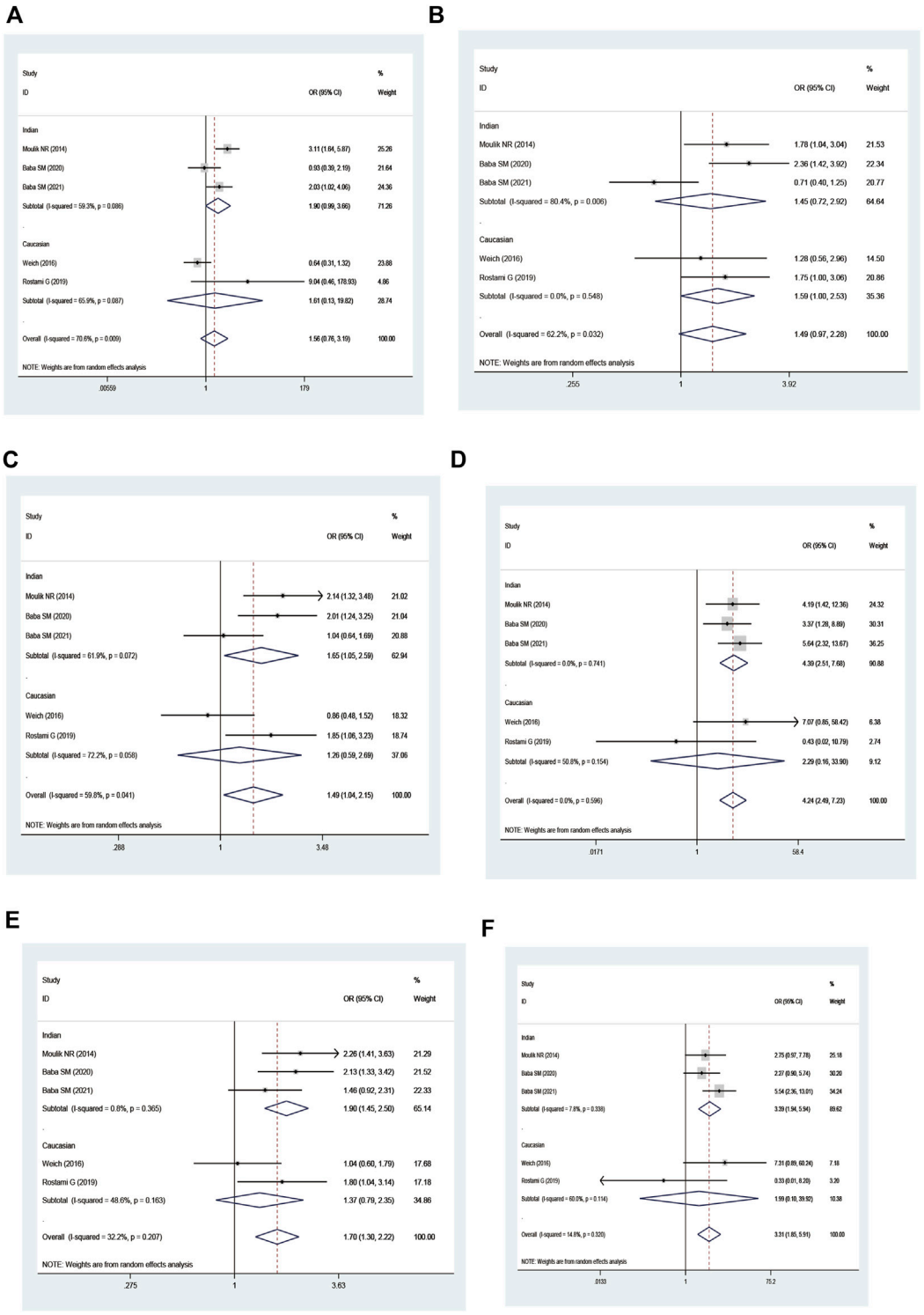


FIGURE 7
Forest plot of the association between the combined effects of GSTT1 present/null and GSTP1 11e105Val polymorphisms and leukemia risk in ethnicity subgroup analysis [**(A)**: Model 1; **(B)** Model 2; **(C)** Model 3; **(D)** Model 4; **(E)** Model 5; and **(F)** Model 6].

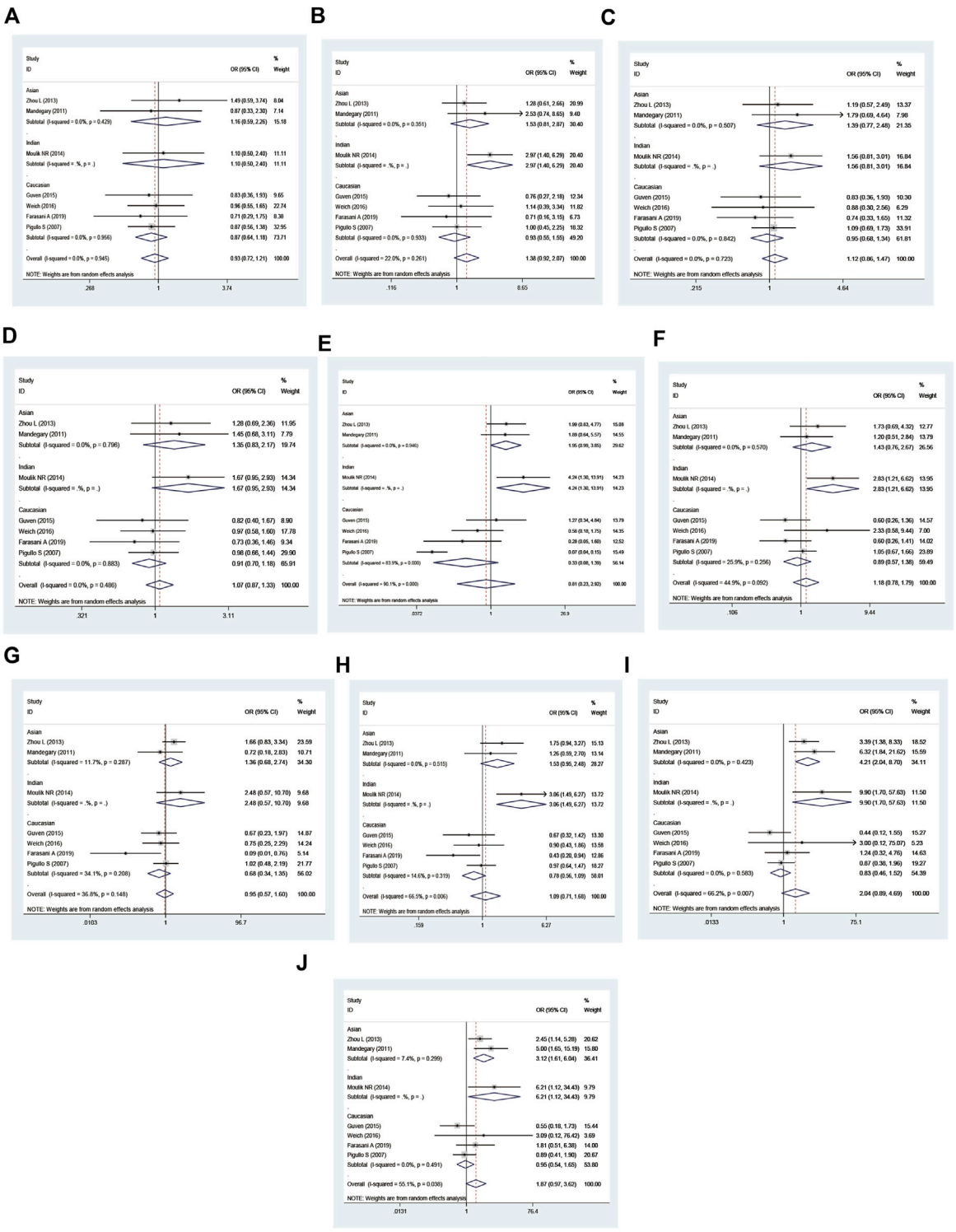


FIGURE 8
Forest plot of the association between the combined effects of GSTM1 present/null, GSTT1 present/null, and GSTP111e105Val polymorphisms and leukemia risk in the ethnicity subgroup analysis [**(A)**: Model 1; **(B)** Model 2; **(C)** Model 3; **(D)** Model 4; **(E)** Model 5; **(F)** Model 6; **(G)** Model 7; **(H)** Model 8; **(I)** Model 9; and **(J)** Model 10].

TABLE 8 Heterogeneity analysis in current meta-analysis.

Variables	Type of leukemia	Age group	Ethnicity	Sample size	Type of control	Matching	HWE	Quality score
P								
Genotype								
GSTM1	0.342	0.957	0.000	0.137	0.777	0.137	—	0.038
GSTT1	0.075	0.781	0.974	0.111	0.913	0.052	—	0.930
GSTP1 IIe105Val								
Val/Val vs. lle/lle	0.144	0.546	0.074	0.134	0.002	0.023	0.005	0.617
lle/Val vs. lle/lle	0.385	0.450	0.767	0.892	0.445	0.190	0.280	0.714
Val/Val vs. lle/lle + lle/Val	0.185	0.648	0.081	0.100	0.001	0.037	0.007	0.642
Val/Val+ lle/Val vs. lle/lle	0.341	0.525	0.575	0.706	0.244	0.098	0.142	0.829
Val vs. lle	0.328	0.616	0.463	0.528	0.106	0.064	0.073	0.878
The combined effects of GSTM1 and GSTT1 polymorphisms								
Model 1	0.648	0.067	0.432	0.015	0.622	0.212	—	0.478
Model 2	0.349	0.281	0.071	0.537	0.234	0.532	—	0.886
Model 3	0.702	0.917	0.792	0.686	0.739	0.714	—	0.699
Model 4	0.341	0.979	0.215	0.161	0.721	0.987	—	0.753
Model 5	0.402	0.939	0.124	0.268	0.850	0.974	—	0.644
Model 6	0.882	0.801	0.956	0.361	0.627	0.667	—	0.796

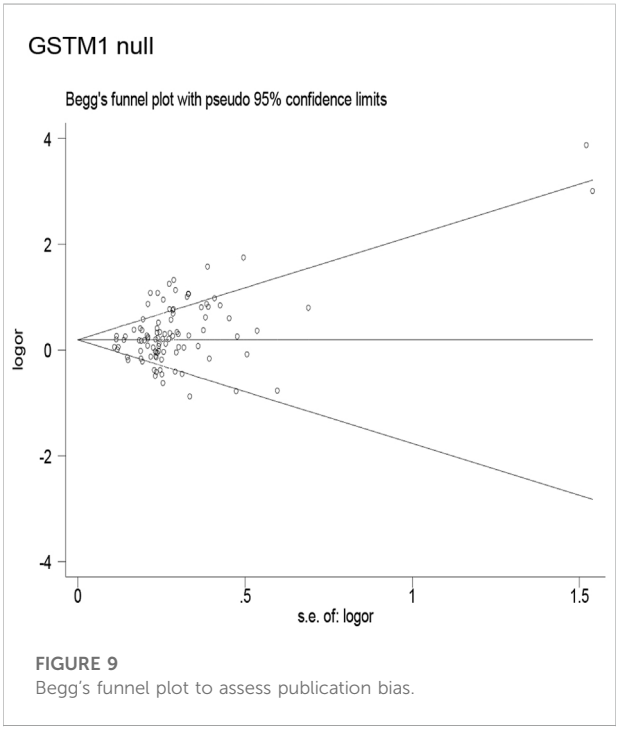


FIGURE 9
Begg's funnel plot to assess publication bias.

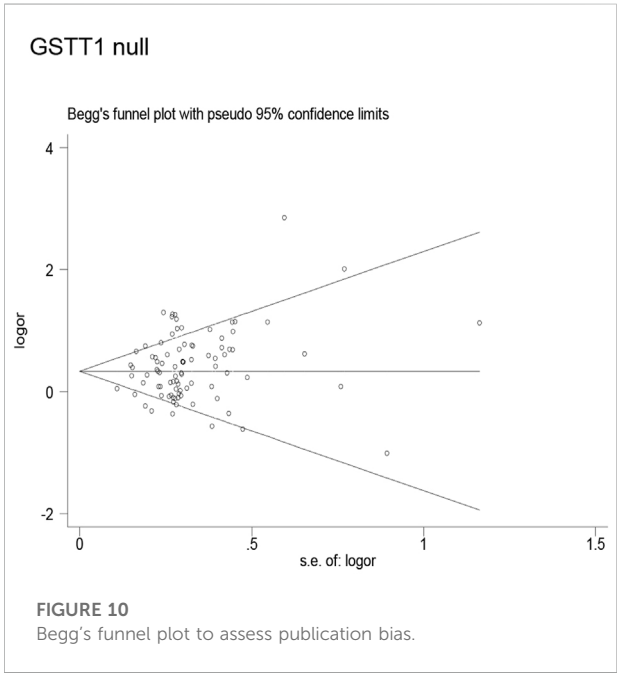


FIGURE 10
Begg's funnel plot to assess publication bias.

for the association between GST polymorphisms and leukemia risk.

Overall, the present study showed that the *GSTM1*, *GSTT1*, and *GSTP1* polymorphisms significantly added the risk of leukemia in the overall and several subgroups. Moreover, with the combined *GSTM1* and *GSTT1*, *GSTM1* and *GSTP1*, and *GSTT1* and *GSTP1* polymorphisms, there were six gene models to explore the association with leukemia risk, and positive results were observed in partial gene models. However, there was no significant contact between the composite effects of these three polymorphisms with

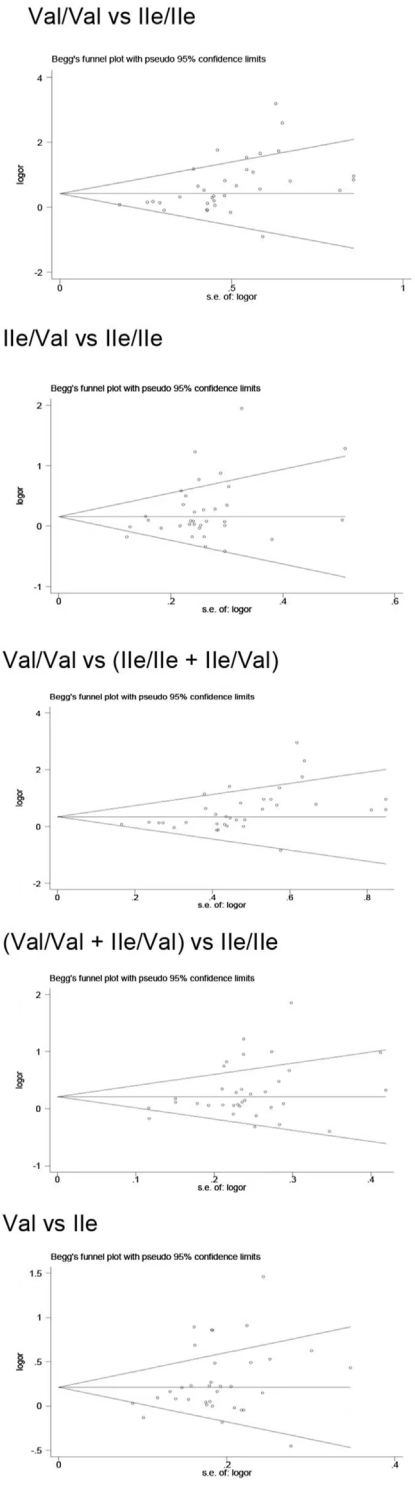


FIGURE 11
Begg's funnel plot to assess publication bias.

leukemia in overall analysis. Furthermore, in sensitivity analysis, when selecting Hardy–Weinberg equilibrium (HWE) and medium and high-quality studies, we had come to a similar conclusion. Finally, in view of the quantities of genomic data being produced currently, we used a more exact Bayesian measure of false-positive found in genetic epidemiological studies in the present study. Using FPRP and BFDP to correct the positive results, in all of these positive results we found previously, only the association between GSTT1 null and leukemia risk was watched in ethnicity (BFDP = 0.367, FPRP = 0.009). Our results indicated that the false-positive associations were common between SNP and disease risk. Moreover, these results further confirmed that the occurrence of leukemia was the result of multiple genes.

Thirteen previous meta-analyses analyzed the links between *GSTM1*, *GSTT1*, and *GSTP1* polymorphisms and the risk of leukemia. Tang et al. (2014), Ye and Song (2005), Wang et al. (2019), Zhang et al. (2017), Das et al. (2009), and He et al. (2014) discussed the association between *GSTM1* and *GSTT1* null genotypes and the risk of leukemia, and their results suggested that there was a significant association between *GSTM1* and *GSTT1* polymorphisms and leukemia risk. The studies of Ma et al. (2014) and Tang et al. (2013) showed that *GSTM1* null genotypes increased the risk of acute leukemia. The results of Moulik et al. (2014) demonstrated that there was a significant connection between *GSTP1* polymorphism with the risk of leukemia; however, Huang et al. (2013) discussed the association between *GSTP1* polymorphism and the risk of leukemia, and the results showed that there was no significant connection. The number of studies and sample sizes in the current study were larger than the published meta-analyses. When comparing to the present meta-analysis, previous studies had several defects. First, none of the previous studies performed quality assessments. Second, HWE was not reported in any published meta-analysis. Third, all previous meta-analyses did not adjust the positive results for multiple comparisons, and only five previous meta-analyses (Ye and Song, 2005; Huang et al., 2013; Tang et al., 2013; Tang et al., 2014; Zhang et al., 2017) conducted subgroup analysis. Fourth, there were no published meta-analyses that performed sensitivity analysis. Moreover, previous meta-analyses had a small sample size; most eligible studies were not assessed for quality assessment; and the reliability of positive results was not evaluated using FPRP, BFDP, and Venice criteria. In addition, they failed to establish a more complete genetic model. Thus, their meta-analyses might have lower credibility.

The current meta-analysis had some advantages over previously published meta-analyses. 1) We explored the credibility by applying the Venice criteria, FPRP, and BFDP. 2) The qualified studies were evaluated for quality. 3) The sample size was larger and the data collected were more detailed over the previous meta-analyses. 4) We conducted more subunit analyses, such as ethnicity, age group, type of control, matching or not, type of leukemia, quality score, and HWE. 5) We established a more complete genetic model. 6) Our study is the first one to explore the combined effects of *GSTM1*, *GSTT1,* and *GSTP1* polymorphisms with leukemia risk. Nonetheless, there are still some potential limitations for this current study. First, in this study, we only studied published research studies, and as we all know, the positive results are more likely to be published than the

TABLE 9 Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
GSTM1						
Overall	Null vs present	1.28 (1.17–1.40)	68.3	1.000	<0.001	0.006
Asian	Null vs present	1.50 (1.29–1.73)	51.2	0.500	<0.001	0.002
Caucasian	Null vs present	1.17 (1.07–1.28)	46.0	1.000	0.381	0.973
African	Null vs present	1.99 (1.30–3.94)	69.0	0.209	0.996	0.998
Adults	Null vs present	1.26 (1.11–1.43)	65.6	0.997	0.257	0.940
Children	Null vs present	1.42 (1.23–1.64)	64.4	0.772	0.002	0.096
HC	Null vs present	1.29 (1.15–1.44)	66.6	0.996	0.006	0.273
NBDC	Null vs present	1.29 (1.13–1.48)	71.9	0.984	0.222	0.924
Matching	Null vs present	1.36 (1.12–1.65)	77.7	0.840	0.684	0.981
Nonmatching	Null vs present	1.25 (1.14–1.38)	63.7	1.000	0.010	0.408
AML	Null vs present	1.20 (1.04–1.38)	71.1	0.999	0.914	0.997
ALL	Null vs present	1.44 (1.25–1.65)	66.8	0.722	<0.001	0.010
Sensitivity analysis						
Quality score ≥10						
Overall	Null vs present	1.16 (1.05–1.27)	62.2	1.000	0.569	0.986
Asian	Null vs present	1.17 (1.05–1.31)	0.0	1.000	0.866	0.996
Caucasian	Null vs present	1.17 (1.06–1.30)	45.5	1.000	0.777	0.993
African	Null vs present	2.01 (1.23–3.30)	75.1	0.124	0.979	0.990
Adults	Null vs present	1.31 (1.15–1.50)	63.2	0.975	0.087	0.816
Children	Null vs present	1.21 (1.06–1.39)	27.0	0.999	0.876	0.996
HC	Null vs present	1.21 (1.05–1.39)	69.4	0.999	0.876	0.996
Matching	Null vs present	1.26 (1.05–1.52)	73.5	0.966	0.942	0.997
Nonmatching	Null vs present	1.13 (1.02–1.27)	57.7	1.000	0.976	0.999
ALL	Null vs present	1.22 (1.01–1.46)	63.7	0.988	0.968	0.998
GSTT1						
Overall	Null vs present	1.46 (1.32–1.60)	62.5	0.710	<0.001	<0.001
Indian	Null vs present	1.74 (1.27–2.38)	71.9	0.177	0.749	0.934
Asian	Null vs present	1.30 (1.16–1.46)	24.2	0.992	0.009	0.367
Caucasian	Null vs present	1.37 (1.17–1.59)	65.0	0.884	0.037	0.619
African	Null vs present	2.08 (1.32–3.26)	66.5	0.720	0.999	0.971
Adults	Null vs present	1.55 (1.32–1.82)	69.6	0.344	<0.001	0.006
Children	Null vs present	1.24 (1.09–1.43)	37.2	0.996	0.754	0.991
Adults and Children	Null vs present	1.59 (1.27–1.99)	67.1	0.305	0.143	0.655
HC	Null vs present	1.45 (1.28–1.66)	63.7	0.688	<0.001	0.005
NBDC	Null vs present	1.46 (1.26–1.69)	62.7	0.641	0.001	0.024
Matching	Null vs present	1.80 (1.44–2.24)	63.7	0.051	0.003	0.008
Nonmatching	Null vs present	1.35 (1.22–1.49)	51.7	0.982	<0.001	<0.001
AML	Null vs present	1.41 (1.19–1.66)	67.7	0.771	0.046	0.622
ALL	Null vs present	1.33 (1.16–1.53)	53.0	0.954	0.065	0.758
CML	Null vs present	1.88 (1.47–2.41)	64.5	0.037	0.017	0.033

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
Sensitivity analysis						
Quality score ≥10						
Overall	Null vs present	1.52 (1.34–1.72)	66.9	0.417	<0.001	<0.001
Indian	Null vs present	1.53 (1.08–2.17)	69.6	0.456	0.974	0.996
Asian	Null vs present	1.15 (1.01–1.31)	21.5	1.000	0.973	0.999
Caucasian	Null vs present	1.64 (1.37–1.96)	64.5	0.163	<0.001	0.003
African	Null vs present	2.12 (1.26–3.58)	71.9	0.098	0.981	0.989
Adults	Null vs present	1.58 (1.33–1.89)	71.3	0.285	0.002	0.030
Adults and Children	Null vs present	1.45 (1.14–1.83)	61.5	0.612	0.741	0.978
HC	Null vs present	1.56 (1.31–1.86)	71.0	0.331	0.002	0.038
NBDC	Null vs present	1.45 (1.23–1.72)	56.7	0.651	0.030	0.475
Matching	Null vs present	1.73 (1.37–2.17)	73.6	0.109	0.019	0.093
Nonmatching	Null vs present	1.41 (1.23–1.62)	59.2	0.809	0.002	0.069
AML	Null vs present	1.35 (1.12–1.63)	68.3	0.863	0.676	0.981
ALL	Null vs present	1.49 (1.19–1.88)	64.2	0.522	0.597	0.956
CML	Null vs present	1.93 (1.44–2.59)	64.9	0.047	0.202	0.332
GSTP1						
Overall	Val/Val vs. lle/lle	1.77 (1.40–2.24)	59.8	0.084	0.023	0.089
	lle/Val vs. lle/lle	1.24 (1.08–1.43)	67.7	0.996	0.757	0.991
	Val/Val vs. lle/lle + lle/Val	1.59 (1.29–1.95)	50.9	0.288	0.028	0.273
	Val/Val+lle/Val vs. lle/lle	1.32 (1.15–1.53)	72.6	0.955	0.193	0.905
	Val vs lle	1.31 (1.16–1.47)	75.0	0.989	0.004	0.220
Indian	Val/Val vs. lle/lle	3.01 (1.60–5.66)	76.8	0.015	0.976	0/961
	lle/Val vs. lle/lle	1.28 (1.08–1.53)	30.3	0.959	0.874	0.994
	Val/Val vs. lle/lle +lle/Val	2.65 (1.47–4.79)	74.8	0.030	0.977	0.974
	Val/Val+lle/Val vs. lle/lle	1.45 (1.17–1.80)	57.2	0.621	0.549	0.957
	Val vs lle	1.47 (1.19–1.80)	72.1	0.578	0.250	0.869
Caucasian	Val/Val vs. lle/lle	1.49 (1.10–2.01)	40.2	0.517	0.946	0.994
	Val/Val vs. lle/lle +lle/Val	1.31 (1.04–1.65)	15.3	0.875	0.961	0.997
	Val/Val+lle/Val vs. lle/lle	1.32 (1.02–1.72)	75.6	0.828	0.980	0.998
	Val vs lle	1.28 (1.05–1.55)	74.0	0.948	0.924	0.996
Adults	Val/Val vs. lle/lle	1.39 (1.06–1.82)	34.1	0.710	0.959	0.996
	Val/Val vs.lle/lle + lle/Val	1.27 (1.01–1.61)	20.2	0.915	0.981	0.999
	Val vs lle	1.18 (1.02–1.37)	64.6	0.999	0.968	0.999
Children	Val/Val vs. lle/lle	1.68 (1.10–2.58)	39.6	0.302	0.983	0.996
	Val/Val vs.lle/lle + lle/Val	1.60 (1.11–2.32)	25.8	0.367	0.973	0.995
Adults and Children	Val/Val vs. lle/lle	3.25 (1.61–6.53)	76.8	0.015	0.984	0.974
	lle/Val vs. lle/lle	1.64 (1.16–2.31)	73.7	0.305	0.938	0.989
	Val/Val vs. lle/lle +lle/Val	2.65 (1.41–5.02)	72.9	0.040	0.986	0.986
	Val/Val+lle/Val vs. lle/lle	1.82 (1.29–2.57)	77.3	0.136	0.831	0.945
	Val vs lle	1.72 (1.29–2.30)	80.4	0.176	0.588	0.883
HC	Val/Val vs. lle/lle	2.38 (1.66–3.41)	61.2	0.006	0.278	0.118
	lle/Val vs. lle/lle	1.27 (1.05–1.54)	65.2	0.955	0.940	0.997

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
	Val/Val vs. lle/lle + lle/Val	2.12 (1.53–2.94)	55.3	0.019	0.259	0.239
	Val/Val+lle/Val vs. lle/lle	1.39 (1.15–1.69)	69.6	0.778	0.552	0.967
	Val vs lle	1.40 (1.19–1.63)	71.5	0.813	0.018	0.419
Nonmatching	Val/Val vs. lle/lle	1.37 (1.07–1.76)	23.1	0.761	0.948	0.996
	Val/Val vs.lle/lle + lle/Val	1.30 (1.03–1.64)	19.2	0.886	0.968	0.998
	Val vs lle	1.13 (1.03–1.24)	6.1	1.000	0.908	0.998
Nonmatching	Val/Val vs. lle/lle	2.13 (1.49–3.06)	69.0	0.029	0.598	0.628
	lle/Val vs. lle/lle	1.36 (1.08–1.71)	78.4	0.799	0.914	0.994
	Val/Val vs. lle/lle + lle/Val	1.86 (1.36–2.54)	60.6	0.088	0.518	0.760
	Val/Val+lle/Val vs. lle/lle	1.47 (1.17–1.86)	81.7	0.567	0.701	0.972
	Val vs lle	1.44 (1.19–1.74)	82.8	0.664	0.193	0.852
AML	Val/Val vs. lle/lle	1.57 (1.10–2.24)	66.4	0.401	0.970	0.995
	lle/Val vs. lle/lle	1.37 (1.02–1.84)	83.1	0.727	0.980	0.998
	Val/Val vs. lle/lle + lle/Val	1.37 (1.01–1.84)	55.1	0.727	0.980	0.998
	Val/Val+lle/Val vs. lle/lle	1.42 (1.06–1.89)	84.8	0.646	0.962	0.996
	Val vs lle	1.34 (1.07–1.68)	85.1	0.836	0.930	0.996
ALL	Val/Val vs. lle/lle	1.90 (1.28–2.81)	52.1	0.118	0.917	0.968
	Val/Val vs. lle/lle +lle/Val	1.77 (1.25–2.53)	44.0	0.182	0.905	0.974
	Val/Val+lle/Val vs. lle/lle	1.26 (1.03–1.53)	51.4	0.961	0.953	0.998
	Val vs lle	1.29 (1.08–1.53)	61.0	0.958	0.782	0.990
CML	Val/Val vs. lle/lle	1.29 (1.08–1.53)	67.3	0.958	0.782	0.990
	Val/Val vs.lle/lle + lle/Val	2.13 (1.08–4.24)	60.1	0.159	0.995	0.997
Sensitivity analysis						
HWE						
Overall	Val/Val vs. lle/lle	1.58 (1.27–1.95)	26.3	0.314	0.061	0.455
	lle/Val vs. lle/lle	1.18 (1.02–1.37)	59.0	0.999	0.968	0.999
	Val/Val vs. lle/lle +lle/Val	1.45 (1.21–1.74)	7.3	0.642	0.092	0.722
	Val/Val+lle/Val vs. lle/lle	1.25 (1.07–1.45)	64.1	0.992	0.764	0.991
	Val vs lle	1.24 (1.10–1.40)	64.4	0.999	0.339	0.959
Indian	Val/Val vs. lle/lle	1.83 (1.11–3.03)	56.3	0.220	0.988	0.996
	lle/Val vs. lle/lle	1.24 (1.01–1.51)	12.7	0.971	0.971	0.998
	Val/Val vs. lle/lle + lle/Val	1.67 (1.05–2.64)	51.3	0.323	0.989	0.997
	Val/Val+lle/Val vs. lle/lle	1.34 (1.06–1.69)	42.1	0.830	0.942	0.996
	Val vs lle	1.33 (1.06–1.66)	61.2	0.856	0.932	0.996
Caucasian	Val/Val vs. lle/lle	1.70 (1.23–2.34)	21.0	0.221	0.837	0.964
	Val/Val vs. lle/lle +lle/Val	1.50 (1.14–1.95)	0.0	0.500	0.831	0.982
	Val/Val+lle/Val vs. lle/lle	1.39 (1.01–1.90)	76.4	0.684	0.983	0.998
	Val vs lle	1.36 (1.08–1.70)	72.8	0.805	0.896	0.993
Adults	Val/Val vs. lle/lle	1.39 (1.07–1.81)	0.0	0.714	0.953	0.996
	Val/Val vs.lle/lle + lle/Val	1.31 (1.02–1.69)	0.0	0.851	0.978	0.998
	Val vs lle	1.22 (1.01–1.46)	67.7	0.988	0.968	0.998
Children	Val/Val vs. lle/lle	1.68 (1.10–2.58)	39.6	0.302	0.983	0.996
	Val/Val vs.lle/lle + lle/Val	1.60 (1.11–2.32)	25.8	0.367	0.973	0.995

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
Adults and Children	Val/Val+lle/Val vs. lle/lle	1.39 (1.01–1.92)	57.0	0.678	0.985	0.998
HC	Val/Val vs. lle/lle	1.86 (1.38–2.50)	34.2	0.077	0.336	0.591
	Val/Val vs. lle/lle +lle/Val	1.71 (1.33–2.21)	16.0	0.158	0.207	0.603
	Val/Val+lle/Val vs. lle/lle	1.31 (1.05–1.62)	68.9	0.894	0.934	0.996
	Val vs lle	1.31 (1.11–1.55)	68.1	0.943	0.637	0.981
Matching	Val/Val vs. lle/lle	1.51 (1.11–2.05)	0.0	0.483	0.945	0.993
	Val/Val vs. lle/lle + lle/Val	1.43 (1.07–1.93)	0.0	0.623	0.969	0.997
	Val/Val+lle/Val vs. lle/lle	1.17 (1.01–1.35)	0.0	1.000	0.969	0.999
	Val vs lle	1.18 (1.05–1.32)	0.0	1.000	0.792	0.994
Non matching	Val/Val vs. lle/lle	1.66 (1.21–2.27)	47.1	0.263	0.851	0.972
	Val/Val vs. lle/lle + lle/Val	1.50 (1.15–1.94)	28.8	0.500	0.800	0.979
	Val/Val+lle/Val vs. lle/lle	1.31 (1.03–1.66)	76.9	0.869	0.967	0.998
	Val vs lle	1.29 (1.07–1.55)	76.6	0.946	0.874	0.994
ALL	Val/Val vs. lle/lle	1.60 (1.15–2.22)	26.8	0.350	0.933	0.989
	Val/Val vs.lle/lle + lle/Val	1.53 (1.14–2.06)	17.8	0.448	0.919	0.990
	Val vs lle	1.21 (1.03–1.43)	52.6	0.994	0.962	0.998
Quality score≥12						
Overall	Val/Val vs. lle/lle	1.62 (1.25–2.11)	45.7	0.284	0.549	0.910
	Val/Val vs. lle/lle + lle/Val	1.49 (1.18–1.89)	35.6	0.522	0.660	0.964
	Val/Val+lle/Val vs. lle/lle	1.23 (1.05–1.44)	64.0	0.993	0.910	0.996
	Val vs lle	1.23 (1.09–1.40)	65.5	0.999	0.633	0.985
Caucasian	Val/Val vs. lle/lle	1.54 (1.09–2.17)	28.7	0.440	0.969	0.995
	Val/Val vs. lle/lle +lle/Val	1.28 (1.01–1.64)	0.0	0.895	0.983	0.999
	Val/Val+lle/Val vs. lle/lle	1.48 (1.04–2.10)	77.2	0.530	0.981	0.997
	Val vs lle	1.38 (1.08–1.77)	73.1	0.744	0.938	0.995
Indian	Val/Val vs. lle/lle	2.15 (1.22–3.76)	62.6	0.103	0.986	0.992
	Val/Val vs. lle/lle + lle/Val	1.97 (1.16–3.35)	60.3	0.157	0.987	0.995
	Val/Val+lle/Val vs. lle/lle	1.29 (1.03–1.63)	47.1	0.897	0.973	0.998
	Val vs lle	1.32 (1.07–1.64)	61.6	0.876	0.933	0.996
Adults	Val/Val vs. lle/lle	1.42 (1.07–1.88)	33.1	0.649	0.957	0.996
	Val/Val vs. lle/lle + lle/Val	1.29 (1.02–1.64)	17.3	0.891	0.977	0.998
	Val/Val+lle/Val vs. lle/lle	1.24 (1.01–1.51)	68.9	0.971	0.971	0.998
	Val vs lle	1.21 (1.04–1.41)	65.5	0.997	0.936	0.997
HC	Val/Val vs. lle/lle	1.94 (1.37–2.76)	44.2	0.076	0.750	0.874
	Val/Val vs. lle/lle +lle/Val	1.77 (1.29–2.44)	36.9	0.156	0.758	0.930
	Val/Val+lle/Val vs. lle/lle	1.30 (1.05–1.62)	64.8	0.899	0.956	0.997
	Val vs lle	1.31 (1.10–1.55)	64.0	0.943	0.637	0.981
Matching	Val/Val vs. lle/lle	1.45 (1.11–1.90)	16.9	0.597	0.922	0.993
	Val/Val vs.lle/lle + lle/Val	1.41 (1.08–1.83)	18.9	0.679	0.935	0.995
	Val vs lle	1.14 (1.04–1.25)	0.0	1.000	0.841	0.996
Non matching	Val/Val vs. lle/lle	1.81 (1.07–3.06)	67.1	0.242	0.991	0.997
	Val/Val vs.lle/lle + lle/Val	1.58 (1.02–2.46)	55.9	0.409	0.991	0.998
CML	Val/Val vs. lle/lle	3.17 (1.89–5.32)	16.7	0.002	0.845	0.489

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
	Val/Val vs.lle/lle + lle/Val	2.80 (1.79–4.39)	0.0	0.003	0.688	0.322
	Val vs lle	1.41 (1.05–1.89)	68.3	0.661	0.970	0.997
HWE and Quality score≥12						
Overall	Val/Val vs. lle/lle	1.63 (1.24–2.13)	35.2	0.271	0.559	0.909
	Val/Val vs. lle/lle +lle/Val	1.49 (1.18–1.88)	20.2	0.522	0.597	0.956
	Val/Val+lle/Val vs. lle/lle	1.27 (1.06–1.53)	66.8	0.960	0.925	0.996
	Val vs lle	1.26 (1.08–1.46)	67.7	0.990	0.680	0.986
Indian	Val/Val vs. lle/lle	1.91 (1.07–3.40)	62.6	0.206	0.993	0.997
	Val/Val vs. lle/lle + lle/Val	1.74 (1.03–2.96)	57.8	0.292	0.993	0.998
	Val/Val+lle/Val vs. lle/lle	1.34 (1.02–1.76)	51.7	0.791	0.978	0.998
	Val vs lle	1.34 (1.04–1.74)	67.4	0.801	0.972	0.998
Caucasian	Val/Val vs. lle/lle	1.87 (1.28–2.74)	0.0	0.129	0.911	0.968
	lle/Val vs. lle/lle	1.55 (1.02–2.34)	71.9	0.438	0.988	0.998
	Val/Val vs. lle/lle +lle/Val	1.59 (1.11–2.30)	0.0	0.379	0.973	0.995
	Val/Val+lle/Val vs. lle/lle	1.63 (1.12–2.37)	68.7	0.332	0.969	0.994
	Val vs lle	1.50 (1.17–1.91)	56.4	0.500	0.668	0.964
Adults	Val/Val vs. lle/lle	1.38 (1.05–1.82)	3.2	0.723	0.969	0.997
	Val vs lle	1.24 (1.02–1.52)	70.4	0.967	0.975	0.999
HC	Val/Val vs. lle/lle	1.83 (1.29–2.58)	40.3	0.128	0.815	0.937
	Val/Val vs. lle/lle + lle/Val	1.66 (1.23–2.25)	28.2	0.257	0.809	0.963
	Val/Val+lle/Val vs. lle/lle	1.33 (1.05–1.68)	66.7	0.844	0.952	0.997
	Val vs lle	1.32 (1.09–1.59)	66.8	0.911	0.791	0.989
Matching	Val/Val vs. lle/lle	1.51 (1.11–2.05)	0.0	0.483	0.945	0.993
	Val/Val vs. lle/lle + lle/Val	1.43 (1.07–1.93)	0.0	0.623	0.969	0.997
	Val/Val+lle/Val vs. lle/lle	1.17 (1.01–1.35)	0.0	1.000	0.969	0.999
	Val vs lle	1.18 (1.05–1.32)	0.0	1.000	0.792	0.994
Nonmatching	Val/Val vs. lle/lle	1.81 (1.07–3.06)	67.1	0.242	0.991	0.997
	Val/Val vs. lle/lle +lle/Val	1.58 (1.02–2.46)	55.9	0.409	0.991	0.998
The combined effects of GSTM1 and GSTT1 polymorphisms						
Overall	Model 1	1.66 (1.37–2.00)	30.3	0.143	0.001	0.006
	Model 3	2.44 (1.86–3.21)	51.2	<0.001	0.001	<0.001
	Model 4	1.29 (1.11–1.50)	52.2	0.975	0.489	0.971
	Model 5	1.44 (1.25–1.66)	51.5	0.713	0.001	0.030
	Model 6	2.16 (1.65–2.81)	55.4	0.003	0.003	0.001
Indian	Model 1	1.92 (1.18–3.12)	52.9	0.159	0.981	0.993
	Model 3	3.16 (1.90–5.25)	0.0	0.002	0.816	0.412
	Model 6	2.83 (1.73–4.64)	0.0	0.006	0.863	0.674
Asian	Model 1	1.43 (1.04–1.97)	22.1	0.615	0.979	0.997
	Model 3	2.47 (1.55–3.95)	57.5	0.019	0.896	0.860
	Model 4	1.35 (1.02–1.80)	45.3	0.764	0.982	0.998
	Model 5	1.57 (1.20–2.05)	44.0	0.369	0.713	0.959
	Model 6	2.05 (1.40–3.00)	50.3	0.054	0.803	0.873
Caucasian	Model 1	1.65 (1.14–2.39)	40.6	0.307	0.963	0.993

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
	Model 3	1.98 (1.16–3.37)	61.5	0.153	0.987	0.994
	Model 4	1.30 (1.05–1.60)	41.3	0.912	0.936	0.996
	Model 5	1.37 (1.17–1.61)	13.7	0.864	0.132	0.843
Adults	Model 1	1.44 (1.18–1.76)	0.0	0.655	0.360	0.923
	Model 2	1.27 (1.04–1.54)	50.7	0.955	0.940	0.997
	Model 3	2.51 (1.71–3.68)	60.0	0.004	0.367	0.131
	Model 4	1.34 (1.15–1.57)	35.3	0.919	0.242	0.919
	Model 5	1.50 (1.29–1.74)	33.0	0.500	< 0.001	0.006
	Model 6	2.26 (1.53–3.33)	65.8	0.019	0.662	0.610
Adults and children	Model 6	1.94 (1.04–4.36)	43.6	0.267	0.998	0.999
HC	Model 1	1.73 (1.31–2.30)	39.2	0.163	0.498	0.835
	Model 3	2.59 (1.71–3.93)	51.1	0.005	0.600	0.310
	Model 5	1.45 (1.16–1.80)	60.3	0.621	0.549	0.957
	Model 6	2.33 (1.54–3.58)	59.7	0.022	0.8361	0.811
NBDC	Model 1	1.60 (1.22–2.10)	25.7	0.321	0.687	0.949
	Model 2	1.29 (1.11–1.50)	0.0	0.975	0.489	0.971
	Model 3	2.31 (1.56–3.43)	55.7	0.016	0.672	0.589
	Model 4	1.36 (1.18–1.57)	1.6	0.909	0.029	0.572
	Model 5	1.49 (1.25–1.78)	33.9	0.529	0.020	0.338
	Model 6	1.86 (1.33–2.61)	46.7	0.107	0.756	0.904
Matching	Model 1	1.60 (1.29–1.99)	0.0	0.281	0.079	0.491
	Model 3	2.57 (1.61–4.12)	64.2	0.463	0.999	0.999
	Model 4	1.31 (1.09–1.58)	41.0	0.922	0.837	0.992
	Model 5	1.46 (1.24–1.73)	33.8	0.623	0.019	0.367
	Model 6	2.33 (1.44–3.76)	69.6	0.036	0.937	0.942
Nonmatching	Model 1	1.67 (1.24–2.27)	48.0	0.247	0.811	0.963
	Model 3	2.38 (1.70–3.33)	37.6	0.004	0.106	0.027
	Model 5	1.43 (1.13–1.80)	62.1	0.658	0.779	0.983
	Model 6	2.07 (1.52–2.81)	36.5	0.019	0.137	0.133
AML	Model 3	2.15 (1.35–3.43)	55.1	0.065	0.953	0.970
	Model 5	1.41 (1.09–1.82)	46.7	0.683	0.924	0.994
	Model 6	1.85 (1.22–2.80)	51.1	0.161	0.957	0.986
ALL	Model 1	2.15 (1.43–3.23)	39.9	0.041	0.846	0.880
	Model 3	2.79 (1.47–5.30)	52.0	0.029	0.983	0.981
	Model 4	1.52 (1.13–2.05)	44.5	0.465	0.929	0.991
	Model 5	1.66 (1.25–2.20)	42.7	0.240	0.636	0.922
	Model 6	2.23 (1.20–4.14)	55.4	0.105	0.991	0.994
CML	Model 1	1.54 (1.18–2.01)	7.2	0.423	0.778	0.973
	Model 3	2.58 (1.57–4.24)	51.4	0.016	0.919	0.880
	Model 5	1.37 (1.06–1.77)	62.5	0.756	0.955	0.996
	Model 6	2.41 (1.45–4.00)	55.8	0.033	0.952	0.953

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
Sensitivity analysis						
Quality score≥10						
Overall	Model 1	1.56 (1.28–1.91)	26.3	0.352	0.045	0.413
	Model 3	2.41 (1.76–3.29)	55.3	0.001	0.021	0.002
	Model 4	1.27 (1.09–1.48)	48.3	0.983	0.691	0.986
	Model 5	1.42 (1.22–1.65)	51.7	0.763	0.006	0.201
	Model 6	2.17 (1.61–2.94)	57.8	0.009	0.063	0.033
Indian	Model 1	1.92 (1.18–3.12)	52.9	0.159	0.981	0.993
	Model 3	3.16 (1.90–5.25)	0.0	0.002	0.816	0.412
	Model 6	2.83 (1.73–4.64)	0.0	0.006	0.863	0.674
Caucasian	Model 4	1.28 (1.04–1.57)	33.6	0.936	0.950	0.997
	Model 5	1.34 (1.13–1.58)	11.5	0.910	0.354	0.947
Adults	Model 1	1.43 (1.16–1.76)	0.0	0.674	0.521	0.956
	Model 2	1.31 (1.08–1.60)	48.8	0.908	0.899	0.995
	Model 3	2.40 (1.61–3.58)	59.9	0.011	0.626	0.463
	Model 4	1.37 (1.16–1.61)	35.8	0.864	0.132	0.843
	Model 5	1.51 (1.29–1.77)	37.4	0.467	0.001	0.022
	Model 6	2.13 (1.43–3.18)	64.7	0.043	0.834	0.875
Adults and children	Model 6	1.94 (1.04–4.36)	43.6	0.267	0.998	0.999
HC	Model 1	1.63 (1.21–2.18)	34.9	0.288	0.774	0.961
	Model 3	2.56 (1.60–4.10)	55.0	0.013	0.875	0.798
	Model 5	1.41 (1.11–1.80)	63.6	0.690	0.894	0.992
	Model 6	2.39 (1.50–3.80)	58.7	0.024	0.904	0.890
NBDC	Model 1	1.47 (1.12–1.94)	15.7	0.557	0.921	0.992
	Model 2	1.24 (1.06–1.45)	0.0	0.991	0.876	0.995
	Model 3	2.17 (1.43–3.29)	54.9	0.041	0.865	0.893
	Model 4	1.30 (1.12–1.50)	0.0	0.975	0.251	0.931
	Model 5	1.41 (1.19–1.66)	17.2	0.771	0.046	0.622
	Model 6	1.85 (1.27–2.69)	51.3	0.136	0.904	0.967
Matching	Model 1	1.60 (1.29–1.99)	0.0	0.281	0.079	0.491
	Model 3	2.57 (1.61–4.12)	64.2	0.463	0.999	0.999
	Model 4	1.31 (1.09–1.58)	41.0	0.922	0.837	0.992
	Model 5	1.46 (1.24–1.73)	33.8	0.623	0.019	0.367
	Model 6	2.33 (1.44–3.76)	69.6	0.036	0.937	0.942
Nonmatching	Model 1	1.51 (1.04–2.19)	48.0	0.486	0.984	0.997
	Model 3	2.26 (1.45–3.53)	46.7	0.036	0.904	0.915
	Model 5	1.37 (1.03–1.82)	65.1	0.734	0.976	0.998
	Model 6	2.04 (1.40–2.98)	37.7	0.056	0.802	0.876
ALL	Model 1	1.92 (1.28–2.86)	29.6	0.112	0.922	0.969
	Model 3	3.10 (1.48–6.49)	58.1	0.027	0.990	0.988
	Model 4	1.43 (1.07–1.91)	39.3	0.627	0.961	0.996
	Model 5	1.59 (1.18–2.14)	45.8	0.350	0.863	0.980
	Model 6	2.66 (1.38–5.15)	54.1	0.045	0.988	0.989

(Continued on following page)

TABLE 9 (*Continued*) Credibility of the current meta-analysis.

Variables	Model	OR (95%CI)	I2 (%)	Statistical power	Credibility	
					Prior probability of 0.001	
					FPRP	BFDP
CML	Model 1	1.55 (1.15–2.09)	16.4	0.415	0.907	0.988
	Model 3	2.39 (1.37–4.16)	54.3	0.050	0.976	0.981
	Model 5	1.38 (1.03–1.83)	66.2	0.719	0.972	0.997
	Model 6	2.21 (1.26–3.87)	27.7	0.088	0.984	0.990
The combined effects of GSTM1 and GSTP1 polymorphisms						
Overall	Model 4	1.95 (1.35–2.80)	21.5	0.078	0.793	0.897
	Model 6	1.95 (1.37–2.77)	30.4	0.071	0.729	0.857
Indian	Model 4	1.72 (1.10–2.70)	33.5	0.276	0.985	0.996
	Model 6	1.65 (1.14–2.40)	19.6	0.309	0.966	0.993
HC	Model 4	1.82 (1.21–2.74)	25.9	0.177	0.959	0.987
	Model 6	1.88 (1.23–2.89)	41.8	0.152	0.964	0.987
Matching	Model 6	2.20 (1.25–3.89)	37.1	0.094	0.986	0.992
Non-matching	Model 4	2.07 (1.34–3.20)	0.0	0.074	0.935	0.964
	Model 5	1.44 (1.08–1.92)	0.0	0.610	0.955	0.995
	Model 6	1.76 (1.05–2.96)	42.1	0.273	0.992	0.997
ALL	Model 4	1.86 (1.01–3.43)	51.9	0.245	0.995	0.998
	Model 6	1.92 (1.30–2.83)	0.0	0.106	0.902	0.960
CML	Model 4	2.08 (1.27–3.40)	1.4	0.096	0.973	0.986
Sensitivity analysis						
HWE and Quality score≥10						
Overall	Model 4	1.95 (1.35–2.80)	21.5	0.078	0.793	0.897
	Model 6	1.95 (1.37–2.77)	30.4	0.071	0.729	0.857
The combined effects of GSTT1 and GSTP1 polymorphisms						
Overall	Model 3	1.50 (1.04–2.15)	59.8	0.500	0.982	0.997
	Model 4	4.24 (2.49–7.24)	0.0	0.000	0.632	0.027
	Model 5	1.70 (1.30–2.22)	32.2	0.179	0.352	0.765
	Model 6	3.31 (1.85–5.91)	14.8	0.004	0.933	0.780
Indian	Model 3	1.65 (1.05–2.59)	61.9	0.339	0.989	0.997
	Model 4	4.39 (2.51–7.68)	0.0	0.000	0.721	0.049
	Model 5	1.91 (1.45–2.50)	0.8	0.039	0.059	0.106
	Model 6	3.39 (1.94–5.94)	7.8	0.002	0.901	0.617
HC	Model 3	1.50 (1.04–2.15)	59.8	0.500	0.982	0.997
	Model 4	4.24 (2.49–7.24)	0.0	0.000	0.632	0.027
	Model 5	1.70 (1.30–2.22)	32.2	0.179	0.352	0.765
	Model 6	3.31 (1.85–5.91)	14.8	0.004	0.933	0.780
Matching	Model 4	4.61 (1.64–12.97)	16.8	0.017	0.996	0.994
	Model 5	1.40 (1.04–1.89)	0.0	0.674	0.976	0.998
CML	Model 2	1.91 (1.35–2.68)	0.0	0.081	0.690	0.849
	Model 4	3.29 (1.37–7.89)	1.9	0.039	0.995	0.995
	Model 5	1.61 (1.05–2.47)	50.4	0.373	0.987	0.997
	Model 6	2.40 (1.21–14.26)	31.8	0.303	0.999	0.999
Sensitivity analysis						
HWE and Quality score≥10						
Overall	Model 3	1.50 (1.04–2.15)	59.8	0.500	0.982	0.997
	Model 4	4.24 (2.49–7.24)	0.0	0.000	0.632	0.027
	Model 5	1.70 (1.30–2.22)	32.2	0.179	0.352	0.765
	Model 6	3.31 (1.85–5.91)	14.8	0.004	0.933	0.780

negative ones. Second, the mechanism of leading to leukemia is greatly sophisticated, and thus a single-gene mutation is not likely to generate remarkably to its development. Third, no consideration was given to if the genotype distribution of *GSTM1* and *GSTT1* polymorphisms in control group was in HWE because we could not calculate the HWE on these two genes. Fourth, the heterogeneity of *GSTM1*, *GSTT1,* and *GSTP1* was large; therefore, the random-effect model was selected, and after subgroup and sensitivity analysis, no source of heterogeneity was found. Hence, the current meta-analysis with a large sample size and enough subgroups will be conducive to confirm our discoveries.

This meta-analysis strongly suggests that only a minority of meaningful associations are credible results. Hence, larger-scale investigations of this topic should be performed in the future to verify or rebut our findings.

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

YZ: research design and performance, data collection, data analysis, and manuscript-writing; DW and C-YZ: data collection; Y-JL, X-HW, M-YS, and WW: data recheck; and X-LS and X-FH: research design and manuscript review.

Acknowledgments

We would like to sincerely thank the authors of the original research studies included in this study.

Conflict of interest

Author WW was employed by the company Beijing Zhendong Guangming Pharmaceutical Research Institute.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.976673/full#supplementary-material

References

Alves, S., Amorim, A., Ferreira, F., Norton, L., and Prata, M. J. (2002)., 16. PubMed, 15651565–15651567. doi:10.1038/sj.leu.2402543The GSTM1 and GSTT1 genetic polymorphisms and susceptibility to acute lymphoblastic leukemia in children from north Portugal*Leukemia*

Arruda, V. R., Lima, C. S. P., Grignoli, C. R. E., de Melo, M. B., Lorand-Metze, I., Alberto, F. L., et al. (2001). Increased risk for acute myeloid leukaemia in individuals with glutathione S-transferase mu 1 (GSTM1) and theta 1 (GSTT1) gene defects. *Eur. J. Haematol.* 66, 383–388. doi:10.1034/j.1600-0609.2001.066006383.x

Baker, W. L., White, C. M., Cappelleri, J. C., Kluger, J., Coleman, C. I., and Health, O (2009). Understanding heterogeneity in meta-analysis: The role of meta-regression. *Int. J. Clin. Pract.* 63 (10), 1426–1434. doi:10.1111/j.1742-1241.2009.02168.x

Begg, C. B., and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50 (4), 1088–1101. doi:10.2307/2533446

Das, P., Shaik, A. P., and Bammidi, V. K. (2009). Meta-analysis study of glutathione-S-transferases (GSTM1, GSTP1, and GSTT1) gene polymorphisms and risk of acute myeloid leukemia. *Leuk. Lymphoma* 50 (8), 1345–1351. doi:10.1080/10428190903003236

Der Simonian, R., and Laird, N. (2015). Meta-analysis in clinical trials revisited. *Contemp. Clin. Trials* 45, 139–145. doi:10.1016/j.cct.2015.09.002

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315 (7109), 629–634. doi:10.1136/bmj.315.7109.629

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

Harries, L. W., Stubbins, M. J., Forman, D., Howard, G. C., and Wolf, C. R. (1997). Identification of genetic polymorphisms at the glutathione S-transferase Pi locus and association with susceptibility to bladder, testicular and prostate cancer. *Carcinogenesis* 18, 641–644. doi:10.1093/carcin/18.4.641

He, H. R., You, H. S., Sun, J. Y., Hu, S. S., Ma, Y., Dong, Y. L., et al. (2014). Glutathione S-transferase gene polymorphisms and susceptibility to acute myeloid leukemia: meta-analyses. *Jpn. J. Clin. Oncol.* 44 (11), 1070–1081. doi:10.1093/jjco/hyu121

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327 (7414), 557–560. doi:10.1136/bmj.327.7414.557

Hollman, A. L., Tchounwou, P. B., and Huang, H. C. (2016). The association between gene-environment interactions and diseases involving the human GST superfamily with SNP variants. *Int. J. Environ. Res. Public Health* 13, 379. doi:10.3390/ijerph13040379

Huang, G. Z., Shan, W., Zeng, L., and Huang, L. G. (2013). The GSTP1 A1578G polymorphism and the risk of childhood acute lymphoblastic leukemia: Results from an updated meta-analysis. *Genet. Mol. Res.* 12 (3), 2481–2491. doi:10.4238/2013.July.24.3

Ioannidis, J. P., Boffetta, P., Little, J., O'Brien, T. R., Uitterlinden, A. G., Vineis, P., et al. (2008). Assessment of cumulative evidence on genetic

associations: Interim guidelines. *Int. J. Epidemiol.* 37 (1), 120–132. doi:10.1093/ije/dym159

Krajinovic, M., Labuda, D., and Sinnett, D. (2001). Childhood acute lymphoblastic leukemia: Genetic determinants of susceptibility and disease outcome. *Rev. Environ. Health* 16, 263–279. doi:10.1515/reveh.2001.16.4.263

Ma, Y., Sui, Y., Wang, L., and Li, H. (2014). Effect of GSTM1 null genotype on risk of childhood acute leukemia: A meta-analysis. *Tumour Biol.* 35 (1), 397–402. doi:10.1007/s13277-013-1055-x

Maia Rda, R., and Wünsch Filho, V. (2013). Infection and childhood leukemia: Review of evidence. *Rev. Saude Publica* 47, 1172–1185. doi:10.1590/s0034-8910.2013047004753

Moulik, N. R., Parveen, F., Kumar, A., and Agrawal, S. (2014). Glutathione-S-transferase polymorphism and acute lymphoblastic leukemia (ALL) in north Indian children: A case-control study and meta-analysis. *J. Hum. Genet.* 59 (9), 529–535. doi:10.1038/jhg.2014.66

Ouerhani, S., Nefzi, M. A., Menif, S., Safra, I., Douzi, K., Fouzai, C., et al. (2011). Influence of genetic polymorphisms of xenobiotic metabolizing enzymes on the risk of developing leukemia in a Tunisian population. *Bull. Cancer* 98 (12), 95–106. doi:10.1684/bdc.2011.1502

Pearson, W. R., Vorachek, W. R., Xu, S. J., Berger, R., Hart, I., Vannais, D., et al. (1993). Identification of class-mu glutathione transferase genes GSTM1-GSTM5 on human chromosome 1p13. *Am. J. Hum. Genet.* 53, 220–233.

Ryberg, D., Skaug, V., Hewer, A., Phillips, D. H., Wolf, C. R., Ogreid, D, et al. (1997). Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. *Carcinogenesis* 18, 1285–1289. doi:10.1093/carcin/18.7.1285

Schüz, J., and Erdmann, F. (2016). Environmental exposure and risk of childhood leukemia: An overview. *Arch. Med. Res.* 47, 607–614. doi:10.1016/j.arcmed.2016.11.017

Strange, R. C., and Fryer, A. A. (1999). The glutathione S-transferases: Influence of polymorphism on cancer susceptibility. *IARC Sci. Pub*, 231–249.

Strange, R. C., Lear, J. T., and Fryer, A. A. (1998). Polymorphism in glutathione S-transferase loci as a risk factor for common cancers. *Arch. Toxicol. Suppl.* 20, 419–428. doi:10.1007/978-3-642-46856-8_37

Strange, R. C., Spiteri, M. A., Ramachandran, S., and Fryer, A. A. (2001). Glutathione-S-transferase family of enzymes. *Mutat. Res.* 482, 21–26. doi:10.1016/s0027-5107(01)00206-8

Tang, Q., Li, J., Zhang, S., Yuan, B., Sun, H., Wu, D., et al. (2013). GSTM1 and GSTT1 null polymorphisms and childhood acute leukemia risk: Evidence from 26 case-control studies. *PLoS One* 8 (10), e78810. doi:10.1371/journal.pone.0078810

Tang, Z. H., Zhang, C., Cheng, P., Sun, H. M., Jin, Y., Chen, Y. J., et al. (2014). Glutathione-S-transferase polymorphisms (GSTM1, GSTT1 and GSTP1) and acute leukemia risk in Asians: A meta-analysis. *Asian pac. J. Cancer Prev.* 15 (5), 2075–2081. doi:10.7314/apjcp.2014.15.5.2075

Theodoratou, E., Montazeri, Z., Hawken, S., Allum, G. C., Gong, J., Tait, V., et al. (2012). Systematic meta-analyses and field synopsis of genetic association studies in colorectal cancer. *J. Natl. Cancer Inst.* 104 (19), 1433–1457. doi:10.1093/jnci/djs369

Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. (2004). Assessing the probability that a positive report is false: An approach for molecular epidemiology studies. *J. Natl. Cancer Inst.* 96 (6), 434–442. doi:10.1093/jnci/djh075

Wakefield, J. (2007). A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* 81 (32), 208–227. doi:10.1086/519024

Wang, J., Wu, D., and Sun, A. (2019). Effects of GST null genotypes on individual susceptibility to leukemia: A meta-analysis. *Exp. Mol. Pathol.* 108, 137–142. doi:10.1016/j.yexmp.2019.01.004

Webb, G., Vaska, V., Coggan, M., and Board, P. (1996). Chromosomal localization of the gene for the human theta class glutathione transferase (GSTT1). *Genomics* 33, 121–123. doi:10.1006/geno.1996.0167

Ye, Z., and Song, H. (2005). Glutathione s-transferase polymorphisms (GSTM1, GSTP1 and GSTT1) and the risk of acute leukaemia: A systematic review and meta-analysis. *Eur. J. Cancer* 41 (7), 980–989. doi:10.1016/j.ejca.2005.01.014

Zhang, H. Y., Zhang, J., Wu, T., and Bai, H. (2017). [Polymorphism of glutathione S-transferases and genetic sensitivity of childhood acute lymphoblastic leukemia: A meta-analysis]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi* 25 (1), 16–23. doi:10.7534/j.issn.1009-2137.2017.01.003

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fgene.2022.1033214

OPEN ACCESS

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

REVIEWED BY
Ning Wang,
Arcus Biosciences, United States
Arvind Mer,
University of Ottawa, Canada

*CORRESPONDENCE
Xingdong Chen,
xingdongchen@fudan.edu.cn

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

RECEIVED 31 August 2022
ACCEPTED 28 October 2022
PUBLISHED 08 November 2022

CITATION
Wang D, Dai J, Suo C, Wang S, Zhang Y and Chen X (2022), Molecular subtyping of esophageal squamous cell carcinoma by large-scale transcriptional profiling: Characterization, therapeutic targets, and prognostic value.
Front. Genet. 13:1033214.
doi: 10.3389/fgene.2022.1033214

Molecular subtyping of esophageal squamous cell carcinoma by large-scale transcriptional profiling: Characterization, therapeutic targets, and prognostic value

Danke Wang[1], Jiacheng Dai[1], Chen Suo[2,3,4], Shangzi Wang[1], Yuting Zhang[1] and Xingdong Chen[1,2,5,6]*

[1]State Key Laboratory of Genetic Engineering, Human Phenome Institute, School of Life Sciences, Fudan University, Shanghai, China, [2]Fudan University Taizhou Institute of Health Sciences, Taizhou, China, [3]Department of Epidemiology, School of Public Health, Fudan University, Shanghai, China, [4]Shanghai Institute of Infectious Disease and Biosecurity, Shanghai, China, [5]Department of Neurology, Huashan Hospital, Fudan University, Shanghai, China, [6]Yiwu Research Institute of Fudan University, Yiwu, Zhejiang, China

The tumor heterogeneity of the transcriptional profiles is independent of genetic variation. Several studies have successfully identified esophageal squamous cell carcinoma (ESCC) subtypes based on the somatic mutation profile and copy number variations on the genome. However, transcriptome-based classification is limited. In this study, we classified 141 patients with ESCC into three subtypes (Subtype 1, Subtype 2, and Subtype 3) *via* tumor sample gene expression profiling. Differential gene expression (DGE) analysis of paired tumor and normal samples for each subtype revealed significant difference among subtypes. Moreover, the degree of change in the expression levels of most genes gradually increased from Subtype 1 to Subtype 3. Gene set enrichment analysis (GSEA) identified the representative pathways in each subtype: Subtype 1, abnormal Wnt signaling pathway activation; Subtype 2, inhibition of glycogen metabolism; and Subtype 3, downregulation of neutrophil degranulation process. Weighted gene co-expression network analysis (WGCNA) was used to elucidate the finer regulation of biological pathways and discover hub genes. Subsequently, nine hub genes (*CORO1A, CD180, SASH3, CD52, CD300A, CD14, DUSP1, KIF14, and MCM2*) were validated to be associated with survival in ESCC based on the RNA sequencing (RNA-seq) data from The Cancer Genome Atlas (TCGA) database. The clustering analysis of ESCC granted better understanding of the molecular characteristics of ESCC and led to the discover of new potential therapeutic targets that may contribute to the clinical treatment of ESCC.

KEYWORDS

ESCC, gene expression profile, subtype, integrate, GSEA, WGCNA

1 Introduction

In 2020, esophageal carcinoma (EC) was the seventh most common cancer worldwide with 604,000 new cases, contributing 3.1% of the total new cancer cases, and was ranked sixth in mortality worldwide (544,000 deaths) (Sung et al., 2021). Esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC) are the two main EC subtypes (Siewert and Ott, 2007; Della Guardia et al., 2022), with ESCC counting for approximately 90% of EC cases worldwide (Smyth et al., 2017). The development of next-generation sequencing technologies has yielded a deeper understanding of ESCC genomic features *via* sequencing and analysis the genomes and transcriptomes of millions of patients with ESCC. The analyses revealed that ESCC has extensive inter- and intra-tumor heterogeneity (Hu et al., 2009; Lin et al., 2018).

Regarding tumor heterogeneity, several studies (Liu et al., 2016; Cancer Genome Atlas Research Network et al., 2017) have identified ESCC subtypes based on the somatic mutation profile and copy number variation on the genome. However, recent single-cell RNA sequencing (RNA-seq) studies have demonstrated that cancer cell state heterogeneity is largely independent of genetic variation (Halbritter et al., 2019; Guo et al., 2020; LaFave et al., 2020; Marjanovic et al., 2020). The transcriptional landscape is reprogrammed with cancer progression, metastasis, and therapy resistance (Hanahan and Weinberg, 2011; Quintanal-Villalonga et al., 2020). Therefore, identifying ESCC subtypes based on the gene expression profile of the tumor sample can reveal some molecular features that cannot be detected in genome-based classification. There are many successful applications for identifying tumor subtypes based on the gene expression profile, including that for colon cancer (Marisa et al., 2013; Guinney et al., 2015), non-small cell lung cancer (Chen et al., 2017) and uterine leiomyosarcoma (An et al., 2017). These transcriptional profile-based classification studies revealed clinically valuable targets. Attempts have also been made to classify ESCC based on its gene expression profile. Wang et al. (Wang et al., 2019) have classified Asian patients with ESCC into two subtypes; the selected genes were clustered and only genes with large standard deviations in the ESCC cohort were selected. Nevertheless, classification based on selected genes may bias the result, neglecting genes that are less varied but that are important in overall regulation. Therefore, it is necessary to characterize ESCC subtypes considering an unbiased transcriptome level.

In this study, we classified 141 patients with ESCC into three subtypes based on the gene expression profile of the patients' tumor samples. The differences in individual gene expression level changes among the three subtypes were identified with differential gene expression (DGE) analysis. Gene set enrichment analysis (GSEA) and weighted gene co-expression network analysis (WGCNA) were used to explore how the gene expression levels co-varied together. *Via* this analysis series, we clearly described the molecular characteristics of each subtype. We discovered important genes and the biological pathways that may affect ESCC prognosis. Our study provides an in-depth understanding of ESCC molecular features and demonstrates potential targets for ESCC clinical treatment.

2 Materials and methods

2.1 Data collection and quality control

The raw microarray gene expression data from 141 ESCC patient tumors and the paired normal samples across seven datasets were obtained from the Gene Expression Omnibus (GEO). The dataset inclusion criteria were: 1) gene expression data from paired tumor and normal tissue samples were available; 2) the patients had not undergone previous treatment. The following datasets were included in this study: GSE17351, GSE20347, GSE23400, GSE38129, GSE77861, GSE161533, and GSE100942 (Table 1) (Hu et al., 2010; Lee et al., 2010; Su et al., 2011; Hu et al., 2015; Erkizan et al., 2017; Ming et al., 2018). Among the seven datasets, the samples in GSE77861 were obtained from African American patients and samples in the remaining six datasets were from Asian patients.

The Affymetrix microarray data were robust multiarray averaging (RMA) normalized (background processing, log2 transformation, quantile scaling, and probe expression measurement) in the R package affy (Gautier et al., 2004). Then, all available biological and technical covariates except for the diagnostic group were regressed from each individual expression dataset. After the above preprocessing had been performed on each dataset, all seven datasets were merged. The batch effect was corrected with the ComBat function of the R sva package (Leek et al., 2012). Outliers were identified with the principal component analysis (PCA) in the R package FactoMineR.

The RNA-seq data of tumor tissue samples from patients with ESCC were downloaded from The Cancer Genome Atlas (TCGA) database using the R package TCGAbiolinks (Colaprico et al., 2016). The screening and elimination yielded the gene expression profiling data of tumor tissues from 80 patients. The TPM (transcripts per million) data that normalized gene length and sequencing depth were used for subsequent analysis.

2.2 Clustering

The function hclust was used to hierarchical clustering the 141 ESCC tumor samples through gene expression profiling. We used the Euclidean method to calculate the Euclidean distance between samples and the ward.D method to cluster the 141 samples. The parameters of the clustering based on the K-means method were centers = 3, nstart = 25.

TABLE 1 Information of the GEO datasets involved in this study.

GEO accession	Platforms	Sample size	Race	PMID
GSE17351	GPL570	10	Japanese	20042640
GSE20347	GPL571	34	Chinese	20955586
GSE23400	GPL96	106	Chinese	21385931
GSE38129	GPL571	60	Chinese	26409826
GDE77861	GPL570	14	African American	28629367
GSE161533	GPL570	56	Chinese	—
GSE100942	GPL570	8	Chinese	29290801

2.3 Differential gene expression analysis

For each subtype, DGE analysis was performed between paired tumor and normal samples, using the R package limma (Ritchie et al., 2015). The differential expression genes (DEGs) were identified as adj.P.Val cutoff <0.05 (Benjamini Hochberg false discovery rate (FDR) correction) (Ge et al., 2021) (FDR <0.05) and $|\log_2$ fold change ($\log_2$FC)| > 1.

2.4 Gene set enrichment analysis and immune cell infiltration analysis

GSEA (Subramanian et al., 2005) was performed with the function gseGO in the R package clusterProfiler (Yu et al., 2012) based on the Gene Ontology (GO) database. GSEA was performed on the three ESCC subtypes and the $\log_2$FC of each gene was used as the basis for the gene ranking. Significantly enriched pathways (FDR <0.05) were available.

The gene expression profile data of the 141 ESCC tumor samples were combined for immune cell infiltration analysis using single-sample GSEA (ssGSEA) in R package GSVA (Hanzelmann et al., 2013). The immune cell marker genes constituted the background gene set for the immune infiltration analysis (Charoentong et al., 2017).

2.5 Weighted gene co-expression network analysis

Network analysis was performed with the WGCNA package (Langfelder and Horvath, 2008) in R. Approximate scale-free topology (R^2 > 0.85) was achieved using a soft threshold power of 7. The network was constructed using all 13,515 gene expression profiling data from the 141 paired ESCC tumor and normal samples. The other network construction parameters were maxBlockSize = 5,000, minModuleSize = 30, TOMType = "unsigned", reassignThreshold = 0, and mergeCutHeight = 0.25.

2.6 Enrichment analysis

GO enrichment analysis was performed with the function enrichGO in the R package clusterProfiler based on the GO database. The significantly enriched pathways had FDR <0.05.

Cell type enrichment analysis was performed based on the marker genes (Xu et al., 2021) of different cell types. Significantly enriched cell types were obtained with $p < 0.05$ (Fisher's exact test).

2.7 Survival analysis

Survival analysis was performed with the RNA-seq data of 80 ESCC tumor tissue samples from TCGA database. Survival analysis and survival curve plotting were performed using the R packages survival and survminer, respectively. For each gene among the 80 samples, samples with expression levels higher than the median value were defined as the high-expression group and those below the median value were defined as the low-expression group.

3 Results

3.1 Characteristics of the mRNA microarray data and analysis pipeline

The analysis pipeline is depicted in Figure 1. After strict data preprocessing, the seven ESCC mRNA microarray datasets were merged, and batch effects were corrected using the function ComBat. PCA was performed with batches as groups before and after batch correction. There was a large distance between the datasets before batch correction (Supplementary Figure S1A), and the data distribution was uniform after batch correction (Supplementary Figure S1B). The boxplots of samples grouped by batch before and after batch correction also reflected this change (Supplementary Figures S1C, D). The results suggested that the batch effects among the seven datasets were eliminated. We used the combined dataset for subsequent analysis.

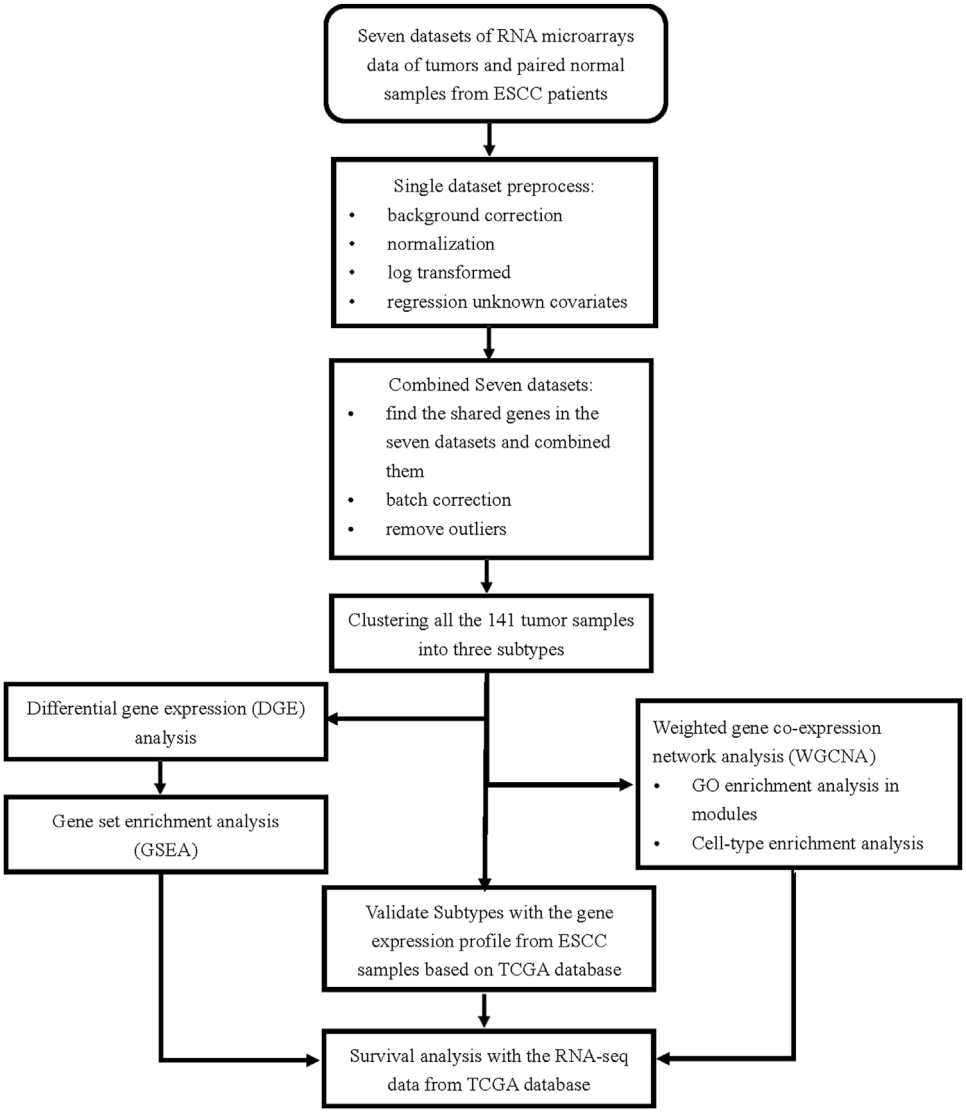


FIGURE 1
The pipeline of this analysis.

When PCA was performed based on tumor and normal grouping, three normal samples which were abnormally grouped with the tumor samples (Supplementary Figure S1E). Therefore, we removed these three samples, which included the match normal (GSM573851, GSM573889, and GSM573852) and tumor samples (GSM573904, GSM573942, and GSM573905). The remaining 282 samples were divided into tumor and normal groups (Supplementary Figure S1F), a distinct transcriptomic pattern was indicated between the two groups.

The 13,515-gene expression profile data of the 141 paired ESCC tumor and normal samples were included in the study. The ESCC subtypes were identified using the gene expression profiles of the 141 ESCC tumor samples. The analyses mainly included: 1) hierarchical clustering of the 141 ESCC tumor sample; 2) DGE analysis of paired tumor and normal samples of each subtype to determine gene expression level changes; 3) GSEA and WGCNA to identify biological pathway regulation and discover hub genes; 4) survival analysis of the genes found in 3) based on the RNA-seq data from 80 ESCC tumor samples in the TCGA database; 5) validation of the subtypes found in 1) using the TCGA RNA-seq data.

3.2 The results of clustering

Hierarchical clustering of the 141 ESCC tumor sample transcriptome profiles revealed a total of three subtypes,

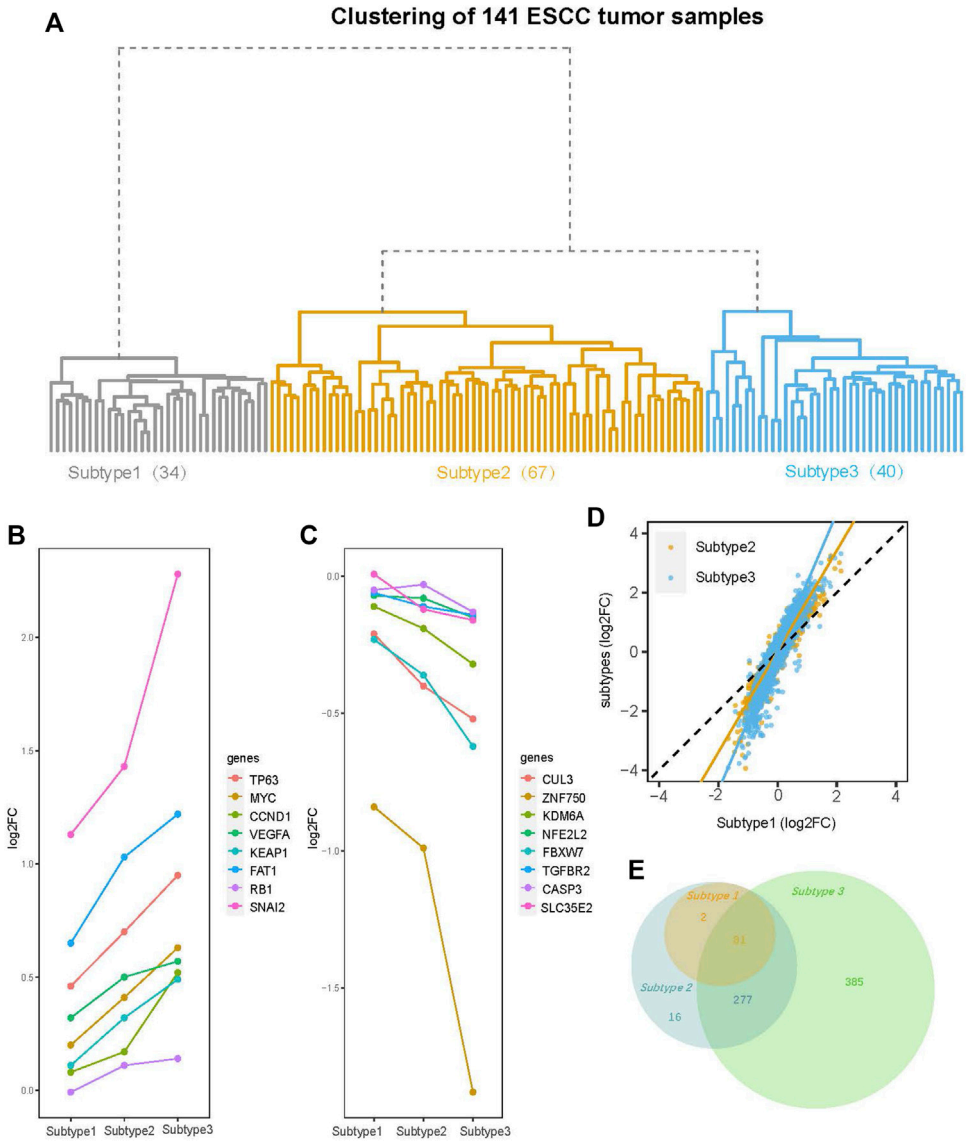


FIGURE 2
Clustering tree of the 141 ESCC tumor samples and an overview of differential gene expression (DGE) analysis. **(A)** The hierarchical clustering tree diagram of 141 ESCC tumor samples constructed by the gene expression profile. The 141 ESCC tumor samples were clustered into three subtypes (Subtype 1, Subtype 2, and Subtype 3), with sample sizes of 34, 67, and 40 for Subtype 1 to Subtype 3, respectively. **(B)** The $\log_2$ fold-change ($\log_2$FC) of the genes (TP63, MYC, CCND1, VEGFA, KEAP1, FAT1, RB1, and SNAI2) with amplification mutations on ESCC genome are increased gradually from Subtype 1 to Subtype 3. **(C)** The $\log_2$FC of ESCC genomic insertion or deletion mutant genes (CUL3, ZNF750, KDM6A, NFE2L2, FBXW7, TGFBR2 CASP3, and SLC35E2) decreased gradually from Subtype 1 to Subtype 3. **(D)** Comparison of $\log_2$FC of all the 13515 genes between subtypes. Among three subtypes, the degree of change in the expression levels of most genes is: Subtype 3 > Subtype 2 > Subtype 1. **(E)** Venn diagram shows overlaps of differential expression genes (DEGs) (FDR <0.05, $|\log_2 FC| > 1$) among three subtypes. The numbers of DEGs include both up- and down-regulated genes in tumor tissue compared to normal tissue.

which we designated Subtype 1 (34 samples), Subtype 2 (67 samples), and Subtype 3 (40 samples). Subtype 1 was under an independent branch of the clustering tree and Subtypes 2 and 3 were two subgroups under the same branch (Figure 2A).

To further confirm that we have eliminated batch effects, we recolored the hierarchical clustering tree using batch as color label. The result showed that the samples from each study were distributed across the three subtypes (Supplementary Figure S2A). So, the subtypes are not driven by different studies.

To explore whether our subtyping is sensitive to the clustering method, we re-clustered the 141 samples into three subtypes (named as K-means_Subtype1, K-means_Subtype2, and K-means_Subtype3) using the K-means method. The

result showed that the three subtypes clustered by the K-means method were consistent with the subtypes obtained by the hierarchical clustering in more than 80% of the samples (Figure 2A, Supplementary Figure S2B). So, it can be concluded that the subtypes we obtained are not sensitive to the clustering method.

3.3 Results of differential gene expression analysis

DGE analysis was performed on the paired tumor and normal samples for each subtype. We first focused on the most commonly mutated genes in ESCC. The log_2FC of these genes exhibited a tendency change from Subtype 1 to Subtype 3. For example, the log_2FC of genes in the ESCC genome with amplification mutations, such as *MYC, TP63, CCND1, VEGFA,* and *SNAI2* (Song et al., 2014; Liu et al., 2016; Sawada et al., 2016; Cancer Genome Atlas Research Network et al., 2017), increased gradually from Subtype 1 to Subtype 3 (Figure 2B). The log_2FC of ESCC genomic insertion or deletion mutant genes [*CUL3, ZNF750, KDM6A, NFE2L2, and SLC35E2* (Song et al., 2014; Liu et al., 2016; Sawada et al., 2016; Cancer Genome Atlas Research Network et al., 2017)] decreased gradually from Subtype 1 to Subtype 3 (Figure 2C). Viewing of the total 13,515 genes, the degree of change in the expression levels of most genes gradually increased from Subtype 1 to Subtype 3 (Figure 2D).

The DEG numbers were greatly different across the three subtypes. Under the thresholds (FDR <0.05; $log_2FC \geq |1|$), Subtypes 1, 2, and 3 had 83, 376, and 743 total DEGs, respectively (Supplementary Figure S3). The dramatic increase in the DEG number from Subtype 1 to Subtype 3 suggested striking differences among the subtypes even if they were identified as the same cancer type. Further exploration of the relationship between the DEGs of the three subtypes revealed that the Subtype 2 DEGs included all Subtype 1 DEGs, while the Subtype 3 DEGs did not completely encompass the Subtypes 1 and 2 DEGs (Figure 2E). Eighteen DEGs were specific to Subtype 1 or 2 rather than Subtype 3 (Figure 2E, Supplementary Figure S4A). The 16 upregulated DEGs among these 18 genes are enriched in the immune-related pathways (Supplementary Figure S4B).

3.4 Results of gene set enrichment analysis and single-sample GSEA

In this study, we used GSEA to determine whether a set of genes involved in a biological pathway demonstrated statistically significant differences between tumor and normal status (Subramanian et al., 2005). GSEA of each subtype enabled the discovery of how various biological pathways were regulated in each subtype. The biological pathways commonly upregulated in the three subtypes were those for chromosomal DNA replication, endodermal cell differentiation, and collagen fibril organization (Figure 3A) (Supplementary Material S1). The commonly downregulated biological pathways included those for fatty acid oxidation metabolism, and keratinocyte differentiation (Figure 3A) (Supplementary Material S1). The shared biological pathway regulation among the three subtypes was consistent with the findings of previous ESCC studies. (Su et al., 2011; Erkizan et al., 2017).

There were representative enriched pathways in tumor samples of each subtype. Subtype 1 was characterized by significant Wnt signaling pathway upregulation (FDR <0.05) (Figure 3B). Glycogen metabolism downregulation was the hallmark of Subtype 2 (Figure 3C). Subtype 3 featured markedly inhibited neutrophil-mediated immunological pathways (FDR <0.05), in which downregulated neutrophil degranulation was the primary manifestation (Figure 3D).

The degree of immune cell infiltration in tumor samples from the three subtypes was assessed with ssGSEA. The results revealed fewer infiltrating T cells in Subtype 1 tumor samples as compared to those of Subtypes 2 and 3 and less neutrophil infiltration in Subtype 3 tumor tissues as compared with that of Subtypes 1 and 2 (Supplementary Figure S5).

3.5 Enrichment results of modules that constructed by weighted gene co-expression network analysis

We performed WGCNA to characterize the involved biological pathways more specifically. The genes regulated in the same pattern were clustered in one co-expression module based on the correlation coefficient weighted value. This approach fully accounted for the genes that change little but that may be important in overall regulation. All gene expression profiles of the 141 paired ESCC tumor and normal samples were considered in the co-expression network construction, which included a total of 14 modules, including module 0 (genes with irregular expression) (Supplementary Material S2).

The correlations between the modules and tumor are depicted in Figure 4A. Seven (modules 1, 3, 7, 8, 10, 11, and 13) and six (modules 0, 2, 4, 5, 9, and 12) modules were positively and negatively correlated with tumor, respectively (FDR <0.05). From Subtype 1 to Subtype 3, eight modules (module 0, 1, 2, 5, 7, 9, 10, and 11) demonstrated a gradually stronger correlation between modules and subtypes, and five modules (module 3, 4, 8, 12, and 13) presented stronger correlations from Subtype 1 to Subtype 2 and weakened correlations from Subtype 2 to Subtype 3. Module 6 was highly distinctive, demonstrating no significant correlation with Subtype 1 or 2, but demonstrating a remarkable negative correlation ($r = -0.26$, $p = 1E\text{-}05$) with Subtype 3.

Module 6 were enriched in a broad range of immune-related pathways. The biological pathways enriched in module 6 included B cell activation, T cell differentiation, lymphocyte

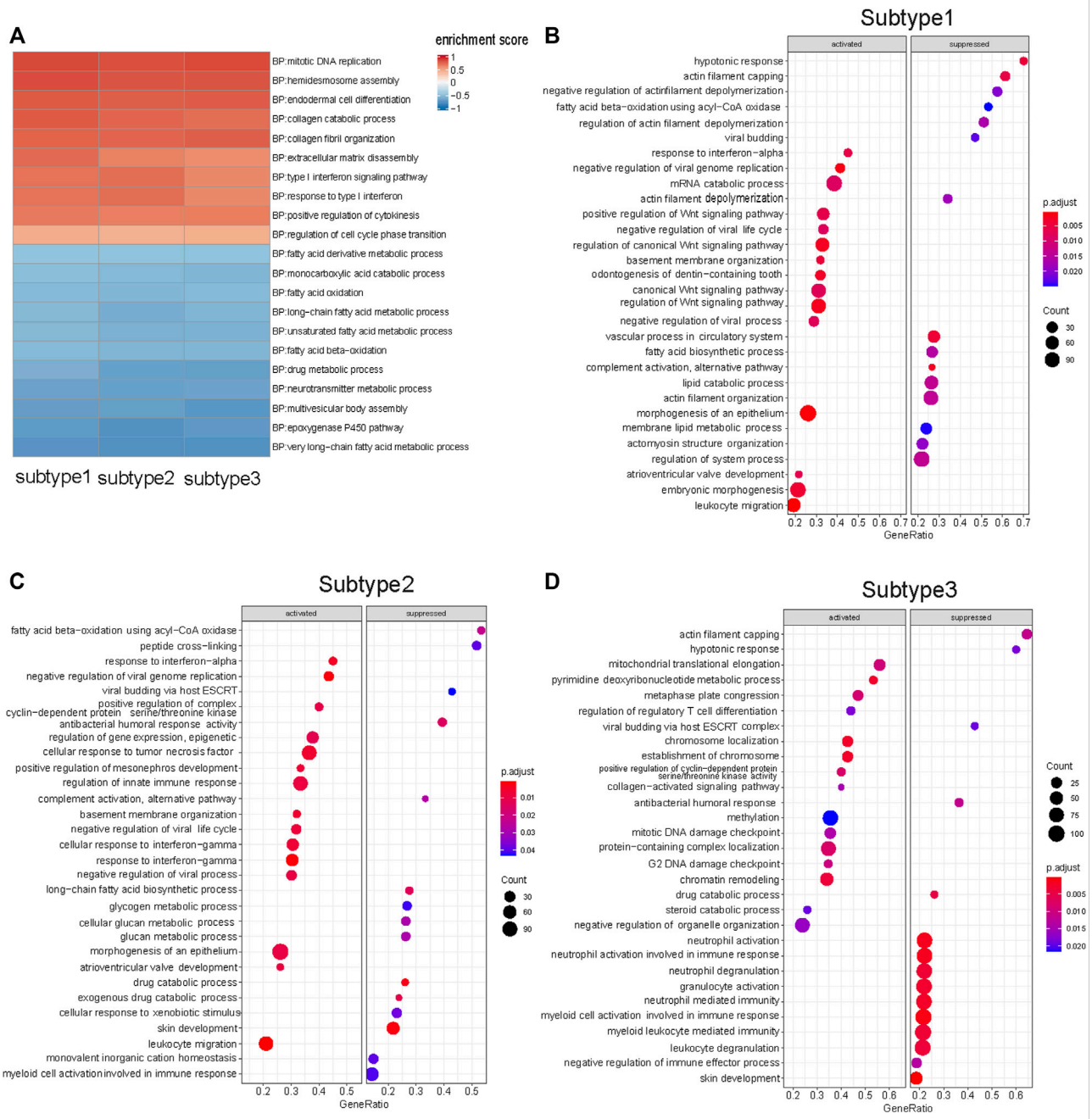


FIGURE 3
GSEA results of three subtypes. **(A)** Ten up-regulated and ten down-regulated biological pathways were selected from the shared enriched pathways in three subtypes tumor samples. **(B)** Top 15 biological pathways that specific up- and down-regulated in Subtype 1 tumor samples. Wnt signaling pathway was significantly up-regulated in Subtype 1. **(C)** Top 15 biological pathways that specific up- and down-regulated in Subtype 2 tumor samples. Glycogen metabolism biological pathway was significantly down-regulated in Subtype 2. **(D)** Top 15 biological pathways that specific up- and down-regulated in Subtype 3 tumor samples. Neutrophil-mediated immunological pathways were down-regulated in Subtype 3. (FDR <0.05).

differentiation and calcium homeostasis (FDR <0.05) (Figure 5A). The cell type enrichment analysis revealed that B cells and dendritic cells were enriched in module 6 (FDR <0.05) (Figure 4B).

Based on the above results, modules 0, 1, and 3 of neutrophil enrichment were the modules of interest (Figure 4B). In modules 1 and 3 (both were positively correlated with the three subtypes), the neutrophil degranulation process was among the top

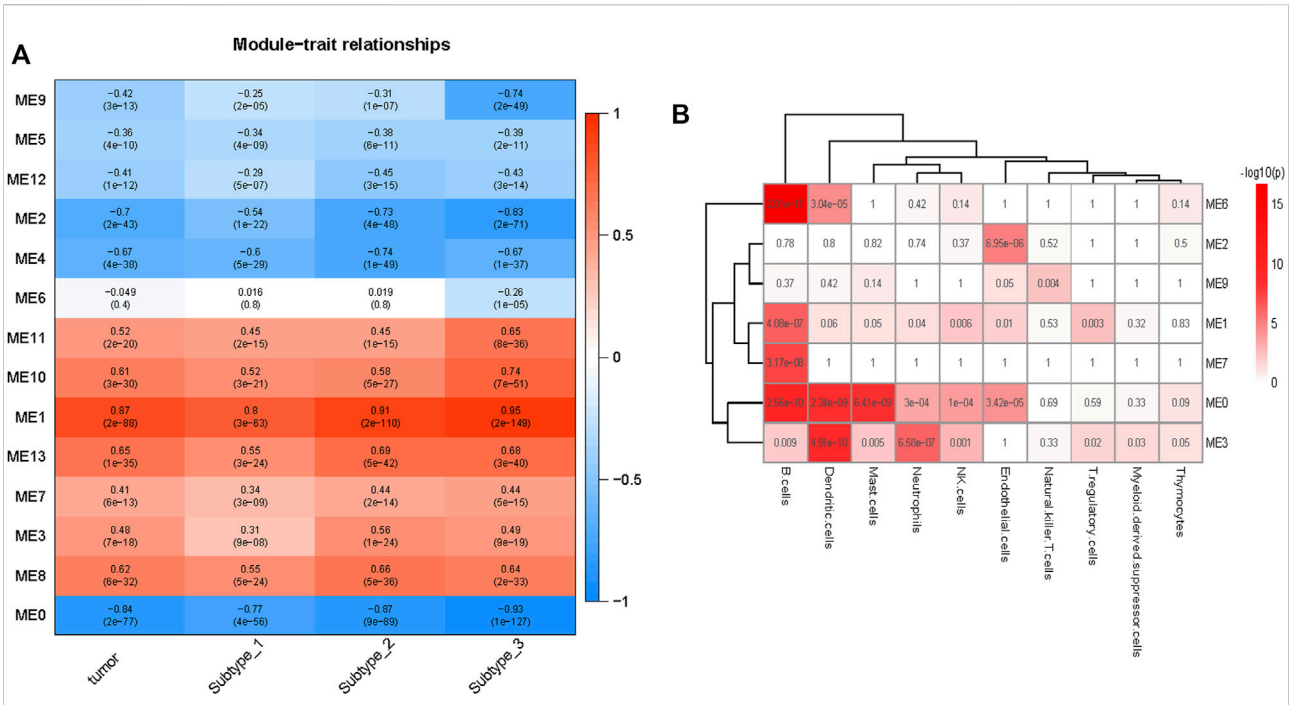


FIGURE 4
Overview of the modules constructed by WGCNA. **(A)** Correlations of modules to tumor. Modules 1, 3, 7, 8, 10,11, and 13 are positive correlated with tumor and modules 0, 2, 4, 5, 9, and 12 are negative correlated with tumor. Module 6 was negatively correlated with Subtype 3 only. **(B)** Enrichment of various cells in different modules (module0, 1, 2, 3, 6, 7, and 9), the *p* values (Fisher's exact test) are shown. B cells were enriched in modules 1, 3, and 6, the endothelial cells were enriched in modules 2 and 9.

30 enriched pathways (Figures 5B, C) (FDR <0.05). In addition, module 1 enriches cell cycle-related biological processes, and module 3 also enriches other immune-related biological pathways. The neutrophil chemotaxis pathway was enriched in module 0 (FDR <0.05) (Supplementary Material S3), which was negatively correlated with the three subtypes (correlation between module 0 and Subtype 3: r = −0.93, p = 1E-127).

Endothelial cells were enriched in modules 2 and 9 ($p \leq 0.05$) (Figure 4B), both of which were negatively correlated with tumor. The negative correlation increased gradually from Subtype 1 to Subtype 3. Biological pathways such as cellular rhythm, response to radiation, and response to oxidative stress were enriched in module 9 (Figure 5D) while, the muscle contraction and myofibril assembly biological pathways were enriched in module 2 (Supplementary Figure S6). We were able to conclude that endothelial cell contractility and responsiveness to external stimuli are affected in ESCC.

3.6 Results of survival analysis

Based on the above module enrichment results, we performed survival analysis of nine genes (*CORO1A*, *CD180*, *SASH3*, *CD52*, *CD300A*, *CD14*, *DUSP1*, *KIF14*, and *MCM2*) from the hub genes in these modules of interest and the genes involved in biological pathways that are important in tumor progression. We validated these genes by TCGA dataset survival analysis. The high levels of seven genes (CORO1A, CD180, SASH3, CD52, CD300A, CD14, and DUSP1) were related with poor survival in ESCC (Figures 6A–G) and KIF14 and MCM2 expression levels were positively correlated with better survival (Figures 6H, I) ($p < 0.05$). *CORO1A*, *CD180*, *SASH3*, and *CD52* were located in module 6, *CD300A* and *CD14* were involved in module 3, module 9 contained *DUSP1*, module 1 contained *KIF14*, and *MCM2* (Table 2).

3.7 Expression levels of the survival-related genes in the three subtypes

The expression levels of the nine survival-related genes in the three subtypes may indicate the survival of subtypes to a certain extent. The gene expression levels of four genes (CD180, SASH3, CD300A, and CD14) inversely associated with survival were significantly higher in Subtype 2 or 3 than in Subtype 1 (Figures 7B, C, E, F). The four genes are involved in the immune-related pathways. So, tumor immunity may be an important factor affecting survival time in Subtype 2 or 3.

Compared with Subtypes 2 and 3, *DUSP1* that negatively correlated with survival, has higher expression in Subtype 1 (Figure 7G), while *KIF14* and *MCM2*, which are positively correlated with survival, have lower expression in Subtype 1

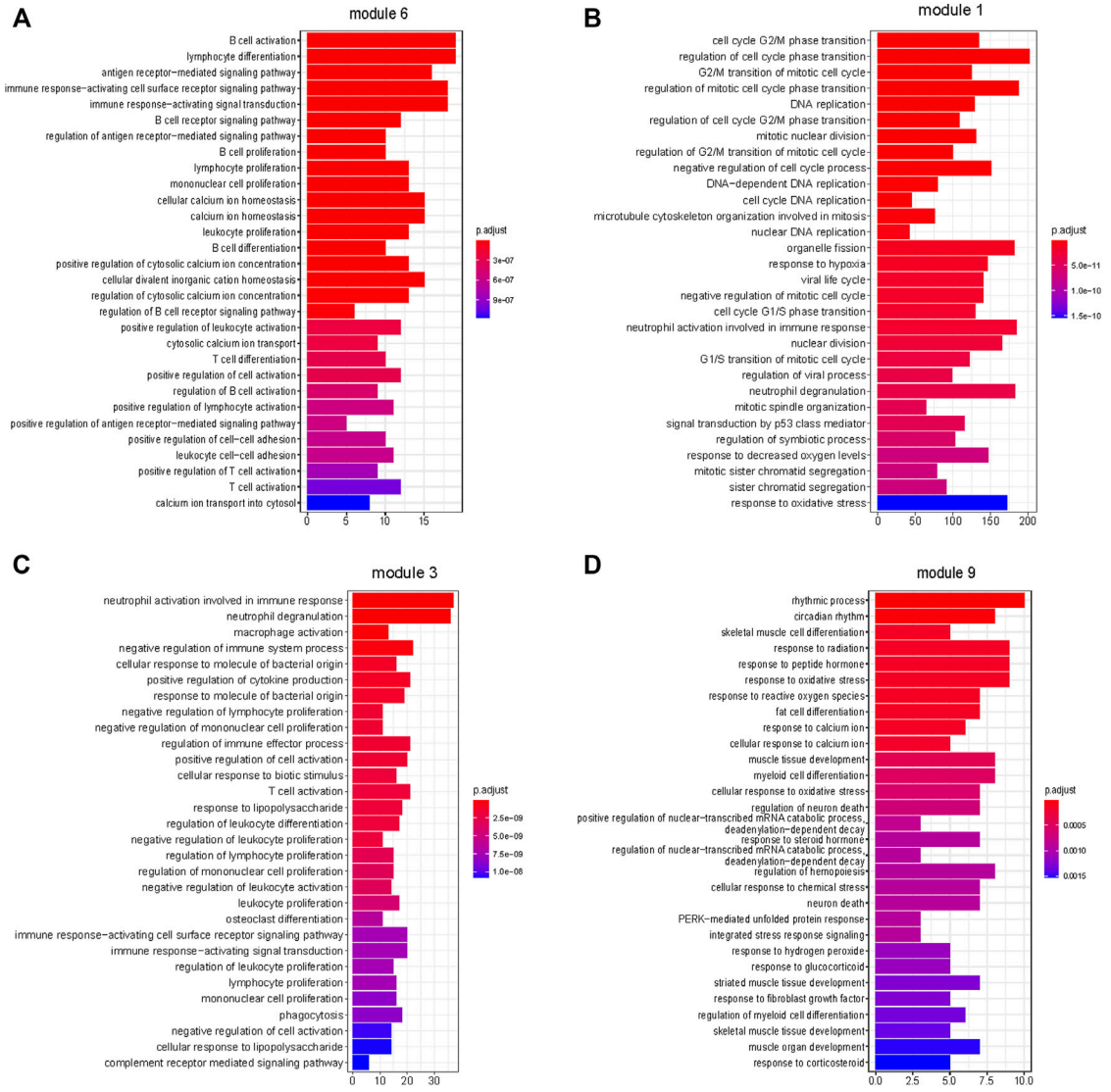


FIGURE 5
Enrichment of biological pathways in modules. **(A–D)** The top 30 pathways that enriched in module 6, module 1, module 3, and module 9, respectively. **(A,C)** Immune-related processes were enriched in modules 6 and 3. **(B)** Cell proliferation-related processes were enriched in module 1. **(D)** Cellular responses to external stimuli were enriched in module 9. (FDR <0.05).

(Figures 7H, I). Therefore, the biological pathways that respond to various cellular stress conditions and the DNA replication pathways play important roles in Subtype 1.

3.8 Validation of the generality of esophageal squamous cell carcinoma subtypes

To explore the generality of our classification, we clustered 80 ESCC tumor samples in the TCGA database and compared the obtained three subtypes with the three subtypes obtained by the 141 ESCC samples above. The 80 ESCC samples were clustered into three subtypes, named as TCGA_Subtype1 (24 samples), TCGA_Subtype2 (3 samples), TCGA_Subtype3 (53 samples) (Figure 8A). These three TCGA_Subtypes correspond one-to-one with the positions of the three Subtypes obtained by 141 ESCC samples on the clustering tree (Figure 1A, Figure 8A).

Through the correlation analysis between 80 TCGA ESCC samples and 141 GEO ESCC samples, it was found that the average correlation between subtypes located at the same position on the clustering tree is significantly higher than the average correlation between subtypes in different positions (Figures 8B–D). For example, the average correlation value

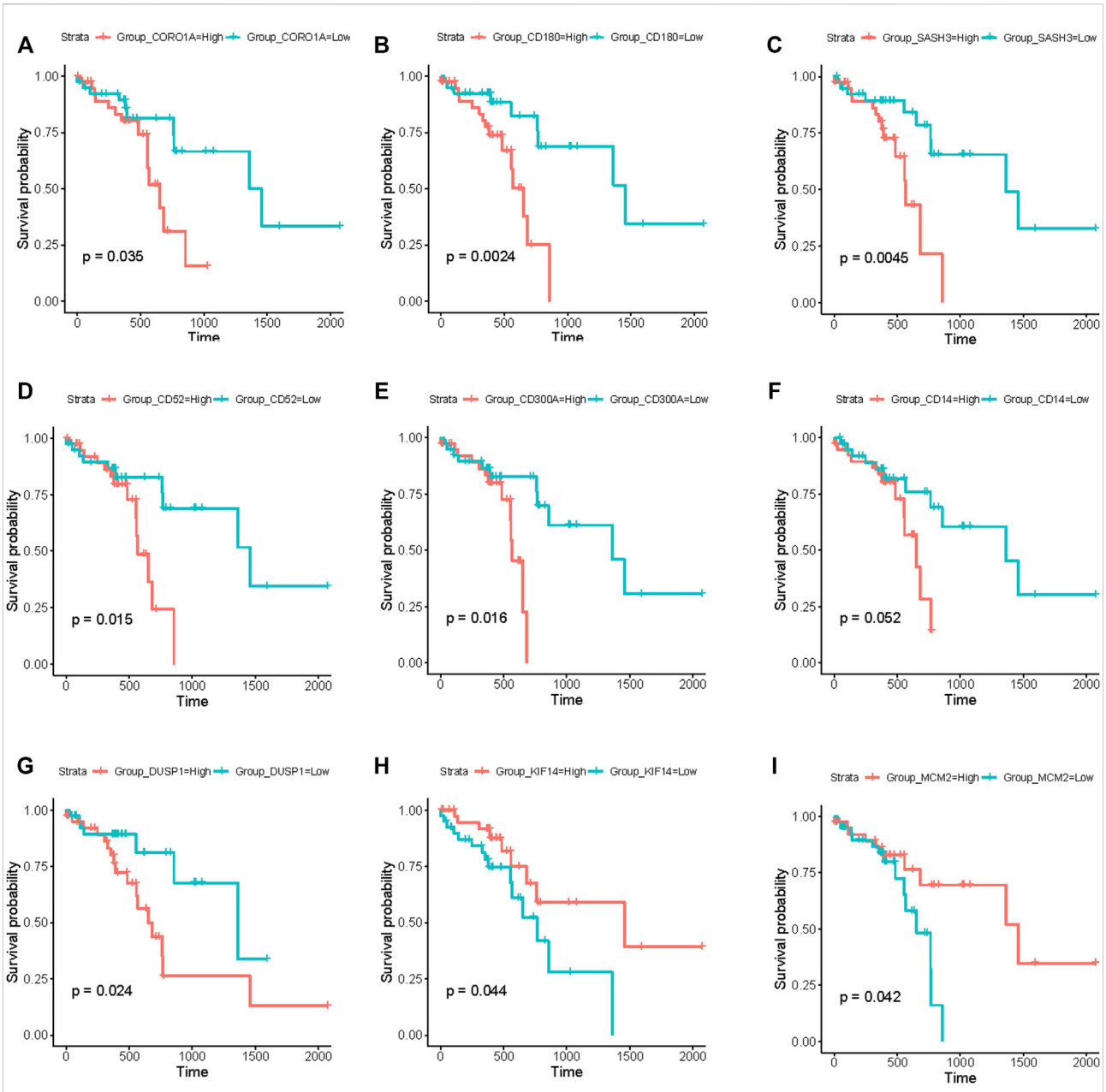


FIGURE 6
Survival analysis using the RNA-seq data from TCGA database. **(A–I)** Kaplan–Meier curves showing that the expression levels of these nine genes (CORO1A, CD180, SASH3, CD52, CD300A, CD14, DUSP1, KIF14, and MCM2) were significantly associated with survival of patients with ESCC. **(A–G)** Seven genes (CORO1A, CD180, SASH3, CD52, CD300A, CD14, and DUSP1) were negatively correlated with survival. **(H,I)** Two genes (KIF14 and MCM2) were positively correlated with survival. ($p < 0.05$).

between TCGA_Subtype3 and Subtype 3 is significantly higher than that TCGA_Subtype3 and Subtype 1 or Subtype 2 (Figure 8D). Therefore, our classification of ESCC into three subtypes by gene expression profiling is of general significance.

The survival analysis revealed no significant differences in survival between TCGA_Subtypes (Figure 8E). This means that the three ESCC subtypes we found may not have differences in survival.

4 Discussion

In this study, we integrated the gene expression profiles from different studies to a comparable level by correcting batch effects. In total, the gene expression profiles of paired tumor and normal samples from 141 patients with ESCC were included in this study. The 141 patients were divided into three subtypes based on

TABLE 2 The profile of the genes that correlated with survival of ESCC patients.

Gene	Module	Correlation with survival	Gene ontology enrichment
CORO1A	Module6	negative	lymphocyte proliferation
			cellular calcium ion homeostasis
			leukocyte proliferation
CD180	Module6	negative	B cell activation
			B cell proliferation
SASH3	Module6	negative	B cell activation; lymphocyte differentiation
CD52	Module6	negative	cellular calcium ion homeostasis
			calcium ion homeostasis
			positive regulation of cytosolic calcium ion concentration
CD300A	Module3	negative	neutrophil activation involved in immune response
			neutrophil degranulation
CD14	Module3	negative	neutrophil activation involved in immune response
			neutrophil degranulation
			cellular response to molecule of bacterial origin
DUSP1	Module9	negative	response to radiation
			response to oxidative stress
KIF14	Module1	positive	cell cycle G2/M phase transition
			regulation of cell cycle phase transition
MCM2	Module1	positive	DNA replication
			DNA-dependent DNA replication

their tumor sample gene expression profiles. Then, detailed characteristics of the three subtypes were described. The results of DGE analysis revealed that subtypes are different to each other. Next, we identified typical phenotype in each subtype; Wnt signaling pathway activation in Subtype 1, downregulation of glycogen metabolism in Subtype 2 and immunosuppression in Subtype 3. WGCNA revealed finer regulation of biological pathways in the three subtypes and revealed hub genes with important regulatory status. Moreover, we validated several hub genes were survival-related based on RNA-seq data from TCGA database. By comparing the survival-related-gene expression level in three subtypes, we suggested that the genes involved in immune-related biological processes or cell proliferation-related processes were responsible for the survival of these subtypes. Finally, in the RNA-seq dataset from TCGA database we verified that the three ESCC subtypes we found were of general significance.

The striking differences among the three ESCC subtypes were manifested in several aspects. For example, the degree of change in the expression levels of most genes between tumor and match normal tissue gradually increased from Subtype 1 to Subtype 3, especially genes with copy number variation (CNV) in ESCC genome. One potential explanation for this phenomenon is the increased overall tumor mutational burden (TMB, the total number of mutations in cancer cell DNA). Several studies have reported that copy number variations on ESCC genome were consistent with changes in gene expression levels (Hu et al., 2009; Shi et al., 2013). Liu et al. (2016) have found that subgroups of ESCC have significantly different somatic mutational burdens, such as subgroup1 (0.75 mutations per Mb) in their study, which showed fewest somatic mutational burden compared with subgroups1a (11.85 mutations per Mb) and subgroup2 (3.71 mutations per Mb). This is consistent with our conjecture that the three subtypes we identified appear to have differences in TMB. A high TMB is associated with poor prognosis (Owada-Ozaki et al., 2018; Hwang et al., 2019; Cui et al., 2020), and TMB has been demonstrated as a selection biomarker of immune checkpoint blockade (ICB) cancer therapy (Chan et al., 2019). In addition, some biological processes also exhibit progressively stronger association from Subtype 1 to Subtype 3, such as upregulation of the cell proliferation and dysfunction of endothelial cells in the results of co-expression modules. The progressively severe endothelial cell dysfunction from Subtype 1 to Subtype 3 means aggravated hypoxic environment and accelerated angiogenesis. Endothelial cells line the vascular systems and play important roles in tumorigenesis. The tumor microenvironment suffers from hypoxia, which will continuously stimulate blood vessel formation (Potente et al., 2011; Jing et al., 2019). These rapidly growing blood vessels are naturally differentiated from normal blood vessels, and tumor endothelial cells (TECs) exhibit different cell proliferation and migration ability compared with normal endothelial cells (NECs) (Hida et al., 2016). Accordingly, it is effective to subtyping ESCC

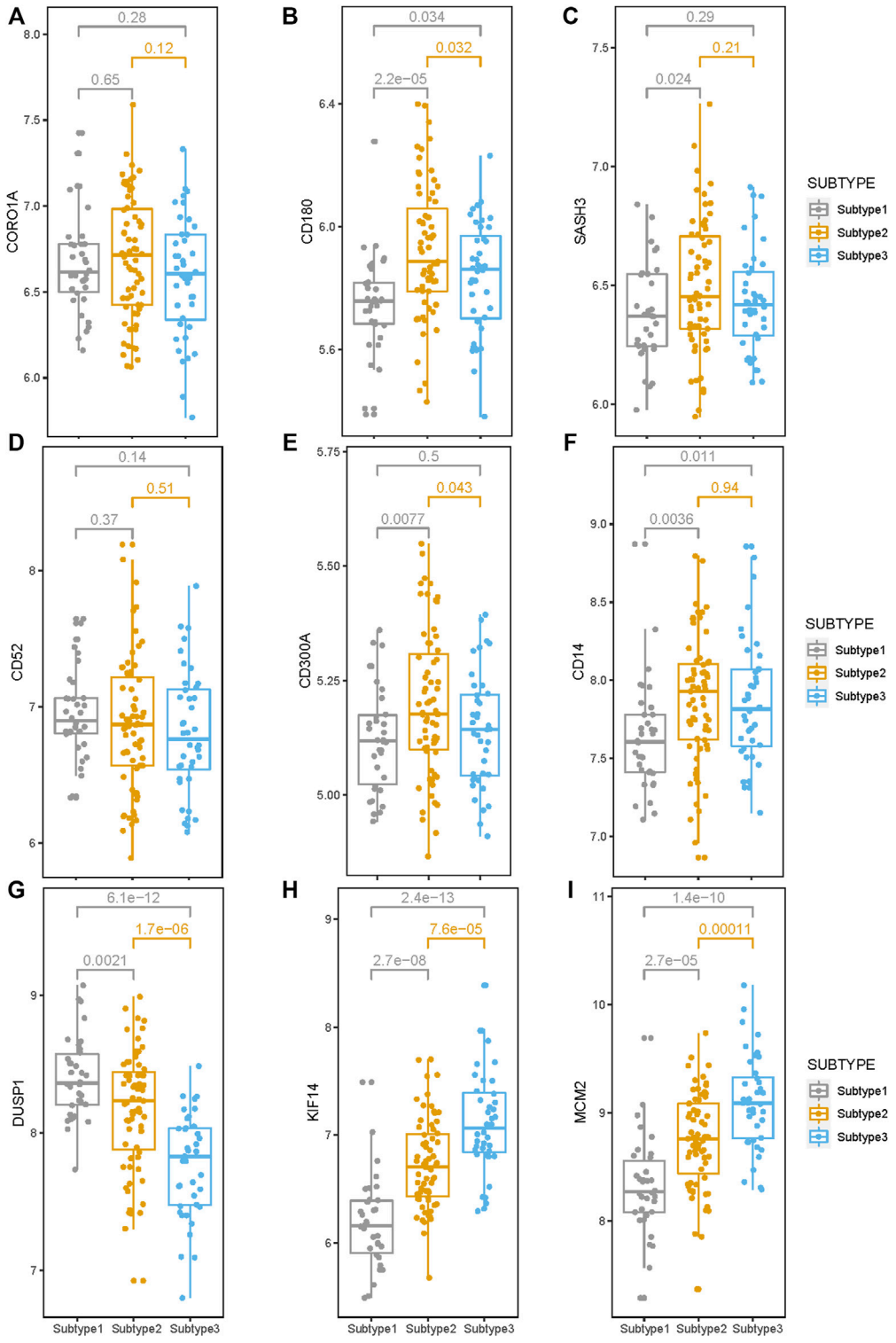


FIGURE 7
Expression levels of the nine survival-related genes in the three subtypes. **(A–I)** The expression level of the nine survival-related genes in the three subtypes. Expression levels of these seven (CD180, SASH3, CD300A, CD14, DUSP1, KIF14, and MCM2) survival-related genes are significantly different among these three subtypes. ($p < 0.05$).

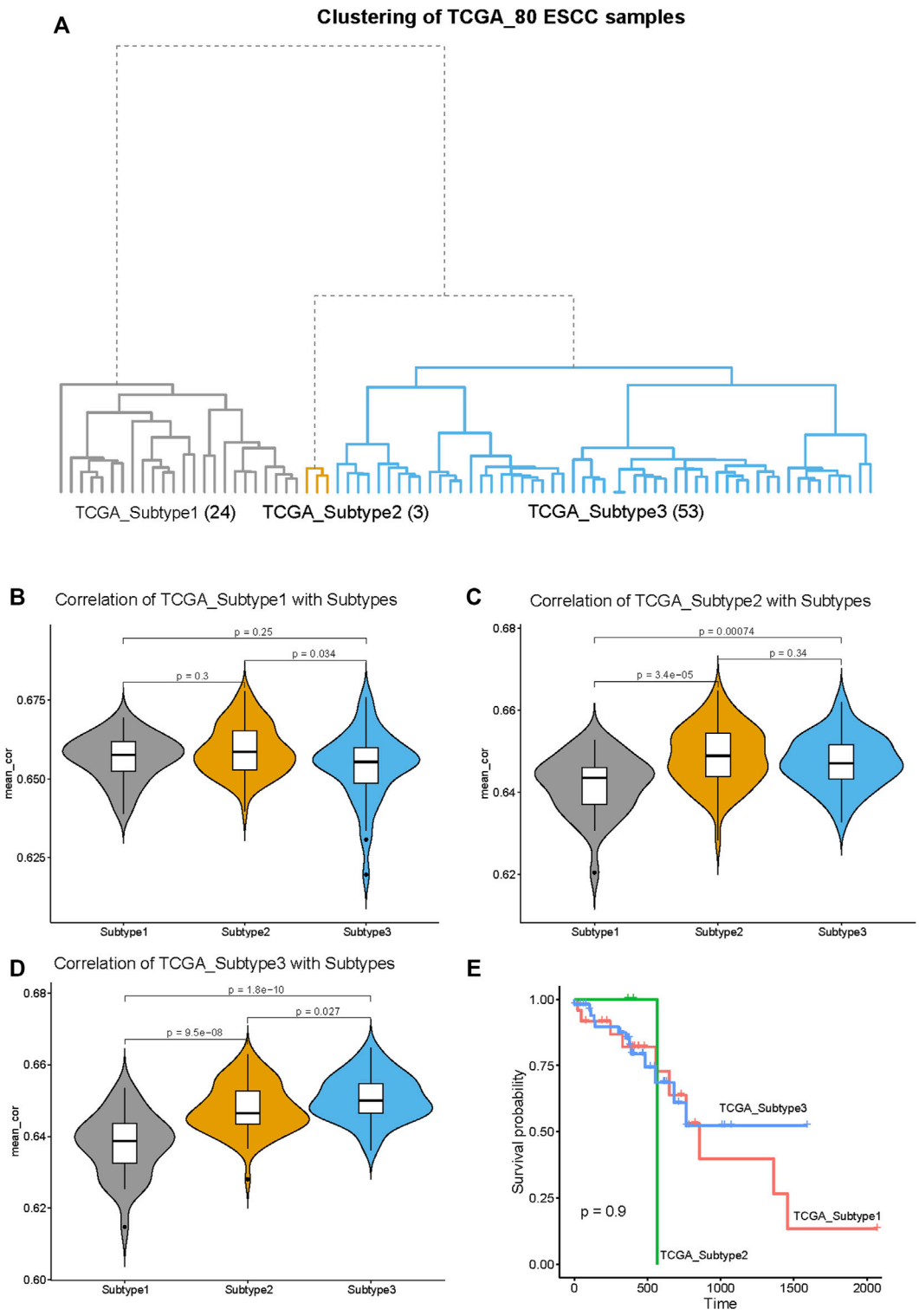


FIGURE 8
Validation of the generality of our ESCC Subtypes based on the RNA-seq data from TCGA database. **(A)** The hierarchical clustering tree of the 80 ESCC samples based on the RNA-seq data from TCGA database. The three subtypes are designated as: TCGA_Subtype1 (24 samples), TCGA_Subtype2 (3 samples), and TCGA_Subtype3 (53 samples). **(B)** The violin plot of the average correlation values between TCGA_Subtype1 and Subtypes (Subtype 1, Subtype 2 and Subtype 3). The average correlation values of TCGA_Subtype1 and Subtype 1 are significantly higher than TCGA_Subtype1 with Subtype 2/Subtype 3. **(C)** The violin plot of the average correlation values between TCGA_Subtype2 and Subtypes. The average
(*Continued*)

FIGURE 8 (Continued)
correlation values of TCGA_Subtype2 and Subtype 2 are significantly higher than TCGA_Subtype2 with Subtype 1/Subtype 3. **(D)** The violin plot of the average correlation values between TCGA_Subtype3 and Subtypes. The average correlation values of TCGA_Subtype3 and Subtype 3 are significantly higher than TCGA_Subtype3 Subtype 1/Subtype 2. **(E)** Survival analysis between TCGA_Subtypes. Kaplan–Meier curves showing there is no significant difference between TCGA_Subtypes in survival.

from the perspective of transcriptome, which helps us to understand the molecular characteristics of ESCC more deeply and provide reference for precise treatment.

Subtype 1 was associated with the activation of Wnt signaling pathway, which will be a promising treatment target in ESCC. As an important pathway, Wnt signaling pathway activation has repeatedly been demonstrated in ESCC (Song et al., 2014; Sawada et al., 2016). Besides, Wnt signaling was reported to be inversely correlated with T cell infiltration in colorectal cancer, (Grasso et al., 2018), as we also found a lower T cell infiltration in subtype 1 than in other subtypes. The Wnt signaling pathway activation is associated with tumorigenesis and progression (Nusse and Varmus, 1992; Zhan et al., 2017; You et al., 2019); tumor proliferation and progression were inhibited by suppressing this pathway (Lee et al., 2012; You et al., 2019; Yu et al., 2021).

Significantly reduced glycogen metabolism in Subtype 2 may lead to glycogen accumulation in tumor tissue. Glycogen accumulation in tumor tissue is a distinguishing feature in various cancers and which promotes tumor development and maintenance (Iida et al., 2012; Maruggi et al., 2019; Xie et al., 2021). Accelerating glycogen metabolism can play a role in suppressing tumors, several enzymes involved in glycogen metabolism exert tumor-suppressive effects, including the glycogen debranching enzyme AGL and the kinase PhK β-subunit (PHKB) (Guin et al., 2016; Richmond et al., 2018; Yang et al., 2019; Liu et al., 2021). So, promoting glycogen metabolism may be a way to inhibit Subtype 2 ESCC.

Subtype 3 is characterized by immunosuppression, including downregulation of neutrophil degranulation and B/T cell related immune pathway. Assuming the downregulation of neutrophil degranulation was caused by the reduced total number of tumor-infiltrating neutrophils, we performed the immune cell infiltration analysis. The result confirmed decreased of neutrophils infiltration in subtype 3 compared to subtype 1 or 2, which possibly lead to downregulation of neutrophil degranulation. Similar down-stream pathways were identified in gene co-expression network. We found the downregulation of neutrophil chemotaxis in Subtype 3 based on the enrichment result of module 0. In view of the crucial role of neutrophils in the pathogenesis of cancer (Mantovani et al., 2011) and its position in the regulation of innate and adaptive immunity (Scapini et al., 2000; Tecchio et al., 2013; Uribe-Querol and Rosales, 2015; Jaillon et al., 2020), the reduction of infiltrating neutrophils may also be part of the reason why B- and T-cell immune-related processes are affected in subtype 3. Actually, several single-cell studies have demonstrated an immunosuppressive microenvironment in ESCC (Zheng et al., 2020; Dinh et al., 2021). Therefore, Subtype 3 is an immunosuppressive ESCC subtype, which also makes it most likely to benefit from immunotherapy. Wang et al. (2019) have found one immune-activate ESCC subtype by comparing two ESCC subtypes they have identified, this is significantly different from our Subtype 3. The difference may arise due to differences in analytical methods. They characterize the subtypes by making comparisons between the subtypes and we characterize the subtypes by comparing the tumor tissue to the normal tissue.

These nine survival-related genes are in key regulatory positions in the gene expression of the three subtypes; they are of great significance for the tumorigenesis and progression. Six of the nine survival-related genes (*CORO1A, CD180, SASH3, CD52, CD300A, and CD14*) were involved in the immune pathways, and were all were negatively correlated with survival, which is consistent with our understanding of the duality of immunity in tumors: immune has dual roles of suppressing and promoting cancer (Schreiber et al., 2011). *CD180* and *SASH3* are involved in the B cell-related immune pathway. The *SASH3*-encoded protein acts as a signaling protein in lymphocytes, and high SASH3 expression was associated with poor survival in cellular renal cell carcinoma (Yin et al., 2020). *CD52* and *CORO1A* are involved in regulating calcium homeostasis in immune cells. Calcium homeostasis is important for regulating the activation and function of macrophages, dendritic cells, and T cells (Zophel et al., 2020). *CD300A* and *CD14* are involved in neutrophil degranulation. CD14 is a key molecule in innate immunity activation, patients with bladder cancer with high CD14 levels may develop a proliferative tumor microenvironment (Cheah et al., 2015). The inhibitory receptor protein CD300A is found on leukocytes and is involved in the immune response signaling pathways; the interaction between CD300A and phosphatidylserine can inhibit the killing effect of natural killer (NK) cells on tumor cells (Lankry et al., 2013). In addition to these six survival-related genes involved in immune-related processes, the remaining three survival-related genes are involved in other biological processes. The mechanism of DUSP1 in organisms is highly complex: it is a transcriptional target of tumor suppressor p53 (Li et al., 2003) and also responds to various cellular stress conditions (Keyse and Emslie, 1992). In invasive ovarian cancer, *DUSP1* expression was significantly associated with shorter progression-free survival (p = 0.019)

(Denkert et al., 2002). *KIF14* and *MCM2* are involved in numerous biological pathways, including G2–M transition, mitotic nuclear division, and DNA replication. *MCM2* has a coregulatory role in ESCC progression and may have core roles during the pathogenesis of ESCC (Wang et al., 2018). MCMs increase genome duplication robustness by restraining the speed at which eukaryotic cells replicate their DNA, where a minor reduction in MCM levels destabilizes the genome and predisposes to increased incidence of tumor formation (Sedlackova et al., 2020). Genes can influence the overall performance of transcriptional profile through different regulatory strategies (Wang et al., 2021). The roles of these nine survival-related genes in ESCC need further explorations and may become potential targets for ESCC therapy in the future.

There are some limitations of this study. The clinical information of these 141 ESCC samples was not comprehensive enough to interpret existing findings in combination with more clinical information. In addition, the genes with prognostic value have not been validated in other independent cohorts. Finally, due to the limited laboratory conditions, some of our results have not been verified in cell experiments.

In conclusion, we have identified three subtypes of ESCC by large-scale gene expression profiling of tumor tissues. Through in-depth exploration of these three subtypes, we have characterized the three subtypes from multiple perspectives and discovered some new potential targets that may be effective in the treatment of ESCC. Taken together, our results suggest that distinct ESCC subtypes defined using transcriptomes may exhibit better responses to specific targeted therapies. Actually, the effectiveness of these targets needs further exploration and verification. Our findings have deepened our understanding of the molecular characteristics of ESCC and provided some references for future clinical treatment research.

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

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

DW collected the data, determined the analysis process, visualized the results, and wrote the draft. DW and JD had the original idea for the study and reviewed the manuscript. CS and XC reviewed the manuscript and got the funding acquisition. All authors have read and approved the final manuscript.

Funding

This study was funded by the National Key Research and Development Program of China (2019YFC1315804 to CS), the National Natural Science Foundation of China (82073637, 82122060, and 91846302 to XC), and Shanghai Municipal Science and Technology Major Project (ZD2021CY001 to CS).

Acknowledgments

Thanks to the efforts of all people involved in this work. Thanks to the GEO and TCGA databases for providing valuable datasets. Thanks for the financial support that from the National Natural Science Foundation of China, the National Key Research and Development Program of China and Shanghai Municipal Science and Technology Major Project.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1033214/full#supplementary-material.

References

An, Y., Wang, S., Li, S., Zhang, L., Wang, D., Wang, H., et al. (2017). Distinct molecular subtypes of uterine leiomyosarcoma respond differently to chemotherapy treatment. *BMC Cancer* 17 (1), 639. doi:10.1186/s12885-017-3568-y

Cancer Genome Atlas Research Network, Analysis Working Group: Asan, U., Agency, B. C. C., BrighamWomen's, H., and Broad, I. (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature* 541 (7636), 169–175. doi:10.1038/nature20805

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

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

Cheah, M. T., Chen, J. Y., Sahoo, D., Contreras-Trujillo, H., Volkmer, A. K., Scheeren, F. A., et al. (2015). CD14-expressing cancer cells establish the inflammatory and proliferative tumor microenvironment in bladder cancer. *Proc. Natl. Acad. Sci. U. S. A.* 112 (15), 4725–4730. doi:10.1073/pnas.1424795112

Chen, F., Zhang, Y., Parra, E., Rodriguez, J., Behrens, C., Akbani, R., et al. (2017). Multiplatform-based molecular subtypes of non-small-cell lung cancer. *Oncogene* 36 (10), 1384–1393. doi:10.1038/onc.2016.303

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

Cui, Y., Chen, H., Xi, R., Cui, H., Zhao, Y., Xu, E., et al. (2020). Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Res.* 30 (10), 902–913. doi:10.1038/s41422-020-0333-6

Della Guardia, A., Lake, M., and Schnitzer, P. (2022). Selective inclusion in cash transfer programs: Unintended consequences for social cohesion. *World Dev.* 157, 105922. doi:10.1016/j.worlddev.2022.105922

Denkert, C., Schmitt, W. D., Berger, S., Reles, A., Pest, S., Siegert, A., et al. (2002). Expression of mitogen-activated protein kinase phosphatase-1 (MKP-1) in primary human ovarian carcinoma. *Int. J. Cancer* 102 (5), 507–513. doi:10.1002/ijc.10746

Dinh, H. Q., Pan, F., Wang, G., Huang, Q. F., Olingy, C. E., Wu, Z. Y., et al. (2021). Integrated single-cell transcriptome analysis reveals heterogeneity of esophageal squamous cell carcinoma microenvironment. *Nat. Commun.* 12 (1), 7335. doi:10.1038/s41467-021-27599-5

Erkizan, H. V., Johnson, K., Ghimbovschi, S., Karkera, D., Trachiotis, G., Adib, H., et al. (2017). African-American esophageal squamous cell carcinoma expression profile reveals dysregulation of stress response and detox networks. *BMC Cancer* 17 (1), 426. doi:10.1186/s12885-017-3423-1

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20 (3), 307–315. doi:10.1093/bioinformatics/btg405

Ge, X., Chen, Y. E., Song, D., McDermott, M., Woyshner, K., Manousopoulou, A., et al. (2021). Clipper: p-value-free FDR control on high-throughput data from two conditions. *Genome Biol.* 22 (1), 288. doi:10.1186/s13059-021-02506-9

Grasso, C. S., Giannakis, M., Wells, D. K., Hamada, T., Mu, X. J., Quist, M., et al. (2018). Genetic mechanisms of immune evasion in colorectal cancer. *Cancer Discov.* 8 (6), 730–749. doi:10.1158/2159-8290.CD-17-1327

Guin, S., Ru, Y., Agarwal, N., Lew, C. R., Owens, C., Comi, G. P., et al. (2016). Loss of glycogen debranching enzyme AGL drives bladder tumor growth via induction of hyaluronic acid synthesis. *Clin. Cancer Res.* 22 (5), 1274–1283. doi:10.1158/1078-0432.CCR-15-1706

Guinney, J., Dienstmann, R., Wang, X., de Reynies, A., Schlicker, A., Soneson, C., et al. (2015). The consensus molecular subtypes of colorectal cancer. *Nat. Med.* 21 (11), 1350–1356. doi:10.1038/nm.3967

Guo, W., Li, L., He, J., Liu, Z., Han, M., Li, F., et al. (2020). Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nat. Genet.* 52 (9), 908–918. doi:10.1038/s41588-020-0642-1

Halbritter, F., Farlik, M., Schwentner, R., Jug, G., Fortelny, N., Schnoller, T., et al. (2019). Epigenomics and single-cell sequencing define a developmental hierarchy in langerhans cell histiocytosis. *Cancer Discov.* 9 (10), 1406–1421. doi:10.1158/2159-8290.CD-19-0138

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

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

Hida, K., Maishi, N., Sakurai, Y., Hida, Y., and Harashima, H. (2016). Heterogeneity of tumor endothelial cells and drug delivery. *Adv. Drug Deliv. Rev.* 99, 140–147. doi:10.1016/j.addr.2015.11.008

Hu, N., Clifford, R. J., Yang, H. H., Wang, C., Goldstein, A. M., Ding, T., et al. (2010). Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. *BMC Genomics* 11, 576. doi:10.1186/1471-2164-11-576

Hu, N., Wang, C., Clifford, R. J., Yang, H. H., Su, H., Wang, L., et al. (2015). Integrative genomics analysis of genes with biallelic loss and its relation to the expression of mRNA and micro-RNA in esophageal squamous cell carcinoma. *BMC Genomics* 16 (1), 732. doi:10.1186/s12864-015-1919-0

Hu, N., Wang, C., Ng, D., Clifford, R., Yang, H. H., Tang, Z. Z., et al. (2009). Genomic characterization of esophageal squamous cell carcinoma from a high-risk population in China. *Cancer Res.* 69 (14), 5908–5917. doi:10.1158/0008-5472.CAN-08-4622

Hwang, W. L., Wolfson, R. L., Niemierko, A., Marcus, K. J., DuBois, S. G., and Haas-Kogan, D. (2019). Clinical impact of tumor mutational burden in neuroblastoma. *J. Natl. Cancer Inst.* 111 (7), 695–699. doi:10.1093/jnci/djy157

Iida, Y., Aoki, K., Asakura, T., Ueda, K., Yanaihara, N., Takakura, S., et al. (2012). Hypoxia promotes glycogen synthesis and accumulation in human ovarian clear cell carcinoma. *Int. J. Oncol.* 40 (6), 2122–2130. doi:10.3892/ijo.2012.1406

Jaillon, S., Ponzetta, A., Di Mitri, D., Santoni, A., Bonecchi, R., and Mantovani, A. (2020). Neutrophil diversity and plasticity in tumour progression and therapy. *Nat. Rev. Cancer* 20 (9), 485–503. doi:10.1038/s41568-020-0281-y

Jing, X., Yang, F., Shao, C., Wei, K., Xie, M., Shen, H., et al. (2019). Role of hypoxia in cancer therapy by regulating the tumor microenvironment. *Mol. Cancer* 18 (1), 157. doi:10.1186/s12943-019-1089-9

Keyse, S. M., and Emslie, E. A. (1992). Oxidative stress and heat shock induce a human gene encoding a protein-tyrosine phosphatase. *Nature* 359 (6396), 644–647. doi:10.1038/359644a0

LaFave, L. M., Kartha, V. K., Ma, S., Meli, K., Del Priore, I., Lareau, C., et al. (2020). Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* 38 (2), 212–228. e213. doi:10.1016/j.ccell.2020.06.006

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

Lankry, D., Rovis, T. L., Jonjic, S., and Mandelboim, O. (2013). The interaction between CD300a and phosphatidylserine inhibits tumor cell killing by NK cells. *Eur. J. Immunol.* 43 (8), 2151–2161. doi:10.1002/eji.201343433

Lee, J. J., Natsuizaka, M., Ohashi, S., Wong, G. S., Takaoka, M., Michaylira, C. Z., et al. (2010). Hypoxia activates the cyclooxygenase-2-prostaglandin E synthase axis. *Carcinogenesis* 31 (3), 427–434. doi:10.1093/carcin/bgp326

Lee, S. Y., Jeon, H. M., Ju, M. K., Kim, C. H., Yoon, G., Han, S. I., et al. (2012). Wnt/Snail signaling regulates cytochrome C oxidase and glucose metabolism. *Cancer Res.* 72 (14), 3607–3617. doi:10.1158/0008-5472.CAN-12-0006

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034

Li, M., Zhou, J. Y., Ge, Y., Matherly, L. H., and Wu, G. S. (2003). The phosphatase MKP1 is a transcriptional target of p53 involved in cell cycle regulation. *J. Biol. Chem.* 278 (42), 41059–41068. doi:10.1074/jbc.M307149200

Lin, D. C., Wang, M. R., and Koeffler, H. P. (2018). Genomic and epigenomic aberrations in esophageal squamous cell carcinoma and implications for patients. *Gastroenterology* 154 (2), 374–389. doi:10.1053/j.gastro.2017.06.066

Liu, Q., Li, J., Zhang, W., Xiao, C., Zhang, S., Nian, C., et al. (2021). Glycogen accumulation and phase separation drives liver tumor initiation. *Cell* 184 (22), 5559–5576.e19. doi:10.1016/j.cell.2021.10.001

Liu, W., Snell, J. M., Jeck, W. R., Hoadley, K. A., Wilkerson, M. D., Parker, J. S., et al. (2016). Subtyping sub-Saharan esophageal squamous cell carcinoma by comprehensive molecular analysis. *JCI Insight* 1 (16), e88755. doi:10.1172/jci.insight.88755

Mantovani, A., Cassatella, M. A., Costantini, C., and Jaillon, S. (2011). Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat. Rev. Immunol.* 11 (8), 519–531. doi:10.1038/nri3024

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

Marjanovic, N. D., Hofree, M., Chan, J. E., Canner, D., Wu, K., Trakala, M., et al. (2020). Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell* 38 (2), 229–246. e213. doi:10.1016/j.ccell.2020.06.012

Maruggi, M., Layng, F. I., Lemos, R., Jr., Garcia, G., James, B. P., Sevilla, M., et al. (2019). Absence of HIF1A leads to glycogen accumulation and an inflammatory response that enables pancreatic tumor growth. *Cancer Res.* 79 (22), 5839–5848. doi:10.1158/0008-5472.CAN-18-2994

Ming, X. Y., Zhang, X., Cao, T. T., Zhang, L. Y., Qi, J. L., Kam, N. W., et al. (2018). RHCG suppresses tumorigenicity and metastasis in esophageal squamous cell carcinoma via inhibiting NF-κB signaling and MMP1 expression. *Theranostics* 8 (1), 185–198. doi:10.7150/thno.21383

Nusse, R., and Varmus, H. E. (1992). Wnt genes. *Cell* 69 (7), 1073–1087. doi:10.1016/0092-8674(92)90630-u

Owada-Ozaki, Y., Muto, S., Takagi, H., Inoue, T., Watanabe, Y., Fukuhara, M., et al. (2018). Prognostic impact of tumor mutation burden in patients with completely resected non-small cell lung cancer: Brief report. *J. Thorac. Oncol.* 13 (8), 1217–1221. doi:10.1016/j.jtho.2018.04.003

Potente, M., Gerhardt, H., and Carmeliet, P. (2011). Basic and therapeutic aspects of angiogenesis. *Cell* 146 (6), 873–887. doi:10.1016/j.cell.2011.08.039

Quintanal-Villalonga, A., Chan, J. M., Yu, H. A., Pe'er, D., Sawyers, C. L., Sen, T., et al. (2020). Lineage plasticity in cancer: A shared pathway of therapeutic resistance. *Nat. Rev. Clin. Oncol.* 17 (6), 360–371. doi:10.1038/s41571-020-0340-z

Richmond, C. S., Oldenburg, D., Dancik, G., Meier, D. R., Weinhaus, B., Theodorescu, D., et al. (2018). Glycogen debranching enzyme (AGL) is a novel regulator of non-small cell lung cancer growth. *Oncotarget* 9 (24), 16718–16730. doi:10.18632/oncotarget.24676

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

Sawada, G., Niida, A., Uchi, R., Hirata, H., Shimamura, T., Suzuki, Y., et al. (2016). Genomic landscape of esophageal squamous cell carcinoma in a Japanese population. *Gastroenterology* 150 (5), 1171–1182. doi:10.1053/j.gastro.2016.01.035

Scapini, P., Lapinet-Vera, J. A., Gasperini, S., Calzetti, F., Bazzoni, F., and Cassatella, M. A. (2000). The neutrophil as a cellular source of chemokines. *Immunol. Rev.* 177, 195–203. doi:10.1034/j.1600-065x.2000.17706.x

Schreiber, R. D., Old, L. J., and Smyth, M. J. (2011). Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. *Science* 331 (6024), 1565–1570. doi:10.1126/science.1203486

Sedlackova, H., Rask, M. B., Gupta, R., Choudhary, C., Somyajit, K., and Lukas, J. (2020). Equilibrium between nascent and parental MCM proteins protects replicating genomes. *Nature* 587 (7833), 297–302. doi:10.1038/s41586-020-2842-3

Shi, Z. Z., Shang, L., Jiang, Y. Y., Hao, J. J., Zhang, Y., Zhang, T. T., et al. (2013). Consistent and differential genetic aberrations between esophageal dysplasia and squamous cell carcinoma detected by array comparative genomic hybridization. *Clin. Cancer Res.* 19 (21), 5867–5878. doi:10.1158/1078-0432.CCR-12-3753

Siewert, J. R., and Ott, K. (2007). Are squamous and adenocarcinomas of the esophagus the same disease? *Semin. Radiat. Oncol.* 17 (1), 38–44. doi:10.1016/j.semradonc.2006.09.007

Smyth, E. C., Lagergren, J., Fitzgerald, R. C., Lordick, F., Shah, M. A., Lagergren, P., et al. (2017). Oesophageal cancer. *Nat. Rev. Dis. Prim.* 3, 17048. doi:10.1038/nrdp.2017.48

Song, Y., Li, L., Ou, Y., Gao, Z., Li, E., Li, X., et al. (2014). Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* 509 (7498), 91–95. doi:10.1038/nature13176

Su, H., Hu, N., Yang, H. H., Wang, C., Takikita, M., Wang, Q. H., et al. (2011). Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin. Cancer Res.* 17 (9), 2955–2966. doi:10.1158/1078-0432.CCR-10-2724

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

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

Tecchio, C., Scapini, P., Pizzolo, G., and Cassatella, M. A. (2013). On the cytokines produced by human neutrophils in tumors. *Semin. Cancer Biol.* 23 (3), 159–170. doi:10.1016/j.semcancer.2013.02.004

Uribe-Querol, E., and Rosales, C. (2015). Neutrophils in cancer: Two sides of the same coin. *J. Immunol. Res.* 2015, 983698. doi:10.1155/2015/983698

Wang, F., Yan, Z., Lv, J., Xin, J., Dang, Y., Sun, X., et al. (2019). Gene expression profiling reveals distinct molecular subtypes of esophageal squamous cell carcinoma in asian populations. *Neoplasia* 21 (6), 571–581. doi:10.1016/j.neo.2019.03.013

Wang, M., Yang, Y. O., Jin, Q., Shang, L., and Zhang, J. (2018). Function of miR-25 in the invasion and metastasis of esophageal squamous carcinoma cells and bioinformatical analysis of the miR-106b-25 cluster. *Exp. Ther. Med.* 15 (1), 440–446. doi:10.3892/etm.2017.5358

Wang, N., Lefaudeux, D., Mazumder, A., Li, J. J., and Hoffmann, A. (2021). Identifying the combinatorial control of signal-dependent transcription factors. *PLoS Comput. Biol.* 17 (6), e1009095. doi:10.1371/journal.pcbi.1009095

Xie, H., Song, J., Godfrey, J., Riscal, R., Skuli, N., Nissim, I., et al. (2021). Glycogen metabolism is dispensable for tumour progression in clear cell renal cell carcinoma. *Nat. Metab.* 3 (3), 327–336. doi:10.1038/s42255-021-00367-x

Xu, Y., Baumgart, S. J., Stegmann, C. M., and Hayat, S. (2021). Maca: Marker-based automatic cell-type annotation for single cell expression data. *Bioinformatics* 38, 1756–1760. doi:10.1093/bioinformatics/btab840

Yang, W., Zhang, C., Li, Y., Jin, A., Sun, Y., Yang, X., et al. (2019). Phosphorylase kinase beta represents a novel prognostic biomarker and inhibits malignant phenotypes of liver cancer cell. *Int. J. Biol. Sci.* 15 (12), 2596–2606. doi:10.7150/ijbs.33278

Yin, X., Zhang, X., Liu, Z., Sun, G., Zhu, X., Zhang, H., et al. (2020). Assessment for prognostic value of differentially expressed genes in immune microenvironment of clear cell renal cell carcinoma. *Am. J. Transl. Res.* 12 (9), 5416–5432.

You, B. H., Yoon, J. H., Kang, H., Lee, E. K., Lee, S. K., and Nam, J. W. (2019). HERES, a lncRNA that regulates canonical and noncanonical Wnt signaling pathways via interaction with EZH2. *Proc. Natl. Acad. Sci. U. S. A.* 116 (49), 24620–24629. doi:10.1073/pnas.1912126116

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

Yu, J., Wang, W., Liu, B., Gu, J., Chen, S., Cui, Y., et al. (2021). Demethylzelasteral inhibits proliferation and EMT via repressing Wnt/β-catenin signaling in esophageal squamous cell carcinoma. *J. Cancer* 12 (13), 3967–3975. doi:10.7150/jca.45493

Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt signaling in cancer. *Oncogene* 36 (11), 1461–1473. doi:10.1038/onc.2016.304

Zheng, Y., Chen, Z., Han, Y., Han, L., Zou, X., Zhou, B., et al. (2020). Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment. *Nat. Commun.* 11 (1), 6268. doi:10.1038/s41467-020-20019-0

Zophel, D., Hof, C., and Lis, A. (2020). Altered Ca(2+) homeostasis in immune cells during aging: Role of ion channels. *Int. J. Mol. Sci.* 22 (1), E110. doi:10.3390/ijms22010110

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 November 2022
DOI 10.3389/fgene.2022.997319

OPEN ACCESS

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

REVIEWED BY
Chenyang Zhao,
Ocean University of China, China
Felipe Carlos Martin Zoppino,
CONICET Instituto de Medicina y Biologia Experimental de Cuyo (IMBECU), Argentina

*CORRESPONDENCE
Yi Yao,
yaoyi2018@whu.edu.cn
Qibin Song,
qibinsong@whu.edu.cn

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

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

RECEIVED 18 July 2022
ACCEPTED 26 October 2022
PUBLISHED 09 November 2022

CITATION
Dong Y, Yi L, Song Q and Yao Y (2022), A pyroptosis-related gene model and its correlation with the microenvironment of lung adenocarcinoma: A bioinformatics analysis and experimental verification.
Front. Genet. 13:997319.
doi: 10.3389/fgene.2022.997319

A pyroptosis-related gene model and its correlation with the microenvironment of lung adenocarcinoma: A bioinformatics analysis and experimental verification

Yi Dong[1†], Lina Yi[1†], Qibin Song[1,2*] and Yi Yao[1,2*]

[1]Cancer Center, Renmin Hospital of Wuhan University, Wuhan, China, [2]Hubei Provincial Research Center for Precision Medicine of Cancer, Wuhan, China

Background: Non-small cell lung cancer, comprising lung adenocarcinoma (LUAD) and lung squamous cell carcinoma, is one of the leading causes of cancer-related mortality. Pyroptosis is a new form of programmed cell death involved in cancer development. The relationship between LUAD and pyroptosis is unclear. This research aims to investigate this relationship and develop a stratified clinical model based on pyroptosis-related genes (PRGs).

Methods: We analyzed the data of LUAD from The Cancer Genome Atlas (TCGA) and evaluated the expression of 48 PRGs to identify the differentially expressed genes. Then, constructing the risk model using the least absolute shrinkage and selection operator and the Cox regression method to find the gene signatures. The functional enrichment, immune cell infiltration, tumor mutational burden (TMB), and expression of immune checkpoints were compared to investigate the potential mechanism. The IC50 of common drugs was evaluated and compared. The inflammasome activation assay and lactate dehydrogenase (LDH) assay of NLR-family CARD-containing protein 4 (NLRC4) were also performed to confirm the role of pyroptosis in LUAD.

Results: The pyroptosis-related model accurately predicted the prognosis of patients with LUAD, with the low-risk group exhibiting a higher survival probability. The risk score was an independent prognostic factor for survival. The stratified patients exhibited distinct tumor microenvironments, TMB, and drug sensitivity. The validation experiments of NLRC4 confirmed its role in inducing pyroptosis *via* promoting IL-1 maturation.

Conclusion: PRGs regulated the tumor microenvironment and influenced the outcome of LUAD. NLRC4 may function as a hub gene in the process of LUAD.

KEYWORDS

pyroptosis, cancer, immunity, microenvironment, NLRC4, bioinformatics

Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide, with 1.8 million newly diagnosed cases and 1.6 million deaths per year (Bray, et al., 2018). More than 85% of patients with lung cancer are diagnosed with non-small cell lung cancer (NSCLC), with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma being the most common subtypes (Duma, et al., 2019). Numerous studies have been conducted on the strategies against lung cancer, leading to a significant decrease in mortality and an increase in survival due to advances in diagnosis and treatment. For those tissue biopsies might not be technically feasible can take liquid biopsies, detect ctDNA in plasma, and obtain molecular information (Thai, et al., 2021). With the advent of novel technologies such as next-generation sequencing, lung cancer treatment has entered the molecular era, the development of target therapy and immunotherapy have marked a turning point in cancer treatment. Early clinical trials with these therapies revealed rapid and long-lasting responses in 14%–20% of patients with pretreated advanced NSCLC (Wu, et al., 2020). Although great progress has been made in this field, significant obstacles remain: the mechanism of target therapy resistance; the treatment for rare somatic activating oncogene mutations; as well as the biomarkers to predict the response of anti-cancer therapy (Hirsch, et al., 2017). During the past decade, although the discovery of predictive biomarkers has created new therapeutic opportunities with targeted therapy and immunotherapy, there are still many limitations. Programmed death-ligand 1 (PD-L1) is a predictive biomarker used to guide treatment decisions for its expression is associated with an increased likelihood of response to programmed death-1 (PD-1) pathway blockade, but responses to immune checkpoint inhibitors (ICIs) can also be seen in patients with no tumor PD-L1 expression. High tumor mutational burden (TMB) might be predictive of response to ICIs without any prospective validation. Moreover, the mutation of on- and off-target resistance has not been solved successfully, which needs further investigation for better clinical practice (Thai, et al., 2021). Thus, new biomarkers need to be urgently discovered to learn more about the pathogenesis of LUAD so that new targets can be developed.

According to the prevailing opinion, the 10 hallmarks of cancer lead to cancer initiation and progression (Hanahan and Weinberg, 2011). Among these, the ability to resist cell death and escape immunological damage was discussed in this study. Cell death is a physiological process that regulates cell proliferation, stress response, and immunological response, as well as inhibits tumor growth. Besides apoptosis and necrosis, autophagy, anoikis, and pyroptosis were also described (Fernandes-Alnemri, et al., 2007). Pyroptosis is an inflammatory form of programmed cell death initiated by caspase 1/4/5/11. It is triggered by certain inflammasomes and results in cell swelling, plasma membrane lysis, chromatin fragmentation, and releasing of intracellular proinflammatory components (Fang, et al., 2020). Recent scientific advances have led to the identification of numerous genes as essential regulators of pyroptosis. The relationship between cancer and pyroptosis is intricate. Paclitaxel and cisplatin triggered pyroptosis in A549 *via* the caspase 3/gasdermin E (GSDME) pathway; the effectiveness was associated with the expression of GSDME (Zhang, et al., 2019). However, another study revealed that GSDME was associated with radioresistant lung cancer cells, and its expression was indicative of a poor prognosis for LUAD (Wei, et al., 2020). Apart from its prognostic value, pyroptosis increases the immunological defenses of the host and contributes to the release of tumor antigens, a previous study demonstrated that pyroptosis stimulated the activation of $CD8^+$ T lymphocytes and inhibited tumor growth and migration (Tang, et al., 2020).

Available evidence indicates that pyroptosis plays an essential but complex role in tumors. Less attention has been paid to its precise mechanism in LUAD, particularly the impact of the hub genes on the microenvironment and anti-cancer immunity. With new technology, the appropriate treatment based on the gene expression pattern to optimize therapeutic benefits has developed. In this study, we examined the expression pattern of pyroptosis-related genes (PRGs) in LUAD, assessed their clinical utility, investigated the relationship between pyroptosis and TME, and provided therapeutic suggestions. We also explored the role of NLR-family CARD-containing protein 4 (NLRC4) in LUAD to find a new biomarker or therapeutic target.

Materials and methods

Dataset collection and processing

The LUAD mRNA sequencing data and corresponding clinical data up to 29 April 2022, were obtained from The Cancer Genome Atlas (TCGA) website (https://portal.gdc.cancer.gov). The gene expression profiles were normalized using the "limma" package. 48 PRGs were extracted from the previous study (Wang, et al., 2021). Their information is shown in Table 1. Figure 1 illustrates the complete workflow of the study.

Identification of differentially expressed PRGs

The differentially expressed genes (DEGs) were extracted based on RNA expression between the normal and tumor samples in the entire cohort. The following criteria were used: $|logFC| > 0.5$ and FDR < 0.05. The "pheatmap" package was

TABLE 1 Differences in expression of pyroptosis-related genes between LUAD and normal samples.

Gene symbol	Full name	logFC	p	FDR
AIM2	Absent in melanoma 2	2.956308	7.49E-15	3.00E-14
CARD8	Caspase recruitment domain-containing protein 8	−0.43852	2.48E-11	7.94E-11
CASP1	Caspase 1	−0.7689	2.17E-17	1.16E-16
CASP3	Caspase 3	0.857751	5.28E-26	8.44E-25
CASP4	Caspase 4	0.300571	0.003537	0.005305042
CASP5	Caspase 5	−1.17359	6.90E-20	5.52E-19
CASP6	Caspase 6	1.078872	3.68E-29	8.83E-28
CASP8	Caspase 8	0.563998	1.51E-13	5.56E-13
DDX3X	DEAD-box helicase 3 X-linked	−0.31273	2.68E-06	6.13E-06
GBP1	Guanylate binding protein 1	−0.14397	0.000133	0.000246347
GBP2	Guanylate binding protein 2	0.272685	0.860409	0.860408749
GBP5	Guanylate binding protein 5	0.732585	0.081692	0.095639142
GSDMA	Gasdermin A	1.113508	1.82E-05	3.64E-05
GSDMB	Gasdermin B	1.851111	4.64E-19	3.18E-18
GSDMC	Gasdermin C	2.735457	2.13E-15	9.30E-15
GSDMD	Gasdermin D	0.297322	0.009873	0.013539723
GSDME	Gasdermin E	0.677512	0.013544	0.018058017
GZMA	Granzyme A	0.155768	0.044481	0.054745779
GZMB	Granzyme B	0.117265	0.006247	0.008819273
HMGB1	High mobility group box 1	−0.10315	0.000176	0.000313767
IFI16	interferon γ-inducible protein 16	0.649383	0.000206	0.000352345
IL18	Interleukin 18	−0.2668	0.000113	0.00021771
IL1B	Interleukin 1β	−0.83492	6.83E-13	2.34E-12
IRF1	Interferon regulatory factor 1	−0.50135	1.76E-07	4.45E-07
IRF2	Interferon regulatory factor 2	−0.09168	0.014351	0.018616998
IRF8	Interferon regulatory factor 8	−0.83976	1.29E-15	6.18E-15
MEFV	Mediterranean fever	−1.50059	2.85E-25	3.43E-24
NAIP	neuronal apoptosis inhibitor protein	−0.0802	0.004813	0.007001444
NEK7	NIMA-related kinase 7	−0.75799	5.68E-22	5.45E-21
NLRC3	NOD-like receptor family CARD domain containing 3	−0.03511	0.048797	0.058556187
NLRC4	NOD-like receptor family CARD domain containing 4	−1.89292	1.90E-32	9.11E-31
NLRC5	NOD-like receptor family CARD domain containing 5	0.459784	0.028978	0.036603825
NLRP1	NOD-like receptor (NLR) family pyrin domain-containing 1	−0.31658	9.75E-06	2.03E-05
NLRP12	NOD-like receptor (NLR) family pyrin domain-containing 12	−0.35668	2.91E-08	8.22E-08
NLRP2	NOD-like receptor (NLR) family pyrin domain-containing 2	1.315508	0.652115	0.665990211
NLRP3	NOD-like receptor (NLR) family pyrin domain-containing 3	−0.67298	5.87E-08	1.57E-07
NLRP6	NOD-like receptor (NLR) family pyrin domain-containing 6	0.307339	0.236802	0.258329677
NLRP7	NOD-like receptor (NLR) family pyrin domain-containing 7	2.136525	5.36E-07	1.29E-06
NLRP9	NOD-like receptor (NLR) family pyrin domain-containing 9	0.665345	0.095985	0.109697104
NOD1	Nucleotide-binding oligomerization domain-containing protein 1	−0.25127	3.69E-06	8.05E-06
NOD2	Nucleotide-binding oligomerization domain-containing protein 2	0.051304	0.450026	0.480027341
NR2C2	nuclear receptor subfamily 2, group C, member 2	0.214582	0.593925	0.619747761
P2RX7	P2X purinoceptor 7	−0.8271	1.67E-17	1.00E-16
PKN1	Serine/threonine-protein kinase N1	−0.16294	0.000444	0.000687913
PKN2	Serine/threonine-protein kinase N2	0.230417	0.163305	0.182294012
PYCARD	Apoptosis-associated speck-like protein containing a CARD	−0.2949	0.000231	0.000382215
TNF	Tumor necrosis factor	−0.23493	0.000321	0.000513452
ZBP1	Z-DNA-binding protein 1	1.384278	1.02E-09	3.06E-09

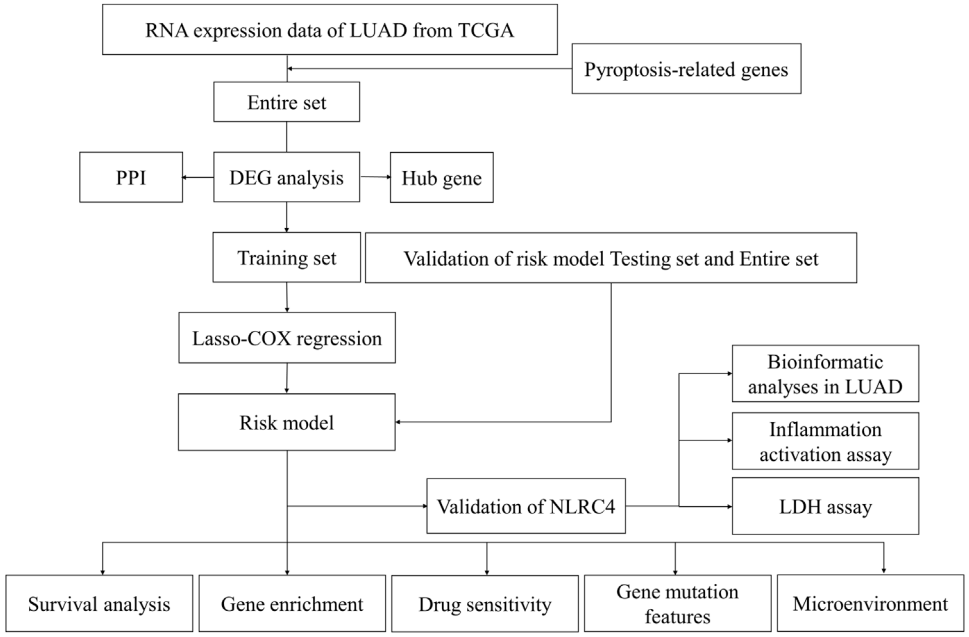


FIGURE 1
Schematic overview of the workflow in this study.

used to create a heatmap of all these genes. In addition to this, we also performed a preliminary analysis of the correlation among PRGs based on their expression. A protein-protein interaction (PPI) network for DEGs was constructed using the Search Tool for the Retrieval of Interacting Genes (https://string-db.org/). Cytoscape was then used for additional display, and the hub genes were obtained using the Maximal Clique Centrality (MCC) method. The association network of PRGs was used to emphasize the significance.

Construction and validation of the pyroptosis-related gene prognostic model

In addition, we randomly separated the data into two groups (the training and testing datasets). The clinical data were compared and summarized in Table 2.The associations between each gene and survival status in the TCGA cohort were investigated using Cox regression analysis to establish the predictive value of PRGs. We chose a cutoff p value of 0.05 to avoid omissions. The least absolute shrinkage and selection operator (LASSO) method was used for variable selection and shrinkage using the "glmnet" R package in the training dataset to filter the candidate genes and generate the prognostic model. After determining the penalty parameter (λ) for the model, the risk scores were computed for each patient based on the expression level of the extracted gene, and the risk score formula was as follows: Risk score = $\sum_i^n X_i \times Y_i$ (X: coefficients, Y: gene expression level).

Prognostic value and validation of the risk model

The clinical data (age and stage) of all patients were retrieved and analyzed in conjunction with the risk score. The univariate and multivariate Cox regression models were used to determine the independence of components. The "prcomp" function in the "stats" R package was used to conduct principal component analysis (PCA) based on the gene signature. The sensitivity and specificity of the prognostic model were tested using time-dependent receiver operating characteristic (ROC) analysis. A 1-, 3-, and 5-year ROC curve study was performed using the "survival" "survminer" and "timeROC" R packages. Patients with LUAD were divided into subgroups with low and high risk based on the median risk score. Kaplan-Meier (K-M) analysis was used to evaluate the prognostic value of the risk model, including the overall survival (OS) in all three datasets, and the progression-free survival (PFS) in the whole dataset. This was done to ensure that the risk model was stable. The OS in patients with different stages was also analyzed.

TABLE 2 Comparison of clinical characteristics. (Chi-squared test).

Parameter	Type	Total	Testing	Training	*p* value
Age	<60	137 (27.57%)	68 (27.42%)	69 (27.71%)	0.2666
	≥60	360 (72.43%)	180 (72.58%)	180 (72.29%)	
Gender	FEMALE	268 (53.92%)	144 (58.06%)	124 (49.8%)	0.0787
	MALE	229 (46.08%)	104 (41.94%)	125 (50.2%)	
Stage	Stage I	265 (53.32%)	145 (58.47%)	120 (48.19%)	0.1572
	Stage II	118 (23.74%)	50 (20.16%)	68 (27.31%)	
	Stage III	80 (16.1%)	40 (16.13%)	40 (16.06%)	
	Stage IV	26 (5.23%)	12 (4.84%)	14 (5.62%)	
	unknow	8 (1.61%)	1 (0.4%)	7 (2.81%)	
T	T1	168 (33.8%)	95 (38.31%)	73 (29.32%)	0.1522
	T2	262 (52.72%)	125 (50.4%)	137 (55.02%)	
	T3	45 (9.05%)	18 (7.26%)	27 (10.84%)	
	T4	19 (3.82%)	9 (3.63%)	10 (4.02%)	
	unknow	3 (0.6%)	1 (0.4%)	2 (0.8%)	
N	N0	320 (64.39%)	167 (67.34%)	153 (61.45%)	0.2467
	N1	93 (18.71%)	41 (16.53%)	52 (20.88%)	
	N2	70 (14.08%)	33 (13.31%)	37 (14.86%)	
	N3	2 (0.4%)	2 (0.81%)	0 (0%)	
	unknow	12 (2.41%)	5 (2.02%)	7 (2.81%)	
M	M0	328 (66%)	165 (66.53%)	163 (65.46%)	0.9883
	M1	25 (5.03%)	12 (4.84%)	13 (5.22%)	
	unknow	144 (28.97%)	71 (28.63%)	73 (29.32%)	

Functional enrichment analysis of the differentially expressed genes between the low- and high-risk groups

The full cohort of patients with LUAD was separated into two subgroups according to the median risk score. And the clinical features of the two risk-groups were also compared. DEGs between low- and high-risk groups were filtered based on the criteria of $|\log_2 FC| \geq 1$ and FDR <0.05. The "clusterProfiler" and "enrichplot" packages were used to conduct Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses based on these DEGs.

Genetic and clinicopathological features based on the risk model

TMB represents the number of mutations per million bases in tumor tissues. The tumor tissues with a higher TMB are detected by the immune system more quickly, enhancing the efficacy of immunotherapy (Chan, et al., 2019). The TMB score for each patient with LUAD was calculated using the somatic mutation data of patients obtained from the TCGA database to investigate the association between the expression pattern of PRGs, TMB, and immunity. In addition, the relationship between the TMB and the risk score derived by the stratified model, as well as the impact of these factors on survival were also evaluated. The infiltration score of 16 immune cells and the activity of 13 immune-related pathways were measured using the single-sample gene set enrichment analysis (ssGSEA) function of "gsva" R package. Furthermore, we analyzed the immune cells infiltration with different methods. The Estimation of Stromal and Immunological Cells in Malignant Tumors using the Expression Data (ESTIMATE, https://bioinformatics.mdanderson.org/estimate/index.html) website provided the stromal score, immune score, and ESTIMATE score of samples in the TCGA database, which were applied for further validation.

Analyses of the sensitivity to anti-cancer therapy based on the model

ICIs are an efficient method for treating various types of cancer. In this study, the expression levels of immune checkpoint molecules, such as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and PD-L1, were compared between the two groups to see whether the stratified model could identify patients with LUAD having a favorable response to ICIs. Besides comparing the expression of immune-checkpoint-related genes, we also

analyzed the Tumor Immune Dysfunction and Exclusion (TIDE) score to identify ICI-beneficial patients. Following the uploading of the gene expression file as the instruction, the TIDE score was acquired from the website (http://tide.dfci.harvard.edu/). The Cancer Immunome Atlas (TCIA) was applied to conduct comprehensive immunogenomic analyses based on the sequencing data from TCGA. We also used "pRRophetic" R package to evaluate drug sensitivity, which was determined by the concentration that could inhibit 50% of cellular growth (IC50) based on the risk level.

Bioinformatics validation of NLRC4

We compared the pattern of NLRC4 expression in pan-cancer. Also, we separated the complete dataset into low- and high-expression groups according to the mRNA level of NLRC4 in LUAD. GO enrichment and KEGG analyses were performed on the basis of the DEGs obtained from the two groups. The correlations with the LUAD microenvironment, including TMB, immune cell infiltrations, and immune checkpoints were also evaluated. Additionally, clinical factors, such as survival and medication sensitivity were applied to investigate its value in clinical. The methods used here were mentioned above. In addition, we acquired the immunophenoscore (IPS) for each LUAD from the Cancer Immunome Database (https://tcia.at/home). It served as a predictor of anti-CTLA-4 and anti-PD-1 treatment response (Charoentong, et al., 2017). And we compared the differences in IPS between the low- and high-expression groups.

Cell culture and inflammasome activation assay

Human embryonic kidney 293 cells (HEK-293T) were cultured on Dulbecco's modified Eagle's medium (DMEM, Gibco, US) containing 10% fetal bovine serum (Gibco, US) and antibiotics (penicillin and streptomycin). H1299 cells were cultured in an incubator at 37°C in the presence of 5% CO_2 with Roswell Park Memorial Institute1640 (RPMI-1640, Gibco, US) containing 10% fetal bovine serum and antibiotics. HEK-293T or H1299 cells were plated into six-well tissue culture plates overnight. HEK-293T cells were transfected with Polyethylenimine and Human influenza hemagglutinin (HA) tagged full-length NLRC4, pro-caspase 1, apoptosis-related specific protein (ASC), and pro-interleukin (IL)-1β to imitate the activation of the inflammasome, while H1299 cells were transfected with HA-tagged full-length NLRC4 for the lactate dehydrogenase (LDH) assay. The cells were collected for further study after 48 h incubation.

Lactate dehydrogenase (LDH) assay

LDH is a stable cytoplasmic enzyme present in every cell. When the plasma membrane is compromised, LDH is rapidly released into the culture supernatant. LDH leakage was evaluated using a colorimetric LDH test kit (Promega, US) following the manufacturer's protocols. The absorbance value of each group was compared with the absorbance value of the control group.

Quantitative polymerase chain reaction and Western blotting

RNA was extracted using TRIzol (ThermoFisher, US) following the manufacturer's protocols. The isolated RNA was reverse-transcribed into cDNA using a first-strand cDNA synthesis kit (ABclonal, CN). Quantitative PCR was performed in triplicate using the MonAmp SYBR Green qPCR Mix kit (Monad, CN), and was performed on an RT-PCR instrument (Bio-RAD, US). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a reference.

The primer sequences for qPCR were as follows:

RT-PCR GAPDH Forward 5′-TGACTTCAACAGCGACACCCA-3′

RT-PCR GAPDH Reverse 5′-CACCCTGTTGCTGTAGCCAAA-3′

RT-PCR NLRC4 Forward: 5′-GTGTTCTCCCACAAGTTTGA-3′

RT-PCR NLRC4 Reverse: 5′-AGTAACCATTCCCCTTGGTC-3′

RT-PCR caspase-1 Forward: 5′-CAGACAAGGGTGCTGAACAA -3′

RT-PCR caspase-1 Reverse: 5′ -TCGGAATAACGGAGTCAATCA-3′

RT-PCR IL-1β Forward 5′-ATGGCAGAAGTACCTGAGCTC-3′

RT-PCR IL-1β Reverse 5′-TTAGGAAGACACAAATTGCATG-3′

The cells were collected and lysed with 2% sodium dodecyl sulfate (SDS) in the presence of complete protease inhibitor mixture. The loading volume was mainly adjusted by the expression of tubulin. After adjusting by 1 × loading buffer, equal amounts of protein were loaded onto SDS-PAGE gels and transferred to NC membranes. The membranes were blocked for 1 h at RT with 5% milk and subsequently incubated at 4°C overnight with the primary antibody. After washing the membrane, incubated with secondary antibody for 1 h at RT. Then, washed the membrane again before chemiluminescence analysis. The following antibodies were used: HA (CST, United States, cat#3724S), caspase-1 (CST, United States, cat#3866T), IL-1β (CST, United States, cat#12242S), and tubulin (proteintech, CN, cat#I0004491). The secondary anti-

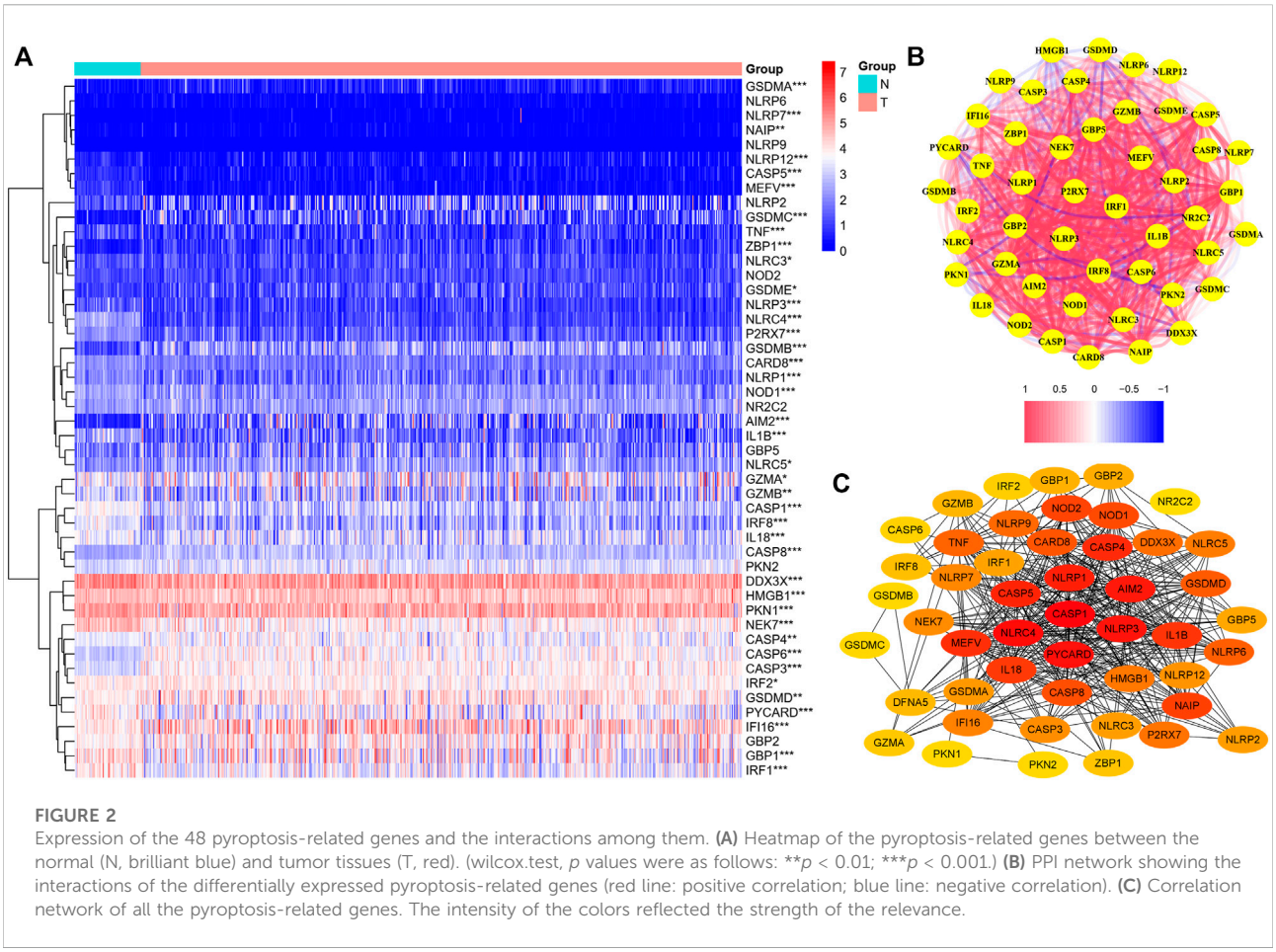


FIGURE 2
Expression of the 48 pyroptosis-related genes and the interactions among them. **(A)** Heatmap of the pyroptosis-related genes between the normal (N, brilliant blue) and tumor tissues (T, red). (wilcox.test, *p* values were as follows: $^{**}p < 0.01$; $^{***}p < 0.001$.) **(B)** PPI network showing the interactions of the differentially expressed pyroptosis-related genes (red line: positive correlation; blue line: negative correlation). **(C)** Correlation network of all the pyroptosis-related genes. The intensity of the colors reflected the strength of the relevance.

mouse (proteintech, CN, cat# SA00001-1) and anti-rabbit immunoglobulins (proteintech, CN, cat# SA00001-4) were used.

Statistical analysis

All statistical analyses in this study were conducted on R software (version 4.1.2). The statistical tests were all two sided. A p value < 0.05 indicated a statistically significant difference. The statistical significance of two groups was evaluated using the Student t test and Wilcoxon test. For correlation analysis was conducted by Spearman and Pearson cor test. Chi-squared test was used to caculate composition difference.

Results

Identification of differentially expressed genes between normal and tumor tissues

We found that 21 of the 48 PRGs were differentially expressed (logFC >0.5, FDR <0.05) in the TCGA dataset consisting of 59 normal and 535 tumor tissues; 10 genes were downregulated (*IL-1β*, *NLRC4*, *CASP5*, *IRF1*, *CASP1*, *NLRP3*, *NEK7*, *P2RX7*, *MEFV*, and *IRF1*), while the remaining were upregulated (*CASP8*, *GSDME*, *CASP3*, *CASP6*, *GSDMA*, *GSDMB*, *NLRP7*, *GSDMC*, *ZBP1*, *IFI16*, and *AIM2*). The RNA levels of these genes are depicted as heatmaps in Figure 2A. The correlation network containing all PRGs is shown in Figure 2B (the red line represents positive correlations, while the blue line represents negative correlations), and most of these genes displayed a positive relationship. We conducted a PPI analysis for these genes to further investigate the interactions of these PRGs. Among these, *CASP1*, *NLRC4*, *NLRP1*, *CASP5*, *NLRP3*, *CASP8*, and *AIM2* were the hub genes, which had the maximum interactions with other genes. The results are presented in Figure 2C.

Development of a prognostic gene model in the TCGA training cohort and validation

A total of 497 samples with complete survival information were collected for further analysis. Initially, the genes associated

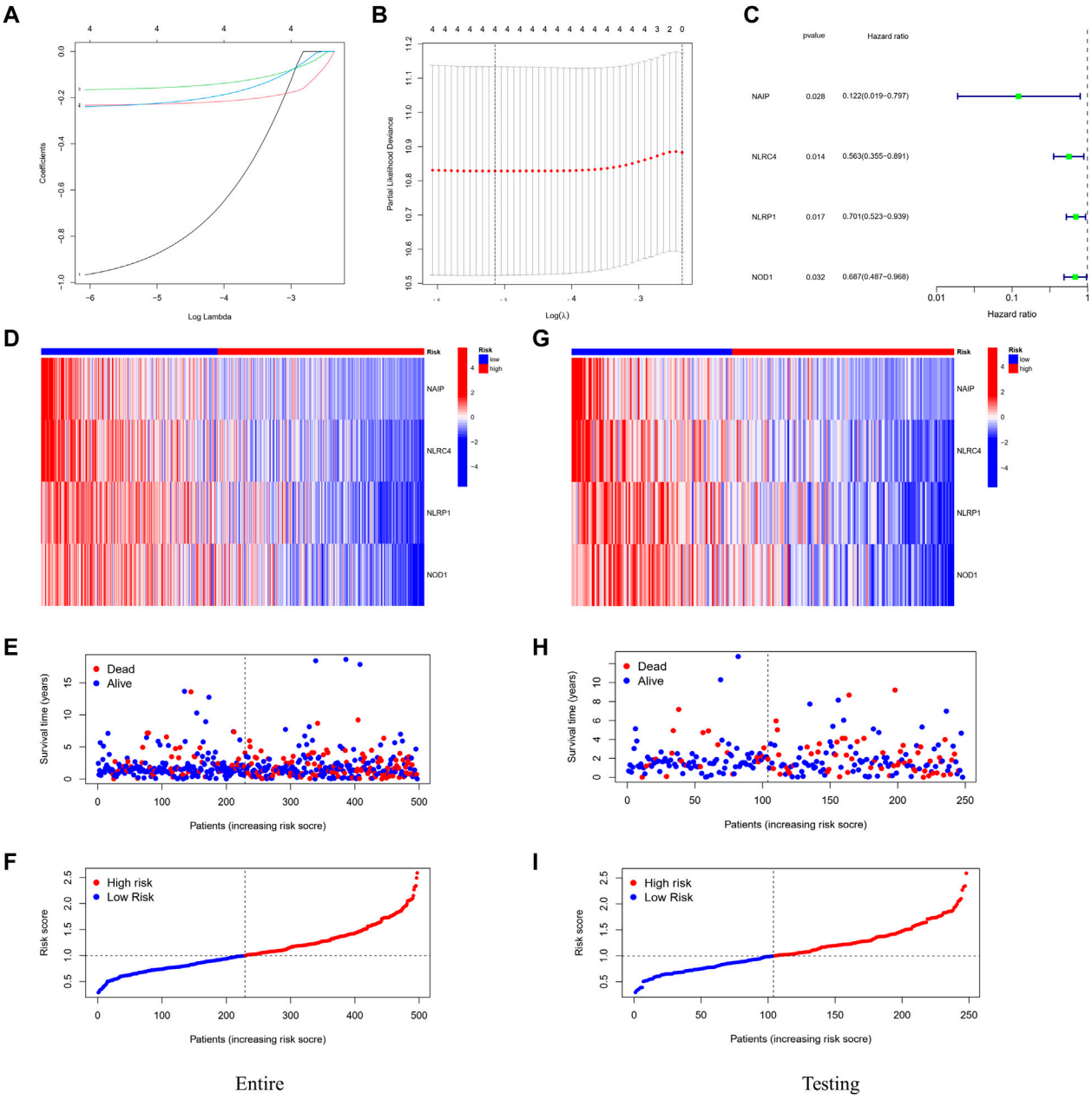


FIGURE 3
Construction of the prognostic stratification model. **(A)** LASSO coefficient profiles. **(B)** Four candidate genes obtained by LASSO regression. **(C)** Four optimal genes obtained by multivariable Cox regression analysis. **(D)** Expression pattern of the four optimal genes between tumor and normal specimens in the entire dataset. **(E)** Survival status for each patient (low-risk population: on the left side of the dotted line; high-risk population: on the right side of the dotted line). **(F)** Distribution of patients based on the risk score. **(G–I)** Expression pattern of the four optimal genes; survival status and distribution of patients based on the risk score between tumor and normal specimens in the testing dataset.

with survival were evaluated using univariate Cox regression analysis on the training dataset. We retained the four genes *NLRC4*, NOD-like receptor (NLR) family pyrin domain-containing 1 (*NLRP1*), Nucleotide-binding oligomerization domain-containing protein 1 (*NOD1*), and NLR family apoptosis inhibitory proteins (*NAIP*) for risk model development based on the optimal value from LASSO Cox regression analysis (Figures 3A–C). The risk score formula was: Risk score = (−1.01608731367098 × NAIP exp) + (−0.235231123651649 × NLRC4 exp) +

TABLE 3 comparison of clinical features between two-risk groups in entire set. (Chi-squared test).

Parameter	Type	Total	High	Low	*p* value
Age	<60	133 (27.2%)	77 (29.39%)	56 (24.67%)	0.2856
	≥60	356 (72.8%)	185 (70.61%)	171 (75.33%)	
Gender	FEMALE	265 (54.19%)	133 (50.76%)	132 (58.15%)	0.1226
	MALE	224 (45.81%)	129 (49.24%)	95 (41.85%)	
Stage	Stage I	265 (54.19%)	125 (47.71%)	140 (61.67%)	0.0115
	Stage II	118 (24.13%)	68 (25.95%)	50 (22.03%)	
	Stage III	80 (16.36%)	51 (19.47%)	29 (12.78%)	
	Stage IV	26 (5.32%)	18 (6.87%)	8 (3.52%)	
T	T1	167 (34.15%)	73 (27.86%)	94 (41.41%)	0.0154
	T2	257 (52.56%)	151 (57.63%)	106 (46.7%)	
	T3	44 (9%)	27 (10.31%)	17 (7.49%)	
	T4	18 (3.68%)	10 (3.82%)	8 (3.52%)	
	unknow	3 (0.61%)	1 (0.38%)	2 (0.88%)	
N	N0	315 (64.42%)	154 (58.78%)	161 (70.93%)	0.0046
	N1	91 (18.61%)	58 (22.14%)	33 (14.54%)	
	N2	70 (14.31%)	46 (17.56%)	24 (10.57%)	
	N3	2 (0.41%)	0 (0%)	2 (0.88%)	
	unknow	11 (2.25%)	4 (1.53%)	7 (3.08%)	
M	M0	322 (65.85%)	184 (70.23%)	138 (60.79%)	0.2148
	M1	25 (5.11%)	18 (6.87%)	7 (3.08%)	
	unknow	142 (29.04%)	60 (22.9%)	82 (36.12%)	

(−0.169828624865761 × NLRP1 exp) + (−0.246566494613787 × NOD1 exp). In addition, patients in various datasets were divided into low- and high-risk subgroups based on the median risk score. And the clinical parameters comparisons in the entire cohort were shown in Table 3. All these genes were relatively suppressed in the population at high-risk. Patients in the high-risk group (on the left side of the dashed line) died more often and lived shorter times than those in the low-risk group (on the right side of the dashed line). This showed that LUAD with a high score might have a worse outcome. Additionally, Figures 3D–F shows the whole dataset, and Figures 3G–I shows the testing dataset.

Independent prognostic value of the risk model and its clinical application

The univariate and multivariable Cox regression analyses were conducted to determine whether the risk score generated by the gene signature model could be employed as an independent prognostic factor. Both the stage and the risk score were found to be independent predictors of poor survival in the TCGA cohorts *via* the univariate Cox regression analysis ($p < 0.001$, Figure 4A,B). We also performed PCA based on tumor and normal specimens and found that the 48 PRGs completely distinguished LUAD samples (Figure 4C). The area under the ROC curve was 0.685 for 1-year survival, 0.610 for 3-year survival, and 0.618 for 5-year survival (Figure 4D). They were further confirmed in the testing and entire datasets, indicating that the pyroptosis-associated risk score had a strong and dependable capacity to predict the prognosis for patients with LUAD. All three datasets showed poorer OS in the high-risk group ($p < 0.05$), as well as the PFS in the entire dataset ($p = 0.026$). These results demonstrated that patients with a high-risk score had a worse prognosis (Figure 4G–J). We further examined the risk model by comparing the OS in patients with different stages. We discovered that patients in the high-risk group had a worse prognosis for both early and advanced stages of lung cancer (Figure 4K,L: $p = 0.021$ in stages I and II, $p < 0.001$ in stages III and IV).

Distinct biological processes, TME, and treatment decision characteristics in LUAD were based on the risk model

Further, we identified DEGs to examine the differences in gene functions and pathways between the subgroups divided by the risk model. Between these, 508 DEGs were identified in the total TCGA cohort. The GO enrichment analysis and KEGG

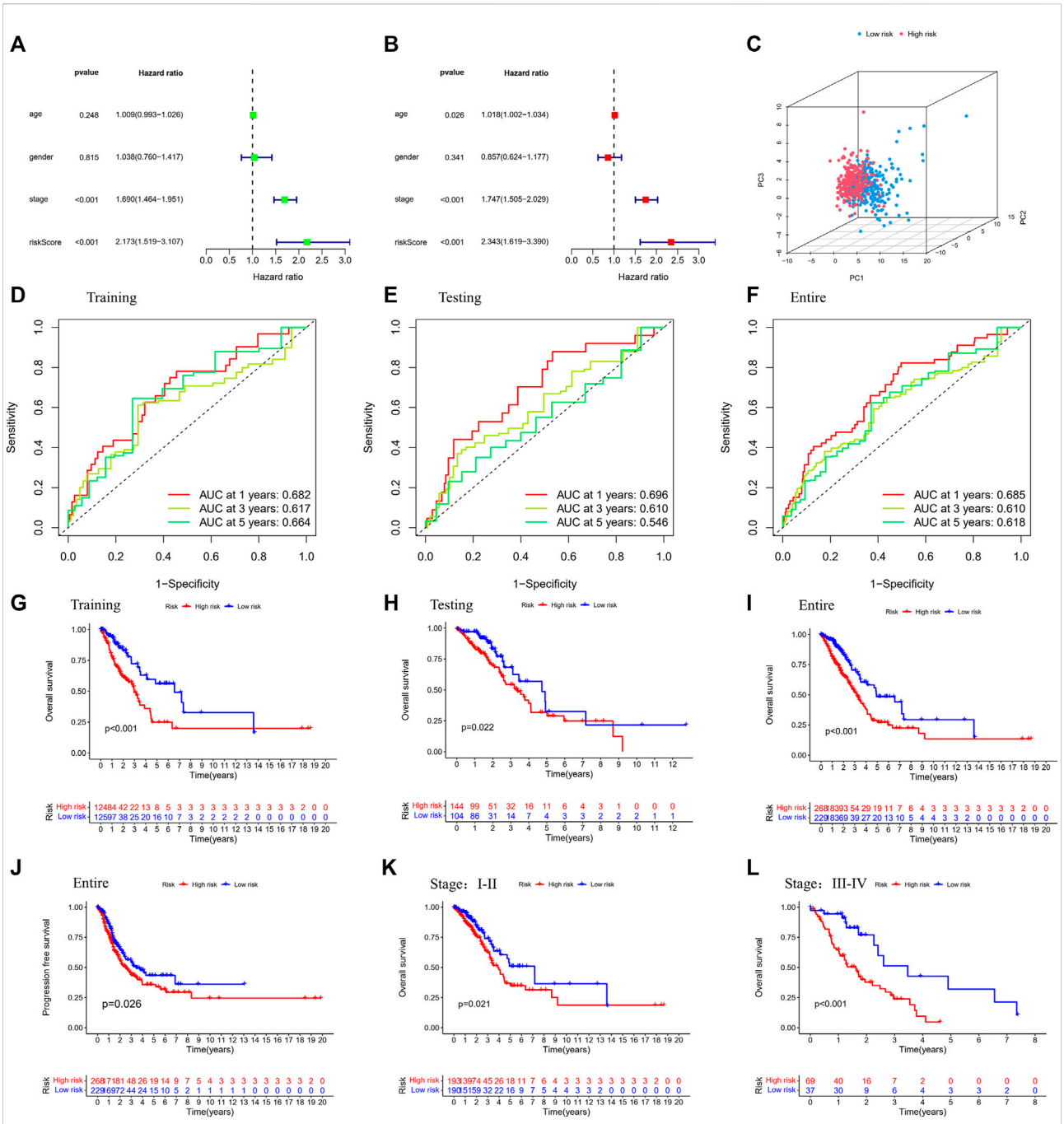


FIGURE 4
Validation of the prognostic stratification model. **(A)** Univariate analysis for the entire dataset. **(B)** Multivariate analysis for the entire dataset. **(C)** PCA plot for LUAD in the entire dataset. **(D–F)** Time-dependent ROC curves for LUAD in the training, testing, and entire datasets. **(G–I)** K-M survival analysis of different risk groups in the training, testing, and entire datasets. **(J)** PFS of the entire dataset. **(K,L)** K-M survival analysis of patients with different stages in low- and high-risk groups. (Chi-squared test).

pathway analyses were performed. The GO enrichment of DEGs was primarily associated with the immune system, including cytokine generation, immune response regulation, chemokine binding, and inflammatory cell chemotaxis (Figure 5A). Regarding the KEGG pathway, we discovered that it corresponded with the GO analysis, which included the chemokine signaling pathway, the B cell receptor signaling pathway, the cytokine-cytokine receptor interaction, and so on (Figure 5B). It was obvious that this pyroptosis-related model might be related to immunity, which could help us differentiate

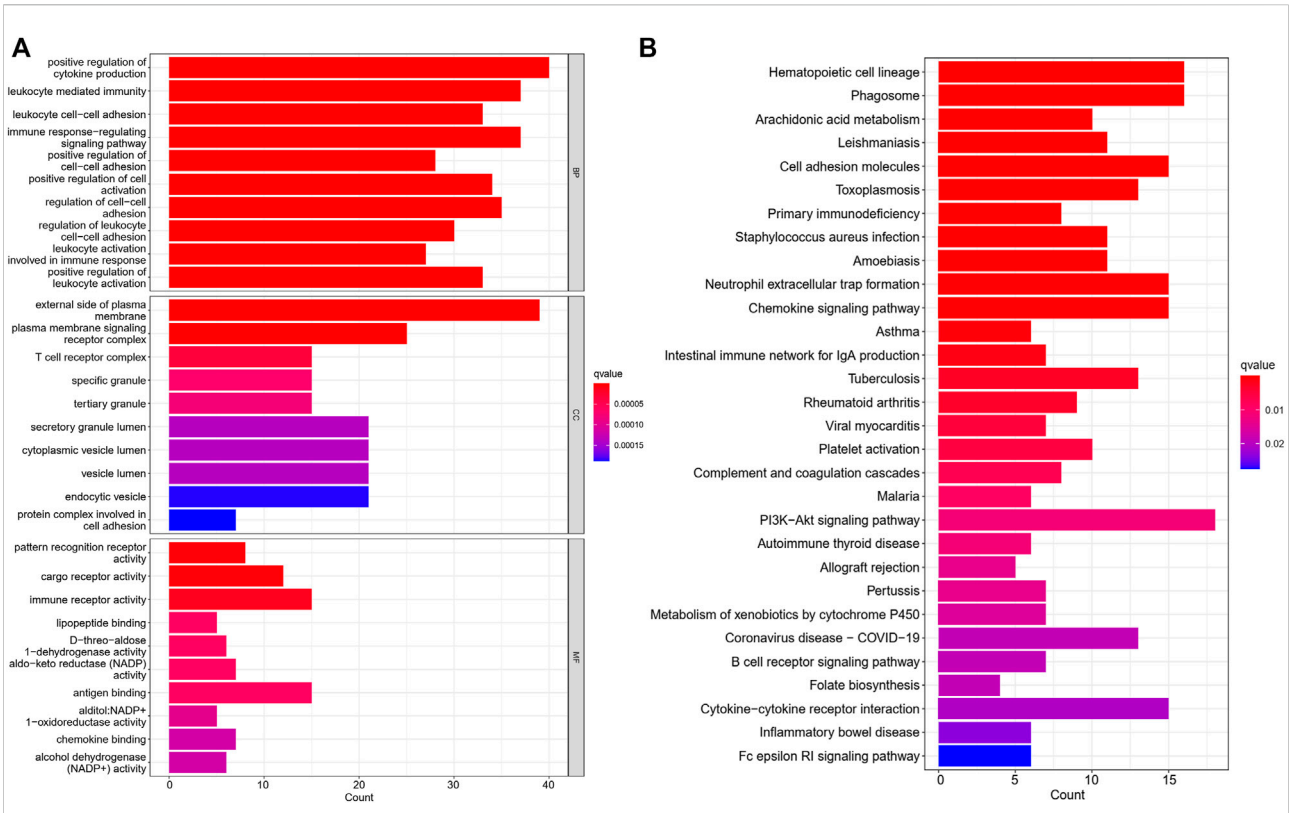


FIGURE 5
The GO and KEGG analyses between the two risk groups **(A)** Bar graph for GO enrichment. **(B)** Bar graph for KEGG pathways (the longer bar means the enrichment of more genes, and the increasing intensity of red color means more obvious differences; q-value: the adjusted *p* value).

LUAD with varying immunological status. The ssGSEA was used to evaluate the enrichment scores of 16 types of immune cells and the activity of 13 immune-related pathways. The low-risk subgroup showed higher proportions of $CD8^+$ T cells, neutrophils, natural killer (NK) cells, T helper (Th) cells (Th1 and Th2), tumor-infiltrating lymphocytes, and regulatory T (Treg) cells compared with the high-risk subgroup (Figure 6A). In the entire TCGA cohort, the activity of all 13 immunological pathways was lower in the high-risk group (Figure 6B). The immune cell infiltration was further investigated using various techniques. Consistent with previous findings, many immune-infiltrating cell subpopulations, including effector memory B cells, $CD8^+$ T cells, $CD4^+$ T cells, and NK cells, were significantly enriched in the low-risk group (Figure 6C).

We analyzed the correlations between the risk model and TMB. To highlight the importance of risk score. We compared the most prevalent mutation genes between the two risk groups by collecting the LUAD mutation data from TCGA. We found that the high-risk group had a significantly greater mutational rate (p = 0.0003, Figure 7A–C). The top list included Tumor protein P53 (TP53) and Kirsten rat sarcoma virus (KRAS). Previous studies found that the TP53 gene was a suppressor gene, its mutation had a significant impact on cancer risk, while KRAS mutation correlated with a low response rate to gefitinib in LUAD (Greathouse, et al., 2018; Reck, et al., 2021). The mutations of these oncogenes based on the risk model might indicate some potential relations in the drug resistance of LUAD. Next, the TMB survival probability was explored. Patients with a low TMB and a high-risk level had the worst outcome (Figure 7D,E).

The correlation between risk score and immune checkpoint expression was further estimated based on the results of GO and KEGG enrichment. As shown in Figure 8A–H, low-risk LUAD exhibited higher expression of various molecular markers (PD-L1, CTLA4, Lymphocyte-activation gene 3: LAG3, CD27, CD80), indicating a superior immunotherapeutic response. Another newly identified predictor TIDE is frequently employed and strongly advised for evaluating the immune response and immune evasion (Jiang, et al., 2018). However, our study revealed that the TIDE expression dramatically increased in the low-risk group, indicating an immune escape phenotype in low-risk group, which is contradictory to immune checkpoint analysis and will be discussed in detail below. In addition, we examined the degree of immune cell infiltration (immune score) and stromal cell infiltration (stromal score) across three unique patterns. The high-risk patients had the lowest immune score compared with the low-risk patients. Also, they had a lower stromal score, indicating that

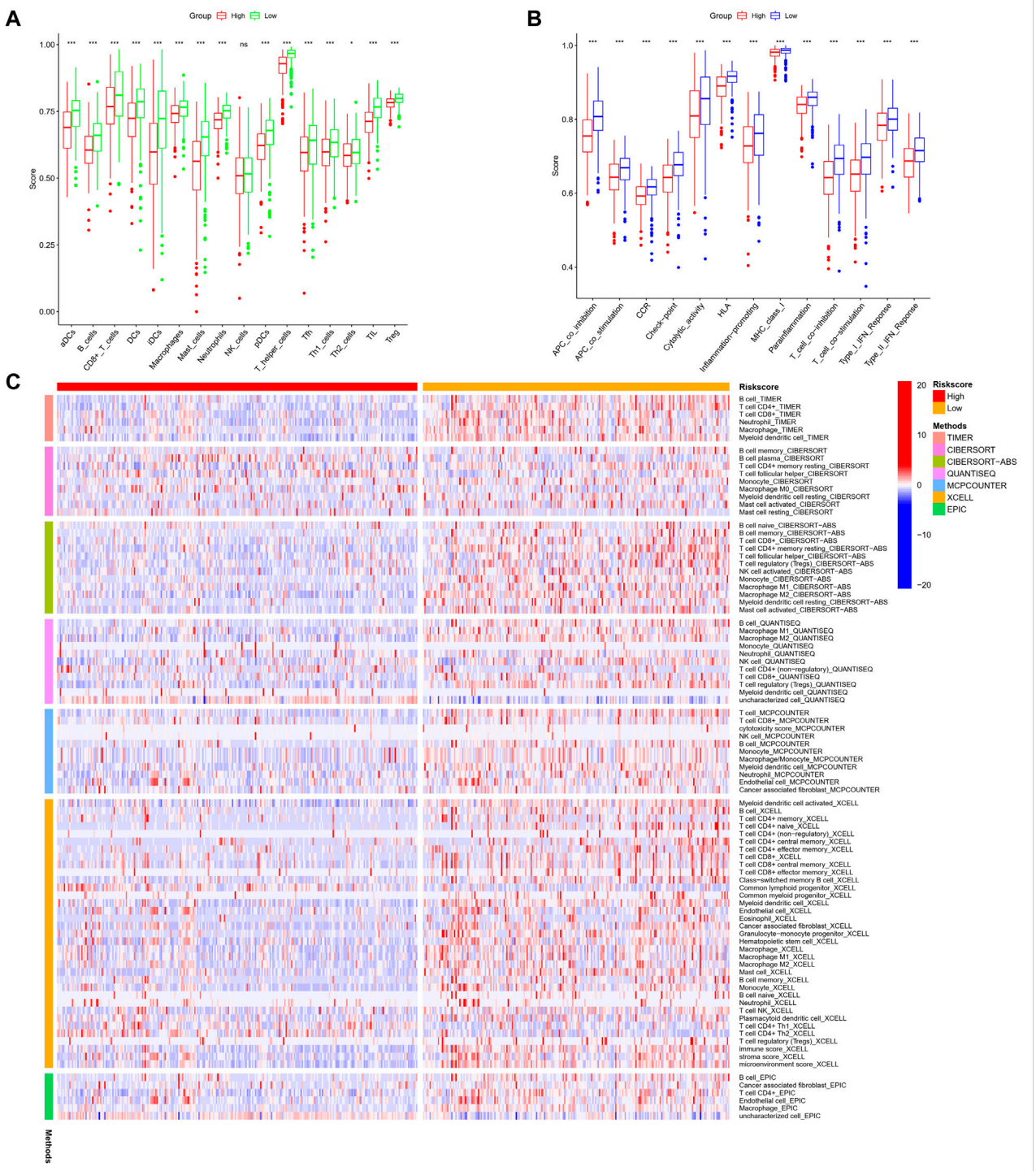


FIGURE 6
(A,B) Comparison of the enrichment scores of 16 types of immune cells and 13 immune-related pathways based on the ssGSEA scores between low- (green box) and high-risk (red box) groups in the TCGA cohort. **(C)** Comparison of the high- and low-risk groups using different methods. (wilcox.test, *p* values were as follows: **p* < 0.05; ***p* < 0.01; ****p* < 0.001).

high-risk LUAD had fewer nontumor components, such as immune cells and stromal cells, which might correspond with a poorer response to immunotherapy (Figure 8I–L). The aforementioned results demonstrated that the difference in tumor pyroptosis patterns might play a crucial role in mediating the clinical response to ICIs treatment through the impact on TMB,

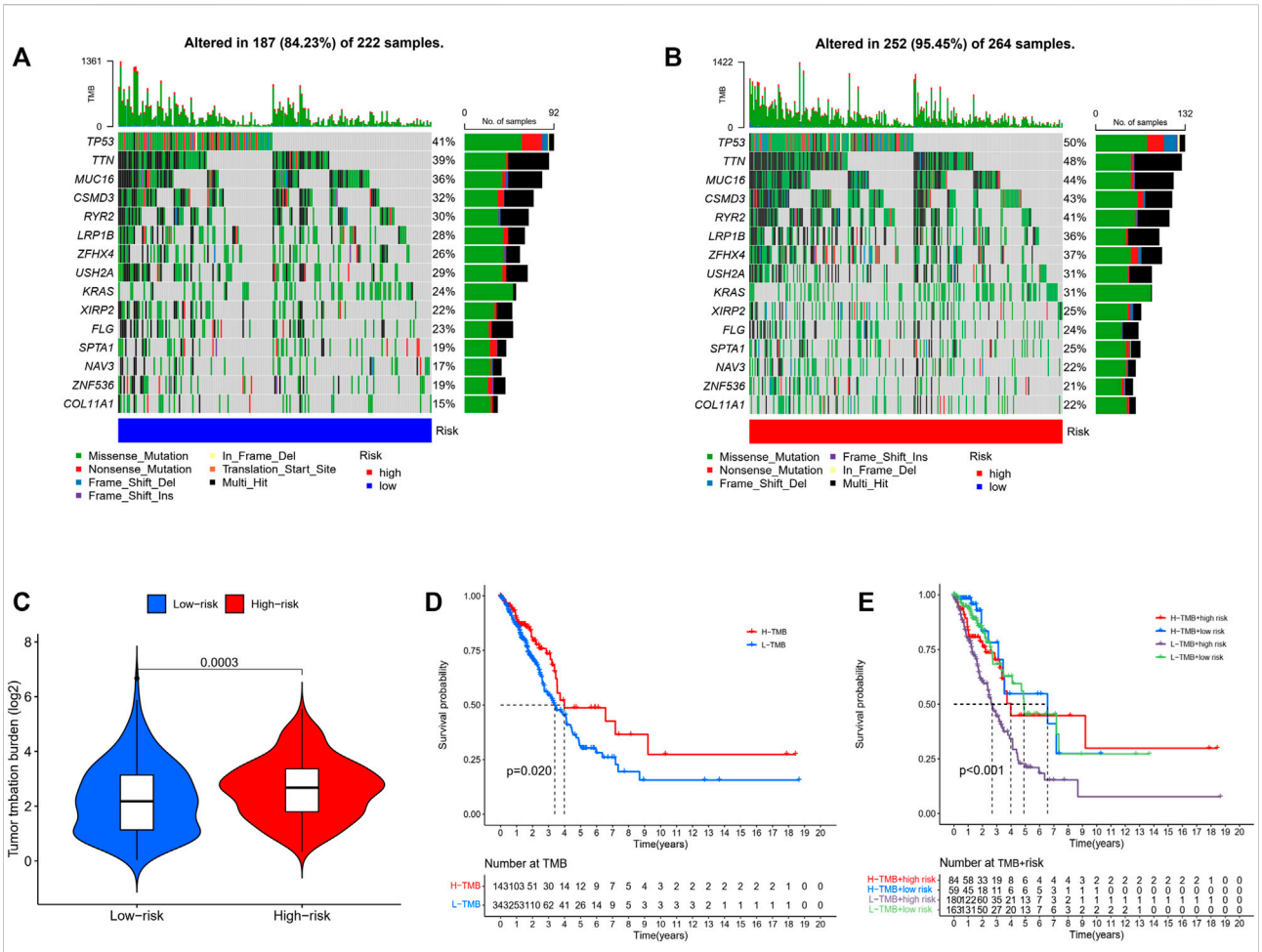


FIGURE 7
(A,B) Mutations in common genes between low- and high-risk groups. **(C)** Comparison of the tumor mutation burden between low- and high-risk groups. (Wilcoxon test) **(D)** Comparison of the survival probability with different levels of tumor mutation burden. **(E)** The survival probability with different levels of tumor mutation burden and risk score. (Chi-squared test).

immune cell infiltration, immunogenicity, and checkpoint expression, providing insights into the crucial role of pyroptosis in the regulation of the immune microenvironment of LUAD.

We investigated common anti-tumor drugs in LUAD to confirm whether the PRGs-related risk model could provide treatment suggestions. We discovered that the high-risk group was more sensitive to erlotinib, gemcitabine, docetaxel, paclitaxel, and rapamycin than the low-risk group, while patients with low-risk were more sensitive to gefitinib and crizotinib (Figure 9).

NLRC4 induced caspase-1-dependent pyroptosis and could arrest the progression of LUAD

We extracted NLRC4 for further investigation to find the potential mechanisms associated with pyroptosis in LUAD. We compared the expression level of NLRC4 in pan-cancer and found that it was much lower in LUAD tissues than in normal tissues (Figure 10). It was validated in the TCGA cohort, where its expression was inversely linked with the outcome (Figure 11A–E). It also had a positive relationship with Toll-like receptor 4 (TLR4), which might promote the synthesis or release of pro- and anti-inflammatory cytokines and chemokines *via* the activation of transcription factors such as NF-κB, as well as the activation of adaptive immunity (Figure 11F) (Pinto, et al., 2011). Furthermore, we investigated the relationship between NLRC4 and immunity. NLRC4 expression negatively correlated with the proportions of Tregs, naive B cells, and plasma cells, and positively correlated with the proportions of dendritic cells, macrophages, and so on (Figure 12A,B). Regarding the tumor tissues, the higher the level of NLRC4, the higher the immune score was (Figure 12C). The expression of NLRC4 was also found to have some correlations with immune checkpoints, especially a

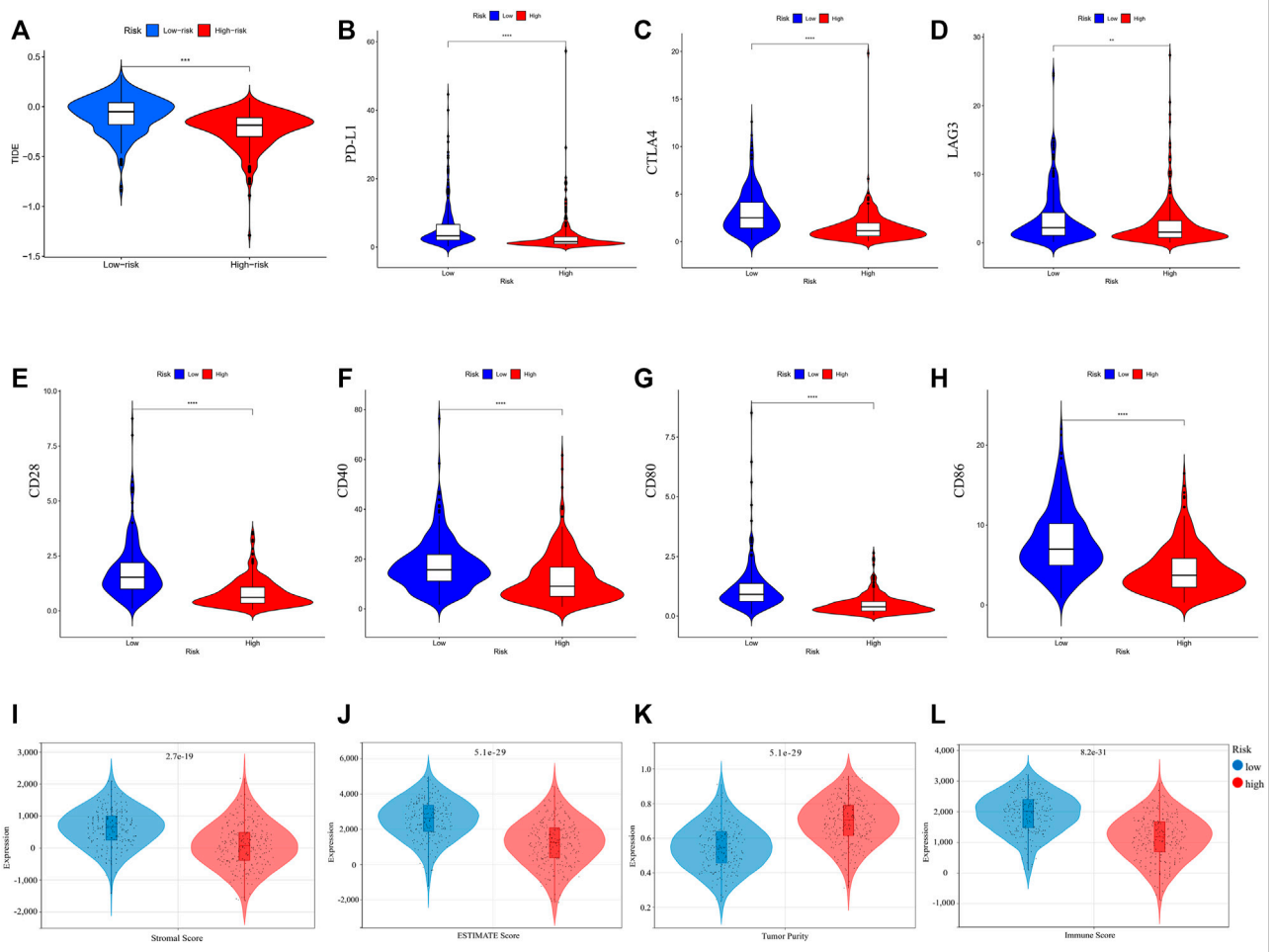


FIGURE 8
(A) TIDE score of different risk groups. **(B–H)** Comparison of immune checkpoint expression between different risk groups. **(I–L)** Tumor immune microenvironment of LUAD comparisons between high-risk and low-risk groups. **(I)** stromal score; **(J)** ESTIMATE score; **(K)** tumor purity; and **(L)** immune score. (wilcox.test, *p* values were as follows: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$).

positive correlation with CD40 and CD28, which participated in T cell activation. However, the correlation between NLRC4 and TMB was negative ($p < 0.01$) (Figure 12D,E). The drug sensitivity analyses revealed that LUAD with a higher level of NLRC4 was more sensitive to crizotinib (Figure 12F,G), which needs clinical data for validation. The immunotherapy prediction indicated that higher expression levels correlated with a better response to anti-PD-1 and anti-CTLA-4 therapy (Figure 12H–K).

An inflammasome activation system was established in HEK293T cells to study further the probable mechanism linking NLRC4, pyroptosis, and LUAD. We found that overexpressing NLRC4, caspase-1, and IL-1 in HEK293T cells enhanced the maturation of caspase-1 and IL-1. As shown in Figure 13, cleaved IL-1, caspase 1, and LDH levels increased in HEK293T and H1299 cells when NLRC4 was overexpressed, implying the death of more cancer cells (Figure 13).

Discussion

As an alternative mode of programmed cell death, necroptosis can elicit strong adaptive immune responses that may defend against tumor progression (Salomon, et al., 2018). Necrosis-induced inflammation facilitates tissue repair responses (which are largely immunoregulatory) but not effective anticancer immunity (Tang, et al., 2020). Activation of the canonical programmed necrosis includes the formation of a complex containing receptor-interacting protein kinase-1 (RIP1), RIP3 and recruitment of mixed lineage kinase domain-like protein (MLKL), leading to lytic cell death accompanied by *de novo* production of proinflammatory mediators (Snyder, et al., 2019). Another key mediator in necroptosis includes cellular inhibitor of apoptosis protein 1 (cIAP1/2), deubiquitinase cylindromatosis (CYLD), and caspase-8. CYLD deubiquitinates RPK1 and subsequently

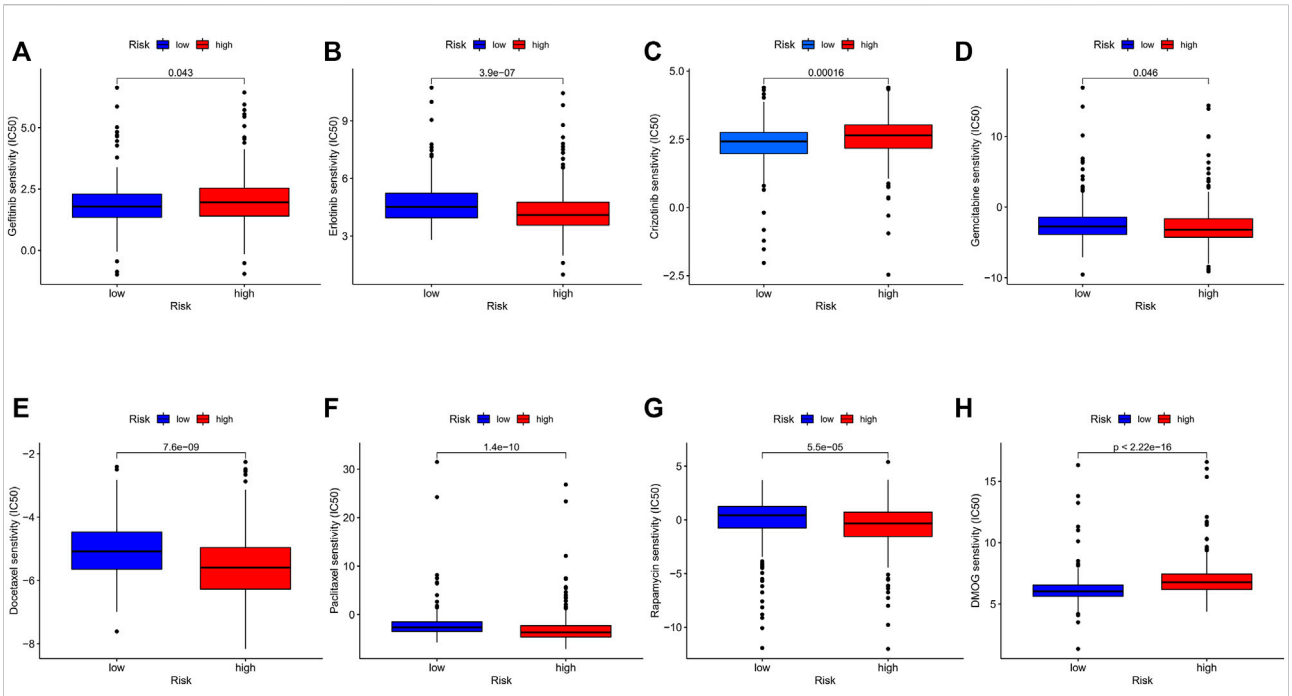


FIGURE 9
(A–H) Comparison of IC50 between different risk groups. (wilcox.test, *p* values were as follows: **p* < 0.05; ***p* < 0.01; ****p* < 0.001).

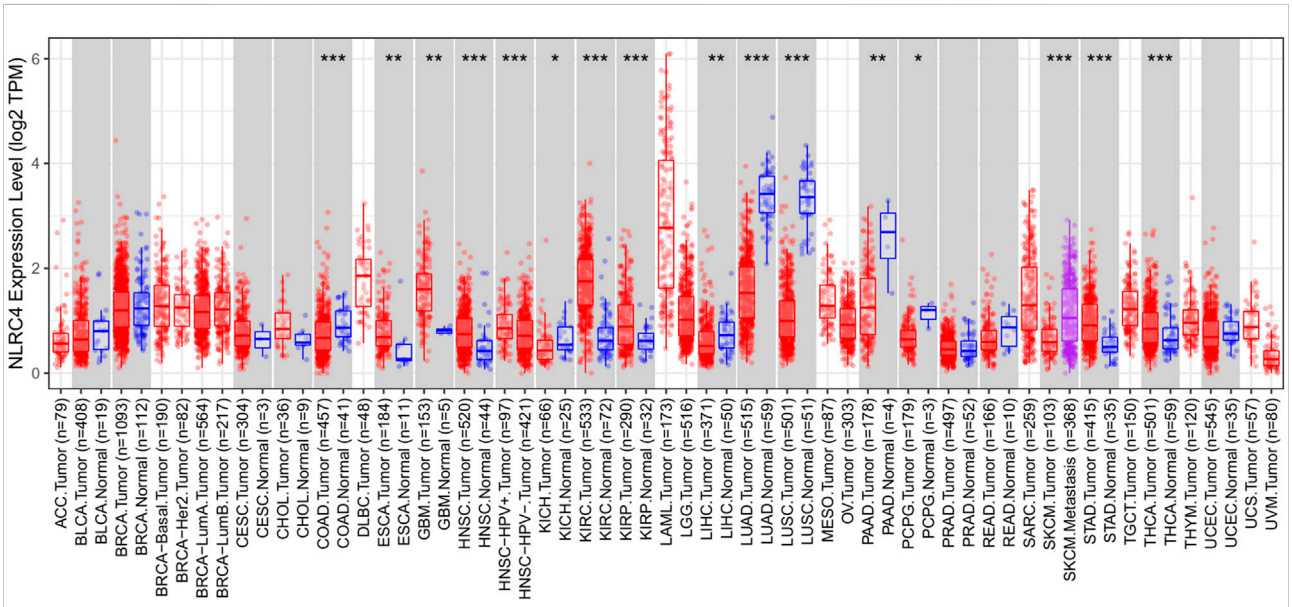


FIGURE 10
Expression of NLRC4 in pan-cancer compared with normal samples. Expression of NLRC4 was significantly lower in LUAD samples than in normal samples.

limits the sustained activation of NF-κB signaling, while cIAP1/2 polyubiquitinates RIPK1 to induce it (Gong, et al., 2019). Although previous studies pointed out its antitumor immunogenicity through $CD8^+$ T cells, recent research proved that immune-mediated tumor control by necroptotic fibroblasts requires nuclear factor κB (NF-κB) activation within dying cells

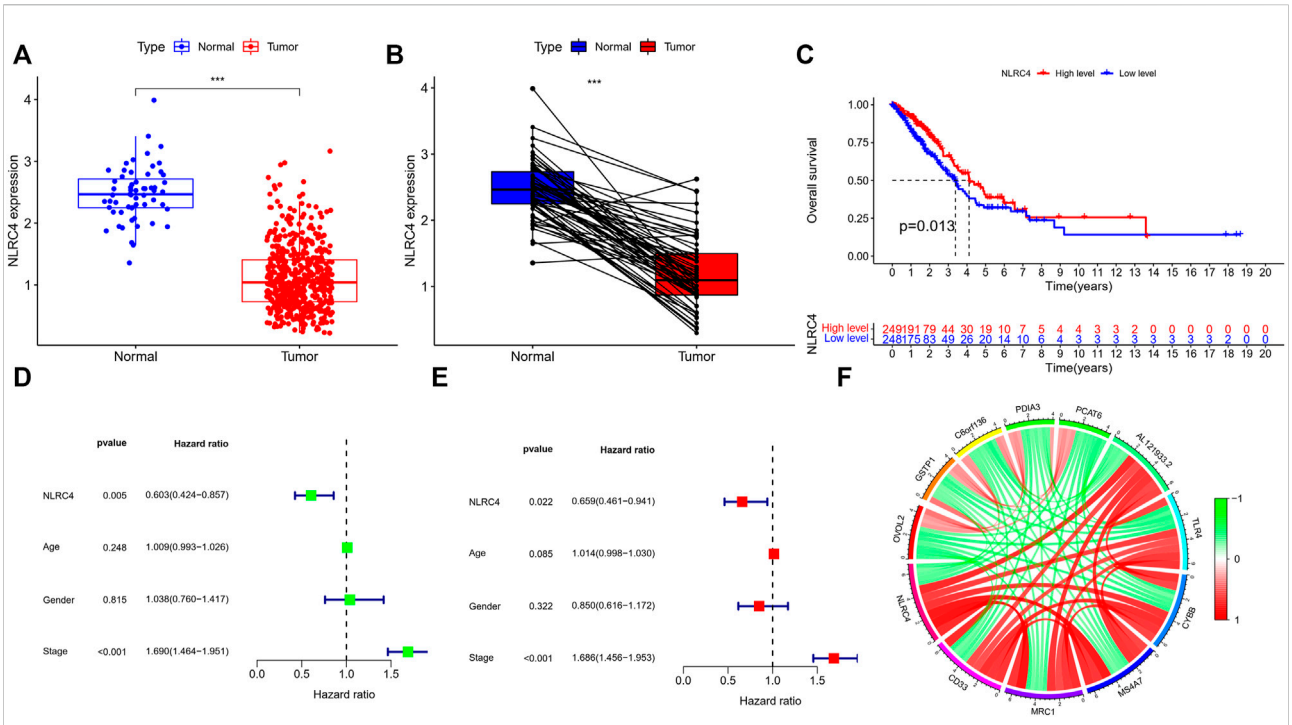


FIGURE 11
(A,B) Expression of NLRC4 in TCGA cohort compared with normal samples (wilcox test, ***$p < 0.001$). **(C)** K-M survival analyses based on the expression of NLRC4 in the LUAD-TCGA cohort. (Chi-squared test) **(D,E)** Univariate and multivariate analyses of different parameters related to survival. **(F)** The correlation of NLRC4 with other genes (red line: positive correlation; green line: negative correlation).

but not MLKL-mediated and cell lysis-dependent DAMP release (Snyder, et al., 2019). And the role of NF-κB activation in necroptosis-provoked antitumor immunity is controversial (Yatim, et al., 2015). MLKL translocation to the plasma membrane is triggered by RIPK3-mediated phosphorylation of MLKL, which results in membrane damage. Consequently, potassium ion efflux may further activate NLRP3 *via* NEK7, which may constitute a cross-talk with the pyroptosis pathway (Tang, et al., 2020). Pyroptosis is a new nonapoptotic form of programmed cell death strongly associated with the inflammatory response by triggering the production of cytokines, such as IL-1 and IL-18, playing a crucial role in modulating the progression of cancer (Hsu, et al., 2021). We developed a PRGs model to analyze the effect of pyroptosis on LUAD progression and possible biochemical pathways to address the potential of integrating these two modalities comprehensively. We also generated a landscape of the differences in the LUAD microenvironment based on these gene signatures to develop personalized anti-cancer therapeutic strategies. In this study, the mRNA levels of 48 PRGs were measured in LUAD and normal tissues, thus obtaining DEGs. We constructed a pyroptosis-related gene model to identify two pyroptosis patterns distinguished by different biological processes and immunological features. When the risk model was applied in a clinical setting, the score accurately predicted the prognosis of individual patients with LUAD. Patients with high scores typically had shorter survival times. The pyroptosis-related score demonstrated substantial correlations with PD-L1, CTLA-4, and immunophenotype, confirming the ability of the risk model to predict the immunotherapy response. In addition, patients with different scores had varying sensitivity to target therapy or chemotherapy, thus providing some suggestions for individualized anti-cancer therapies. Overall, this study showed how pyroptosis influenced the microenvironment in LUAD and highlighted its value in predicting the response to anti-cancer treatment.

Inflammation by innate immune cells designed to fight infections and heal wounds can contribute to the initiation and progression of cancer by secreting growth factors and reactive oxygen species, which induce genomic alterations, chronic and uncontrolled inflammation, and proliferation of malignant cells (Hanahan and Weinberg, 2011). Also, cytokines, chemokines, and a few other substances may enhance proliferation, prevent cell death, and facilitate the migration of cancer through their influence on the TME (Grivennikov, et al., 2010). The inflammasome is a cytosolic immunological signaling complex that causes inflammation and pyroptosis. It is composed of a sensor receptor and an adaptor protein (apoptosis-associated speck-like protein containing a

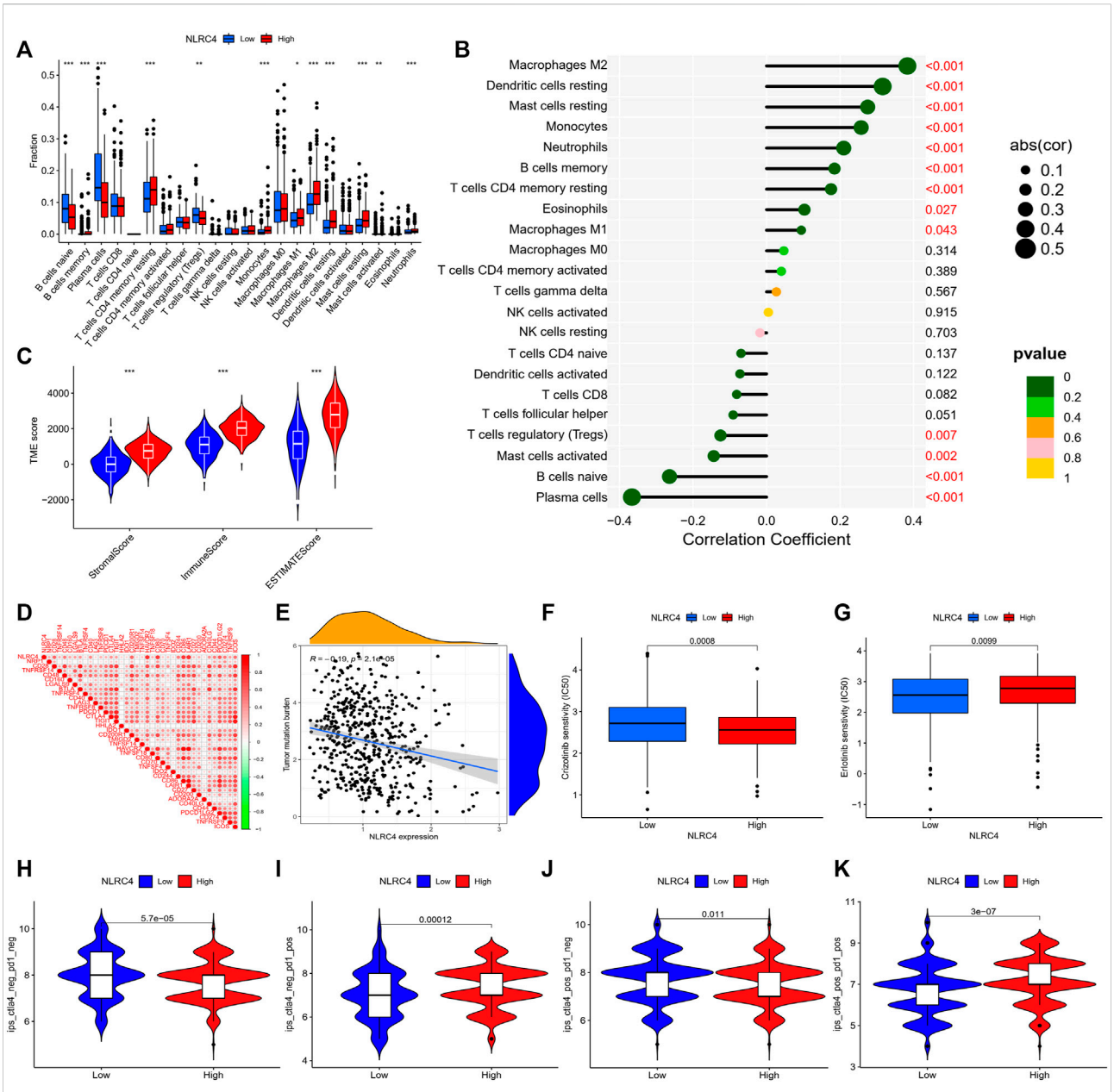


FIGURE 12
(A) Infiltrating levels of 22 immune cell types between high- and low-expression groups. **(B)** Correlation coefficient analyses of NLRC4 and different immune cells. **(C)** Tumor microenvironment between high- and low-NLRC4-expression groups (wilcox test, ***$p < 0.001$). **(D)** Correlation of NLRC4 and other immune checkpoints (Pearson cor test). **(E)** Correlation coefficient analyses of NLRC4 and TMB (Spearman cor test). **(F,G)** Drug sensitivity for LUAD with different NLRC4 expression levels. **(H–K)** TCIA analyses of LUAD with different NLRC4 expression levels. (wilcox test, p values were as follows: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

caspase activation and recruitment domains complex). A functional inflammasome is initiated by pattern recognition receptors that can detect pathogen-associated molecular patterns, danger-associated molecular patterns, and homeostasis-altering molecular processes, including nucleotide-binding domain-like receptors (NLRs), absent in melanoma 2-like (AIM2) receptors, and the newly identified pyrin domain–containing receptors (Xue, et al., 2019). Multiple inflammasomes, including NLRP3, NLRC4, NLRP1, and AIM2, may inhibit tumor initiation by influencing innate and adaptive immunity, apoptosis, and differentiation (Di Virgilio, 2013). A previous study discovered that NLRP1, NLRP3, NLRC4, and AIM2 inflammasome complex proteins had pro- or antitumoral properties, especially in breast cancer (Jin and Kim, 2020).

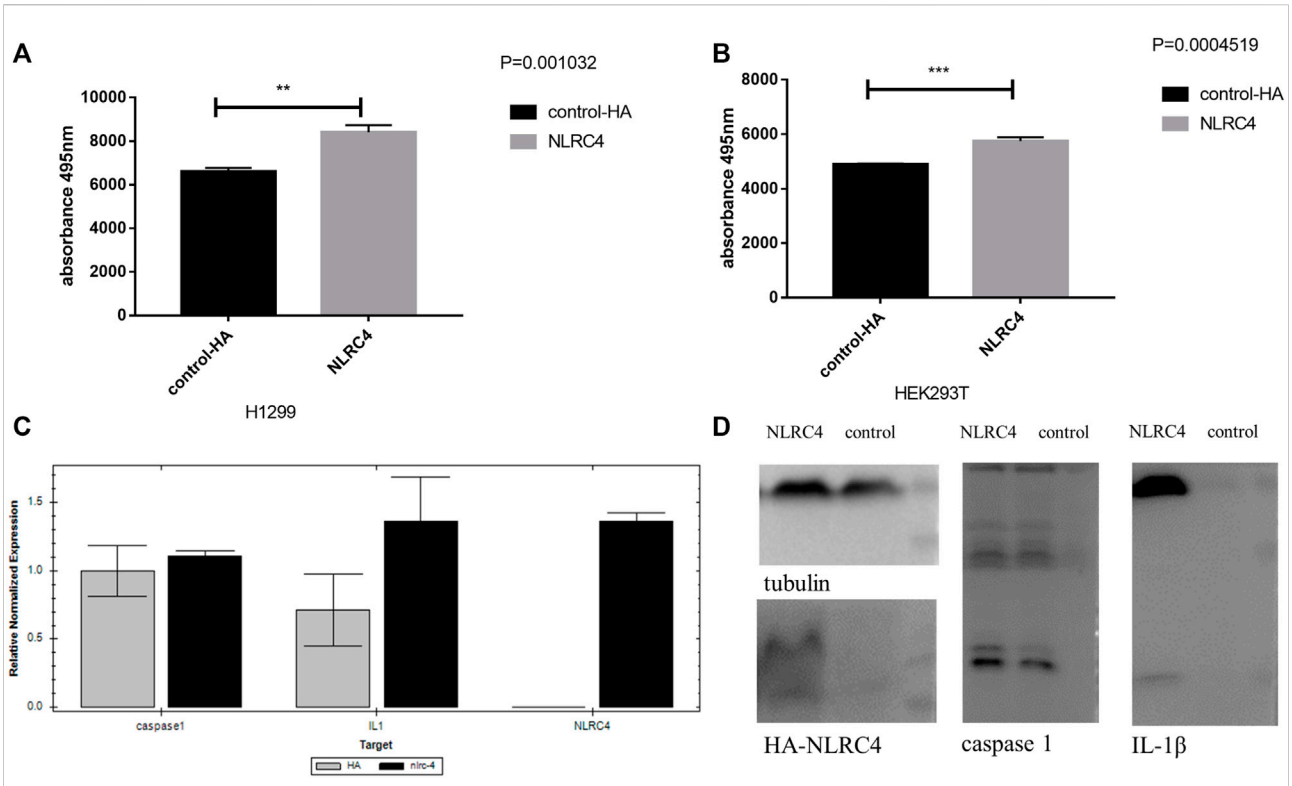


FIGURE 13
Experimental validation of NLRC4 in pyroptosis and lung adenocarcinoma. **(A,B)** LDH assay. NLRC4 was overexpressed in H1299 and HEK293T cells. We collected the supernatant in each group to test the level of LDH at the absorbance of 495 nm. **(C)** Comparison of mRNA levels of downstream related genes when NLRC4 was overexpressed, the difference of caspase 1 was not very obvious, but a higher level of IL-1β in NLRC4 overexpression group. **(D)** Activation of inflammasome in HEK293T cells by overexpressing NLRC4, pro-caspase 1, ASC and pro-IL-1β. We found that NLRC4 responded to the expression and maturation of caspase-1 and IL-1β. (*t*-test, *p* values were as follows: $^{**}p < 0.01$; $^{***}p < 0.001$).

Human NLRP1 was identified as the first protein capable of forming an inflammasome complex. Recent studies indicated that NLRP1 expression was higher in primary breast cancer tissue than in adjacent noncancerous tissue and was associated with lymph node metastasis, tumor-node-metastasis (TNM) stage, and Ki-67. Moreover, NLRP1 enhanced breast cancer cell proliferation, migration, and invasion by inducing epithelial–mesenchymal transition (EMT) (Wei, et al., 2017). The N-terminal oligomerization domain (NOD) proteins, NOD1 and NOD2, are members of the intracellular NOD-like receptor family, which can induce proinflammatory responses. NOD1 was found to be constitutively expressed in epithelial cells, helping in monitoring cytosol integrity and avoiding malignant transformation (Shin, et al., 2018). The overexpression of NOD1 significantly inhibited carcinogenesis *in vivo* and increased the sensitivity of hepatocellular carcinoma cells to chemotherapy *via* blocking the mitogen-activated protein kinase (MAPK) pathway (Ma, et al., 2020).

As with NLRC4, the inflammasome regulated the expression of adipocyte-mediated vascular endothelial growth factor A and angiogenesis, which accelerates breast cancer progression (Kolb, et al., 2016). A previous study showed that NAIP could form an inflammasome with NLRC4, which was related to protection against colitis-associated cancer (Allam, et al., 2015). It was also found to inhibit the hyperactivation of the transcription factor STAT3 as well as the generation of anti-apoptotic and proliferation-related enzymes (Allam, et al., 2015). NLRC4 was demonstrated to be essential for cytokine and chemokine production in macrophages associated with tumors, as well as the generation of IFN-producing $CD4^+$ and $CD8^+$ T cells that reduced the growth of melanoma tumors in mice (Janowski, et al., 2016). In the 293T inflammatory experiment, we discovered that NLRC4 contributed to the cleavage of pro-IL-1, resulting in pyroptosis. The activation of pro-caspase-1 was responsible for the cleavage of pro-IL-1 and pro-IL-18 proteins into mature active forms and the generation of cytokines in response to pathogen-associated molecular patterns and damage-associated molecular patterns (Rathinam and Fitzgerald, 2016). Moreover, the K-M survival curve showed that a higher level of NLRC4 expression in LUAD was associated with a more favorable prognosis.

And from the risk model, we found that all of these factors were up-expressed in low-risk group, particularly, the level of NLRC4 in lung cancer was also significantly lower in the tumor sample. To investigate its value as a prognostic factor, we compared the survival differences in the two risk groups divided by our model. In clinical practice, we observed that the risk score was an independent factor related to prognosis, with a correlation between increased risk and a worse prognosis, as measured by OS and PFS. It was also effective when we applied it to patients with different stages; high-risk patients had a low probability of survival in both early and advanced stages. Patients with a low-risk score had significantly longer longevity, suggesting that those with a high-risk score should receive more frequent clinical surveillance and appropriate treatment to avoid disease recurrence and progression.

As we discussed above, pyroptosis can trigger crosstalk between innate and adaptive immunity, modulating the cancer microenvironment to induce an immunostimulatory response (Hsu, et al., 2021). As the previous study demonstrated that the context of TME was critical to tumor development and treatment (Thorsson, et al., 2018). Our risk model had numerous similarities with a recent study in which the high-risk group displayed an immune desert and a reduced degree of immune checkpoint expression. Consistent with previous findings, the low-risk group possessed a highly active immune status, including cytokine production, immune receptor activation, and the phosphatidyl-inositol 3-kinase/serine-threonine kinase (PI3K-Akt) signaling pathway, all of which corresponded to a hot tumor phenotype. In the high-risk group, the number of essential anti-tumor-infiltrating immune cells was low, indicating an overall decrease in immune activity. Moreover, the immune microenvironment analyses indicated a global enhancement of immune cell infiltration as well as immune score in low-risk group. These findings indicated that the low-risk group might have a better immune environment.

However, whether it meant a better response to immunotherapy needed further clinical investigation. Since not all patients benefit from immunotherapy, a considerable amount of research has been devoted to the selection of the potentially beneficial population for immunotherapy. From the mechanism of immunotherapy, we can see that T cells are the soldiers of the immune response, while the activation of it requires two kinds of signals: TCR engagement with the MHC–peptide antigen complex (MHC-Ag) on an APC or a target cell, and interaction of the costimulatory receptor CD28 on the T cell with costimulatory B7 molecules (CD80/CD86). However, in response to T-cell activation, the immune checkpoints CTLA4 and PD-1 are upregulated on the T cell and bind to B7 and PD-L1/L2, respectively, to inhibit T-cell activation (Sharma, et al., 2021). Thus, PD-L1 expression is associated with an increased likelihood of response to PD-1 pathway blockade, but responses to ICIs can also be seen in patients with no tumor PD-L1 expression. Moreover, a minority of somatic mutations in tumor DNA can give rise to neoantigens, mutation-derived antigens that are recognized and targeted by the immune system. And TMB can represent a useful estimation of tumor neo-antigenic load, evolving as a relevant tool for the identification of patients likely to respond to immunotherapy (Chan, et al., 2019). Recent investigations pointed out that TMB failed to show predictive accuracy for ICIs response due to a lack of broad ICIs approval. And based on these, we wanted to find if these biomarkers could be improved combined with our risk model. According to the model, we found that patients in different risk groups showed significantly different characteristics, both in terms of TMB and immune checkpoints. The TMB was relatively high in the high-risk group, while the expression of their immune checkpoints was generally low. And the expression of TP53 was relatively higher in the high-risk group. Combined with the risk model, we find that high-risk level with lower TMB has a significantly worst survival probability, which might be an amplification effect of these two parameters. Moreover, the mutation rate of KRAS was also higher in patients with high-risk. Previous research showed that the mutation of KRAS correlated with a low response rate to gefitinib in LUAD, which was consistent with the result of drug sensitivity based on the risk model. For the potential relationship among pyroptosis, TMB and anti-cancer immunity, we need further experimental validation, but for its application in prognosis, it works well. As for these immune checkpoints, take PD-L1 for example, the high-risk group showed relatively low expression. However, from the analysis, we cannot judge the intrinsic mechanism, as the analyses of ESTIMATE showed there were more stromal components and less tumor cells. We assumed that LUAD with higher expression of particular PRGs (genes constructed the risk model) may have a higher infiltration of immune cells, which stimulated the expression of immune related markers, including PD-L1, CD80, LAG3, and so on. The expression of immune checkpoints differed significantly between the high- and low-risk groups, indicating that patients in the low-risk group may benefit more from immune checkpoint inhibitors. However, it is not simply a cause-and-effect relationship, but a mixture with predictive value. As the risk model showed, all of these PRGs' coefficients were negative, and we could not simply conclude whether the cancer immunity was induced by pyroptosis, or the pyroptosis was facilitated by anti-cancer immunity. For example, paclitaxel, a microtubule-stabilizing agent used in cancer therapy, has been demonstrated to enhance innate immunity by activating the NLRP3 inflammasome in macrophages (Zeng, et al., 2019). From the mechanism we can conclude that cancer immunity is a complex network, while the activation of T cells may be one of the key points. The expression of immune checkpoints showed significant differences between the high and low risk groups, suggesting the potential relationship between pyroptosis and cancer immunity, especially the function and activation of T cells. However, the result from TIDE reflected the profiles at the late stage of T cell dysfunction. The higher score in low-risk group indicated the

signatures of tumor immune evasion, which means a worse response to immunotherapy. This has been contradicted by the expression level of PD-L1. As we have discussed, there was no perfect biomarker for predicting the efficacy of immunotherapy. A previous study revealed that immunophenoscore claimed to have good ICIs response in melanoma but worse in TIDE. The reliable TIDE signatures were computed in five cancer types without lung cancer, and only melanoma has publicly available data on tumor expression and clinical outcome of patients treated with anti-PD1 or anti-CTLA4. The mouse tumor models revealed two stages of T cell dysfunction; anti-PD1 treatment can revive the early-stage dysfunctional T cells, and the late-stage dysfunctional T cells are resistant to ICIs reprogramming. Apart from mutation or neo-antigen load, multiple factors could affect immune checkpoint inhibitors' effectiveness, such as PD-L1 level, degree of cytotoxic T cell infiltration, antigen presentation defects, interferon signaling, mismatch repair deficiency, tumor aneuploidy, intestinal microbiota, and so on. A previous study revealed that uncontrolled activation of the PI3K-Akt pathway at the cellular level might create an immunologically tolerant TME and alter the response to ICIs (Giannone, et al., 2020). Other biomarker types can also predict T cell infiltration and ICIs response, it might achieve higher predictive performance if the risk model could be applied jointly with them.

In addition, the result of TMB revealed a substantial correlation between risk score and TMB. The combination of risk score and TMB could be used as a tool for prognostic stratification. Our study found a higher rate of mutations in high-risk patients, including TP53 and KRAS, suggesting that patients with high-risk scores may activate potentially oncogenic pathways that promote tumor initiation and proliferation. And we also evaluated the chemotherapy as well as target therapy. From the gene mutation, we found that some oncogenes mutation rates were different in the two risk groups. The drug sensitivity revealed that the sensitivity to chemotherapy and targeted therapy differed in different risk groups. Low-risk patients were more sensitive to gefitinib and crizotinib than erlotinib, gemcitabine, docetaxel, and paclitaxel.

To better understand the potential mechanism of pyroptosis in LUAD, we chose NLRC4 for validation. We discovered that its expression in the survival database was negatively connected with survival, corresponding with the negative coefficient in our construction model. NLRC4 is expressed in immune and non-immune cells, including monocytes, macrophages, and neutrophils; nevertheless, differential expression of NLRC4 has been reported in many types of tumor tissue. Studies have shown normal levels in lung cancers, but our findings based on TCGA reveal that it was significantly lower in LUAD and that its expression was associated with prognosis. And our experimental data from H1299 (lung adenocarcinoma cell line) showed that overexpression of NLRC4 could promote pyroptosis by measuring LDH released from dead cells. And the inflammasome activation assay initially validated that NLRC4 could promote IL-1β maturation, which leads to pyroptosis of lung cancer. IL-1β is a key pro-inflammatory cytokine that regulates the expression of several genes involved in the inflammatory process. Besides, previous studies showed that many clinical drugs stimulated and modulated pyroptotic pathways to inhibit tumor growth. Erlotinib decreased the phosphorylation of extracellular signal-regulated kinase (ERK) 1/2 through the PI3K–Akt signaling pathway after lipopolysaccharide treatment and downregulated the expression of TLR4 on macrophages, thereby regulating the microenvironment or systemic anti-tumor immunity (Xue, et al., 2021). Animal models of colorectal cancer have shown that $Nlrc4^{-/-}$ mice displayed increased tumor formation, reduced apoptosis in tumors, and increased proliferation of colonic epithelial cells during the early-stage (Kay, et al., 2020). Consequently, comprehensive studies of pyroptosis and the characteristics of TME in each patient could help us identify the tumor immunological characteristic and guide a more accurate treatment strategy.

Despite providing some novel insights into the immune-oncology correlations of pyroptosis in LUAD, our study had several limitations. First, our pyroptosis signature was derived from public datasets; however, its prognostic value in patients with LUAD receiving immunotherapy requires more validation. Second, the TIDE score was significantly higher in the low-risk group, indicating a poor immunotherapy response based on a previous study. Also, it appeared to be contradictory with other results, such as the PD-L1 expression and the immune score. We wondered whether TIDE or immune checkpoints alone could accurately predict the immunological efficacy of LUAD. The TIDE score combined T cell dysfunction and elimination characteristics to simulate tumor immune escape with varying proportions of tumor-infiltrating cytotoxic T cells. However, abnormalities in antigen presentation, interferon signaling, and mismatch repair can compromise the efficacy of immune checkpoint inhibition therapy, and hence combining them with other immune cells or factors may be preferable.

Conclusion

The current understanding of pyroptosis, particularly its mechanism in LUAD, is limited. In this study, we investigated the predictive value of PRGs in LUAD. Numerous PRGs were differentially expressed in normal and LUAD tissues, showing a direct correlation between pyroptosis and LUAD. Moreover, the risk score derived from our risk signature was identified as an independent risk factor for LUAD prognosis. The pyroptosis-related risk model outlined the crosstalk and regulatory roles in tumor immunity, as well as their application in cancer treatment. Our model might help in developing personalized cancer treatments for patients with LUAD.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov.

Author contributions

YD and LY performed most of the experiments and data analysis and drafted the manuscript. YY provided professional advice on the experiment design and reviewed the paper. QS managed the experimental design, reviewed the manuscript, and provided funding support. All authors read and approved the final manuscript.

Funding

Beijing Medical and Health Foundation (No. YWJKJJHKYJJ-BXS5-22006); Wu Jieping Medical Foundation (No. 320.6750.19094-18); Beijing Health Alliance Charitable Foundation (No. YXKY-WS834B); Key Youth Training Foundation of Renmin Hospital of Wuhan University (No. 2013-18).

Conflict of interest

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

Publisher's note

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

References

Allam, R., Maillard, M. H., Tardivel, A., Chennupati, V., Bega, H., Yu, C. W., et al. (2015). Epithelial NAIPs protect against colonic tumorigenesis. *J. Exp. Med.* 212, 369–383. doi:10.1084/jem.20140474

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

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

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

Di Virgilio, F. (2013). The therapeutic potential of modifying inflammasomes and NOD-like receptors. *Pharmacol. Rev.* 65, 872–905. doi:10.1124/pr.112.006171

Duma, N., Santana-Davila, R., and Molina, J. R. (2019). Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment. *Mayo Clin. Proc.* 94, 1623–1640. doi:10.1016/j.mayocp.2019.01.013

Fang, Y., Tian, S., Pan, Y., Li, W., Wang, Q., Tang, Y., et al. (2020). Pyroptosis: A new frontier in cancer. *Biomed. Pharmacother.* 121, 109595. doi:10.1016/j.biopha.2019.109595

Fernandes-Alnemri, T., Wu, J., Yu, J. W., Datta, P., Miller, B., Jankowski, W., et al. (2007). The pyroptosome: A supramolecular assembly of ASC dimers mediating inflammatory cell death via caspase-1 activation. *Cell Death Differ.* 14, 1590–1604. doi:10.1038/sj.cdd.4402194

Giannone, G., Ghisoni, E., Genta, S., Scotto, G., Tuninetti, V., Turinetto, M., et al. (2020). Immuno-metabolism and microenvironment in cancer: Key players for immunotherapy. *Int. J. Mol. Sci.* 21, E4414. doi:10.3390/ijms21124414

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The role of necroptosis in cancer biology and therapy. *Mol. Cancer* 18, 100. doi:10.1186/s12943-019-1029-8

Greathouse, K. L., White, J. R., Vargas, A. J., Bliskovsky, V. V., Beck, J. A., Von Muhlinen, N., et al. (2018). Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol.* 19, 123. doi:10.1186/s13059-018-1501-6

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

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Jr., Wu, Y. L., et al. (2017). Lung cancer: Current therapies and new targeted treatments. *Lancet* 389, 299–311. doi:10.1016/S0140-6736(16)30958-8

Hsu, S. K., Li, C. Y., Lin, I. L., Syue, W. J., Chen, Y. F., Cheng, K. C., et al. (2021). Inflammation-related pyroptosis, a novel programmed cell death pathway, and its crosstalk with immune therapy in cancer treatment. *Theranostics* 11, 8813–8835. doi:10.7150/thno.62521

Janowski, A. M., Colegio, O. R., Hornick, E. E., Mcniff, J. M., Martin, M. D., Badovinac, V. P., et al. (2016). NLRC4 suppresses melanoma tumor progression independently of inflammasome activation. *J. Clin. Invest.*. 126, 3917–3928. doi:10.1172/JCI86953

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558. doi:10.1038/s41591-018-0136-1

Jin, H., and Kim, H. J. (2020). NLRC4, ASC and caspase-1 are inflammasome components that are mediated by P2Y2R activation in breast cancer cells. *Int. J. Mol. Sci.* 21, E3337. doi:10.3390/ijms21093337

Kay, C., Wang, R., Kirkby, M., and Man, S. M. (2020). Molecular mechanisms activating the NAIP-NLRC4 inflammasome: Implications in infectious disease, autoinflammation, and cancer. *Immunol. Rev.* 297, 67–82. doi:10.1111/imr.12906

Kolb, R., Phan, L., Borcherding, N., Liu, Y., Yuan, F., Janowski, A. M., et al. (2016). Obesity-associated NLRC4 inflammasome activation drives breast cancer progression. *Nat. Commun.* 7, 13007. doi:10.1038/ncomms13007

Ma, X., Qiu, Y., Zhu, L., Zhao, Y., Lin, Y., Ma, D., et al. (2020). NOD1 inhibits proliferation and enhances response to chemotherapy via suppressing SRC-MAPK pathway in hepatocellular carcinoma. *J. Mol. Med.* 98, 221–232. doi:10.1007/s00109-019-01868-9

Pinto, A., Morello, S., and Sorrentino, R. (2011). Lung cancer and Toll-like receptors. *Cancer Immunol. Immunother.* 60, 1211–1220. doi:10.1007/s00262-011-1057-8

Rathinam, V. A., and Fitzgerald, K. A. (2016). Inflammasome complexes: Emerging mechanisms and effector functions. *Cell* 165, 792–800. doi:10.1016/j.cell.2016.03.046

Reck, M., Carbone, D. P., Garassino, M., and Barlesi, F. (2021). Targeting KRAS in non-small-cell lung cancer: Recent progress and new approaches. *Ann. Oncol.* 32, 1101–1110. doi:10.1016/j.annonc.2021.06.001

Salomon, B. L., Leclerc, M., Tosello, J., Ronin, E., Piaggio, E., and Cohen, J. L. (2018). Tumor necrosis factor alpha and regulatory T cells in oncoimmunology. *Front. Immunol.* 9, 444. doi:10.3389/fimmu.2018.00444

Sharma, P., Siddiqui, B. A., Anandhan, S., Yadav, S. S., Subudhi, S. K., Gao, J., et al. (2021). The next decade of immune checkpoint therapy. *Cancer Discov.* 11, 838–857. doi:10.1158/2159-8290.CD-20-1680

Shin, W. G., Park, B. J., Lee, S. J., and Kim, J. G. (2018). Infection of human intestinal epithelial cells by invasive bacteria activates NF-κB and increases ICAM-1 expression through NOD1. *Korean J. Intern. Med.* 33, 81–90. doi:10.3904/kjim.2015.409

Snyder, A. G., Hubbard, N. W., Messmer, M. N., Kofman, S. B., Hagan, C. E., Orozco, S. L., et al. (2019). Intratumoral activation of the necroptotic pathway components RIPK1 and RIPK3 potentiates antitumor immunity. *Sci. Immunol.* 4, eaaw2004. doi:10.1126/sciimmunol.aaw2004

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

Thai, A. A., Solomon, B. J., Sequist, L. V., Gainor, J. F., and Heist, R. S. (2021). Lung cancer. *Lancet* 398, 535–554. doi:10.1016/S0140-6736(21)00312-3

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830. doi:10.1016/j.immuni.2018.03.023

Wang, X., Lin, W., Liu, T., Xu, Z., Wang, Z., Cao, Z., et al. (2021). Cross-talk of pyroptosis and tumor immune landscape in lung adenocarcinoma. *Transl. Lung Cancer Res.* 10, 4423–4444. doi:10.21037/tlcr-21-715

Wei, J., Xu, Z., Chen, X., Wang, X., Zeng, S., Qian, L., et al. (2020). Overexpression of GSDMC is a prognostic factor for predicting a poor outcome in lung adenocarcinoma. *Mol. Med. Rep.* 21, 360–370. doi:10.3892/mmr.2019.10837

Wei, Y., Huang, H., Qiu, Z., Li, H., Tan, J., Ren, G., et al. (2017). NLRP1 overexpression is correlated with the tumorigenesis and proliferation of human breast tumor. *Biomed. Res. Int.* 2017, 4938473. doi:10.1155/2017/4938473

Wu, Z., Man, S., Sun, R., Li, Z., Wu, Y., and Zuo, D. (2020). Recent advances and challenges of immune checkpoint inhibitors in immunotherapy of non-small cell lung cancer. *Int. Immunopharmacol.* 85, 106613. doi:10.1016/j.intimp.2020.106613

Xue, Q., Liu, X., Chen, C., Zhang, X., Xie, P., Liu, Y., et al. (2021). Erlotinib protests against LPS-induced parthanatos through inhibiting macrophage surface TLR4 expression. *Cell Death Discov.* 7, 181. doi:10.1038/s41420-021-00571-4

Xue, Y., Enosi Tuipulotu, D., Tan, W. H., Kay, C., and Man, S. M. (2019). Emerging activators and regulators of inflammasomes and pyroptosis. *Trends Immunol.* 40, 1035–1052. doi:10.1016/j.it.2019.09.005

Yatim, N., Jusforgues-Saklani, H., Orozco, S., Schulz, O., Barreira Da Silva, R., Reis E Sousa, C., et al. (2015). RIPK1 and NF-κB signaling in dying cells determines cross-priming of $CD8^+$ T cells. *Science* 350, 328–334. doi:10.1126/science.aad0395

Zeng, Q. Z., Yang, F., Li, C. G., Xu, L. H., He, X. H., Mai, F. Y., et al. (2019). Paclitaxel enhances the innate immunity by promoting NLRP3 inflammasome activation in macrophages. *Front. Immunol.* 10, 72. doi:10.3389/fimmu.2019.00072

Zhang, C. C., Li, C. G., Wang, Y. F., Xu, L. H., He, X. H., Zeng, Q. Z., et al. (2019). Chemotherapeutic paclitaxel and cisplatin differentially induce pyroptosis in A549 lung cancer cells via caspase-3/GSDME activation. *Apoptosis* 24, 312–325. doi:10.1007/s10495-019-01515-1

Zhang, J., Zhang, Q., Lou, Y., Fu, Q., Chen, Q., Wei, T., et al. (2018). Hypoxia-inducible factor-1α/interleukin-1β signaling enhances hepatoma epithelial-mesenchymal transition through macrophages in a hypoxic-inflammatory microenvironment. *Hepatology* 67, 1872–1889. doi:10.1002/hep.29681

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 12 December 2022
DOI 10.3389/fgene.2022.992995

OPEN ACCESS

EDITED BY
Prakashbabu Phanithi,
University of Hyderabad, India

REVIEWED BY
Zheng He,
Qilu Hospital of Shandong University, China
Emil Bulatov,
Kazan Federal University, Russia

*CORRESPONDENCE
MeiHua Li,
limeihua2000@sina.com
YeYu Zhao,
zyp19850922@126.com

†These authors have contributed equally to this work

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

RECEIVED 13 July 2022
ACCEPTED 02 December 2022
PUBLISHED 12 December 2022

CITATION
Zhu H, Wan Q, Tan J, Ouyang H, Pan X, Li M and Zhao Y (2022), A novel prognostic signature of cuproptosis-related genes and the prognostic value of FDX1 in gliomas.
Front. Genet. 13:992995.
doi: 10.3389/fgene.2022.992995

A novel prognostic signature of cuproptosis-related genes and the prognostic value of FDX1 in gliomas

HuaXin Zhu[1,2†], Qinsi Wan[3†], Jiacong Tan[1,2], Hengyang Ouyang[4], Xinyi Pan[4], MeiHua Li[1*] and YeYu Zhao[1*]

[1]Department of Neurosurgery, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China, [2]Medical Innovation Center, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China, [3]Department of Gastroenterology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China, [4]Huankui Academy, Nanchang University, Nanchang, Jiangxi, China

Background: Gliomas are the most common malignant tumors of the central nervous system, with extremely bad prognoses. Cuproptosis is a novel form of regulated cell death. The impact of cuproptosis-related genes on glioma development has not been reported.

Methods: The TCGA, GTEx, and CGGA databases were used to retrieve transcriptomic expression data. We employed Cox's regressions to determine the associations between clinical factors and cuproptosis-related gene expression. Overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) were evaluated using the Kaplan-Meier method. We also used the least absolute shrinkage and selection operator (LASSO) regression technique.

Results: The expression levels of all 10 CRGs varied considerably between glioma tumors and healthy tissues. In glioma patients, the levels of CDKN2A, FDX1, DLD, DLAT, LIAS, LIPT1, and PDHA1 were significantly associated with the OS, disease-specific survival, and progression-free interval. We used LASSO Cox's regression to create a prognostic model; the risk score was (0.882340) *FDX1 expression + (0.141089) *DLD expression + (−0.333875) *LIAS expression + (0.356469) *LIPT1 expression + (−0.123851) *PDHA1 expression. A high-risk score/signature was associated with poor OS (hazard ratio = 3.50, 95% confidence interval 2, −4.55, log-rank $p < 0.001$). Cox's regression revealed that the FDX1 level independently predicted prognosis; FDX1 may control immune cell infiltration of the tumor microenvironment.

Conclusion: The CRG signature may be prognostic in glioma patients, and the FDX1 level may independently predict glioma prognosis. These data may afford new insights into treatment.

KEYWORDS

gliomas, cuproptosis, FDX1, prognosis, prognostic signature

1 Introduction

Malignant tumors of the central nervous system (CNS) have one of the poorest prognoses of all cancers; life is shortened by about 20 years (Rouse et al., 2010; Reifenberger et al., 2017). More than 70% of all malignant CNS tumors are gliomas; such tumors are the most common CNS tumors (Gilbert et al., 2014). Half of all newly diagnosed gliomas are very malignant glioblastomas; the median patient survival is approximately 12 months (Gramatzki et al., 2016; Quinones and Le, 2018). Over the last decade, isocitrate dehydrogenase (IDH) mutations, chromosome 1p/19q deletions, MGMT and TERT promoter methylations, and histone mutations have all served as glioma biomarkers; these markers play key roles in glioma classification and treatment decisions (Chen et al., 2017; Brito et al., 2019). Despite great advances in surgery, radiotherapy, chemotherapy, and targeted therapy, almost all malignant gliomas recur, associated with poor prognoses. Better prognostic models are urgently required.

Although heavy metal ions are crucial micronutrients, ion levels that are too low or excessive may trigger controlled cell death (Wang et al., 2022a) *via* activation of various subprograms. For example, ferroptosis is an iron-dependent form of uncontrolled, lipid peroxidation-induced, oxidative cell death (Dixon et al., 2012; Liang et al., 2019). Recently, Tsvetkov et al. (2022) found that intracellular copper (Cu) triggered a unique form of controlled cell death termed "cuproptosis," which differed from apoptosis, necrosis, autophagy, and ferroptosis. The lipoylated acylated components of the tricarboxylic acid (TCA) cycle bind directly to copper, imparting protein stress and eventual cell death (Hatori et al., 2016). Recent studies have shown that cancer patients exhibit much higher serum and tumor tissue copper levels than healthy people (Blockhuys et al., 2017; Ge et al., 2022). Dysregulation of copper homeostasis may be cytotoxic; changes in intracellular copper levels may influence cancer development and spread (Ishida et al., 2013; Babak and Ahn, 2021). However, any relationship between cuproptosis and glioma progression remains unknown. This is our topic here; glioma cuproptosis-related gene (CRG) expression is of clinical and potential prognostic utility.

2 Methods

2.1 Data acquisition

The glioma RNA-seq data of TCGA and the corresponding normal tissue data of GTEx derived *via* uniform toil processing were downloaded from UCSCXENA (https://xenabrowser.net/datapages/). The Human Protein Atlas database (https://www.proteinatlas.org/) was used to examine protein expression in normal and glioma tissues. We also downloaded RNA-seq data and clinical information (DataSet ID: mRNAseq_693) of glioma samples from the CGGA database (http://www.cgga.org.cn/); we used this information to externally validate the survival analyses.

2.2 Differential expression analysis

We integrated the TCGA and GTEx databases and then sought differences in CRG expression levels between glioma and normal tissue samples. We drew receiver operating characteristic (ROC) curves to assess the predictive accuracies of CRG levels. The relationships between CRG levels and clinicopathological features were explored using the TCGA database and validated employing the CGGA database. The R package DESeq2 (ver. 1.26.0) was used to distinguish the expression levels of mRNAs associated with low and high risks of progression; we identified differentially expressed genes (DEGs) using the thresholds of $\log_2$ (FC) > 2.0 and an adjusted *p*-value <0.05. The DEGs were displayed using volcano plots and heat maps.

2.3 Prognostic signatures of cuproptosis-related genes

Useful features were selected using the least absolute shrinkage and selection operator (LASSO) regression algorithm; 10-fold cross-validation followed and the R package glmnet was used to analyze the data. First, multifactorial Cox's regression was performed, followed by iteration using a step function. The optimal model was selected. We used the log-rank test and Cox's proportional hazards regression to derive *p*-values and hazard ratios with 95% confidence intervals (CIs); then we drew Kaplan-Meier curves. We used the decision curve analysis R package ggDCA to construct the prognostic signature of CRG levels.

2.4 Functional and pathway enrichment analyses

We employed the R clusterProfiler package to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses and displayed the findings using the R packages dplyr and ggplot2. We uploaded the DEGs to Metascape (an online tool for gene function/annotation analysis; https://metascape.org/).

2.5 Construction of a prognostic model and external validation

We sought prognostic factors among clinical variables, performed univariate and multivariate Cox's regressions, and

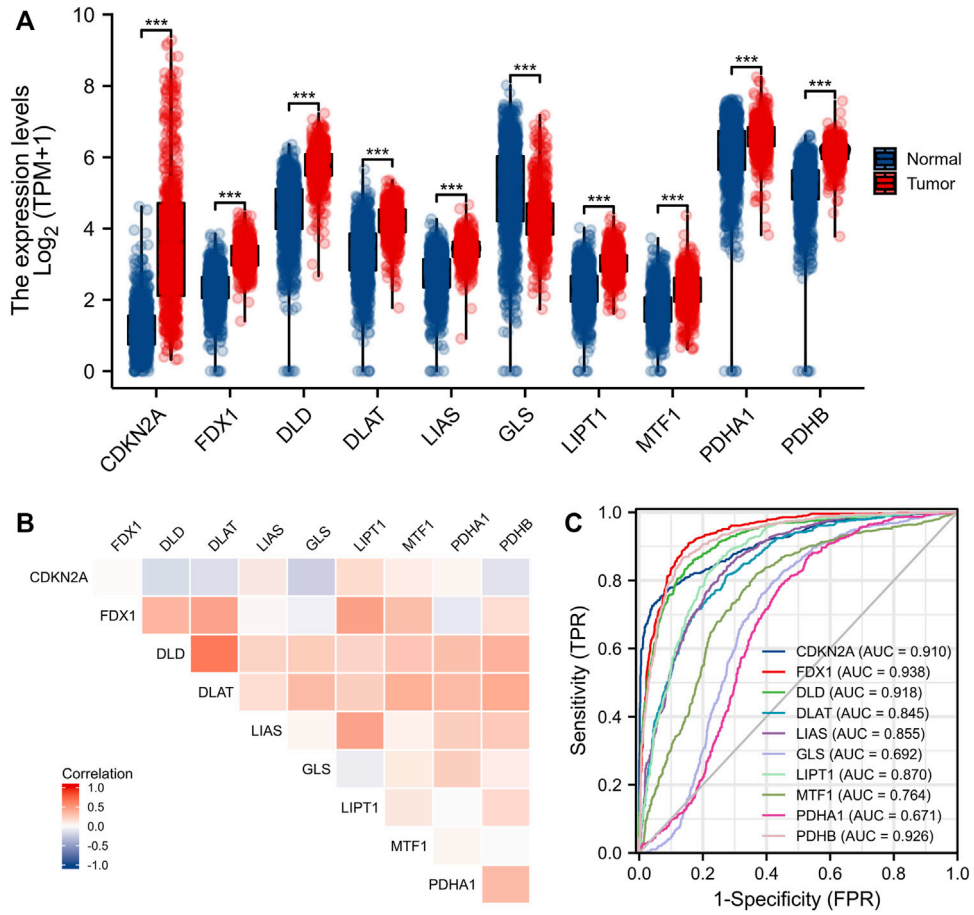


FIGURE 1
Expression of CRGs in gliomas. **(A)** Expression of 10 CRGs in gliomas and normal tissues. **(B)** Correlation between CRGs expression. **(C)** Diagnostic value of CRGs in gliomas. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. CRGs: cuproptosis-related genes.

derived an optimal prognostic model. The R package rms was used to create a nomogram that predicted prognosis. Harrell concordance index (C-index) calibration plots were drawn to evaluate the reliabilities and accuracies of the prognostic models. Kaplan-Meier curves were drawn to compare the differences between paired groups in terms of overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI).

2.6 Evaluation of immune cell infiltration

The Tumor Immune Estimation Resource (Timer) method of the R package immunedeconv was used to seek correlations between immune cell activities, and the constructed models and the gene expression levels *per se*. The R package ggstatsplot was employed to derive correlations between gene expression levels and immune system scores; the R package pheatmap was used to identify multi-gene correlations. We used the single-sample Gene Set Enrichment Analysis (ssGSEA) technique of the R GSVA package to measure the levels of 24 immune cell types associated with glioma infiltration. The R package is based on the TCGA database. We used Spearman analyses to derive correlations between quantitative variables that were not normally distributed.

2.7 Statistical analysis

Paired samples were evaluated using the Wilcoxon signed-rank test; unpaired samples were assessed employing the Wilcoxon rank-sum test. We used the Kruskal–Wallis test, the Wilcoxon signed-rank test, and logistic regression to explore the relationships between clinical characteristics and CRG expression levels. Either the chi-square or the Fisher exact test was used to investigate relationships between CRG expression levels and the clinical characteristics. R software ver. 4.0.3 was employed for all statistical analyses (R Foundation for Statistical Computing, Vienna, Austria). A p-value <0.05 was considered significant.

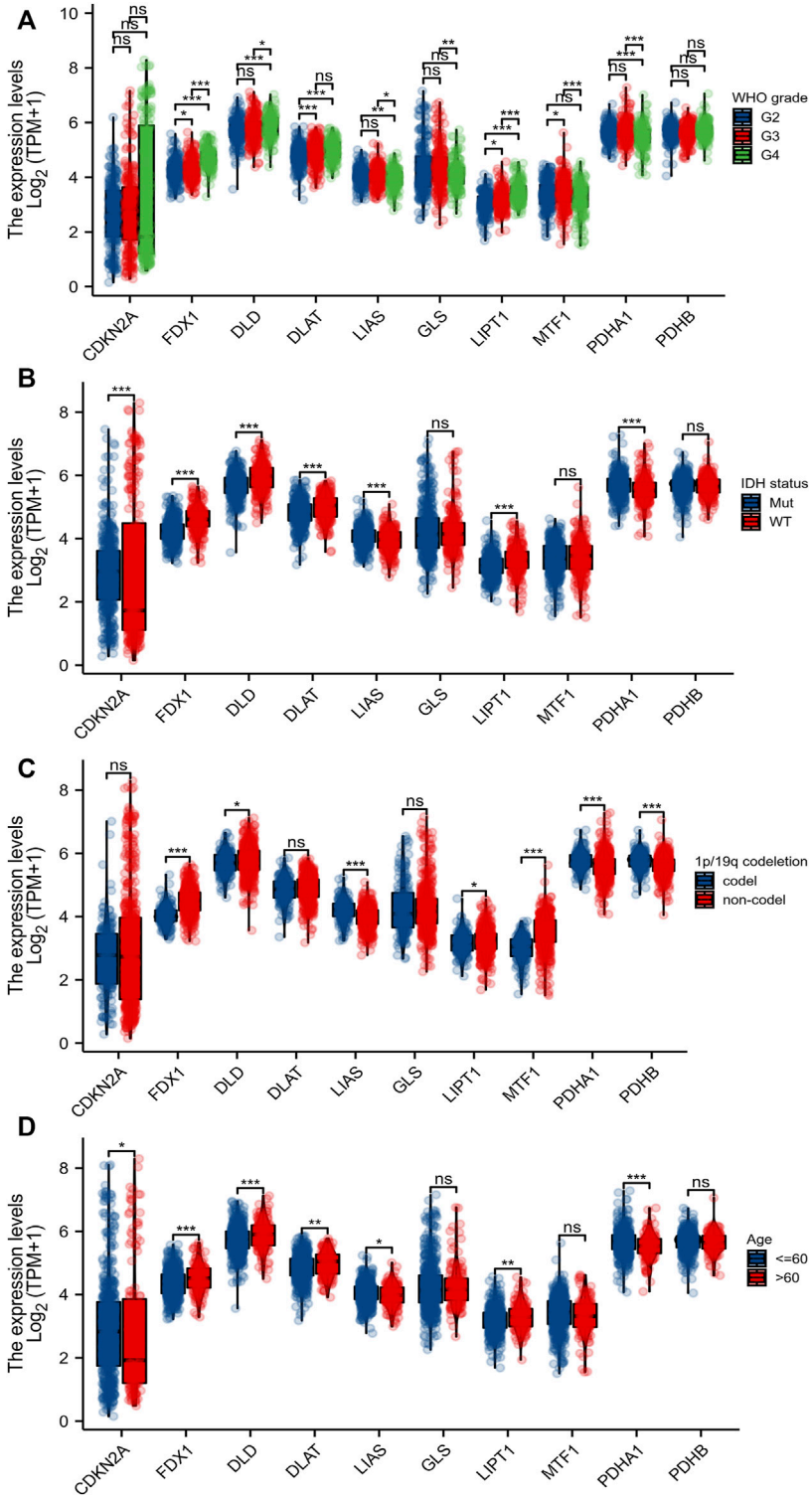


FIGURE 2
Relationship between CRGs and clinicopathological features. The association of 10 CRGs expression with WHO grade **(A)**, IDH status **(B)**, 1p/19q codeletion **(C)**, and age **(D)** in gliomas from TCGA database. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. CRGs: cuproptosis-related genes.

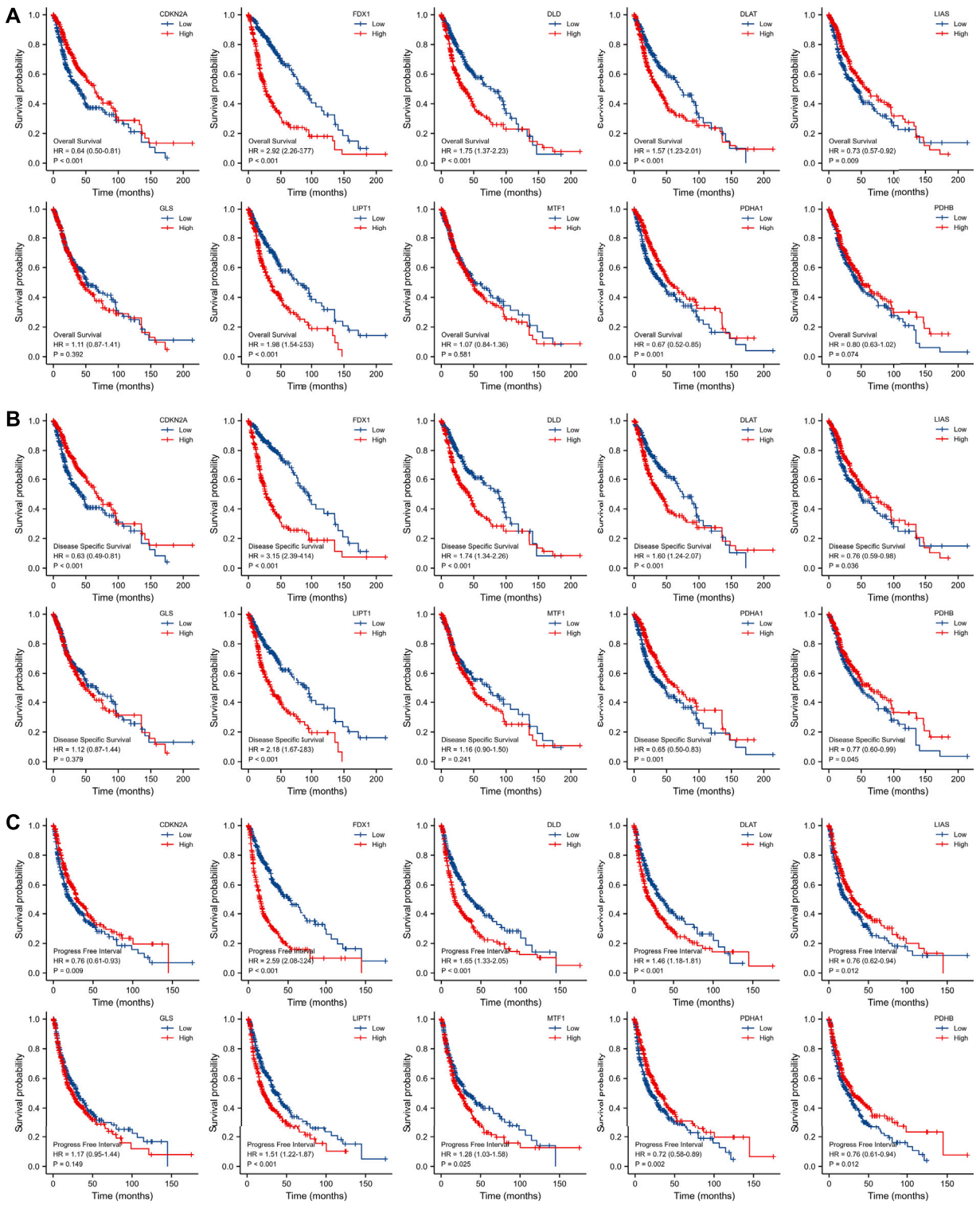


FIGURE 3
Survival analysis of OS **(A)**, DSS **(B)** and PFI **(C)** of 10 CRGs in gliomas patients from TCGA database. OS: overall survival, DSS: disease specific survival, PFI: progress free interval, CRGs: cuproptosis-related genes.

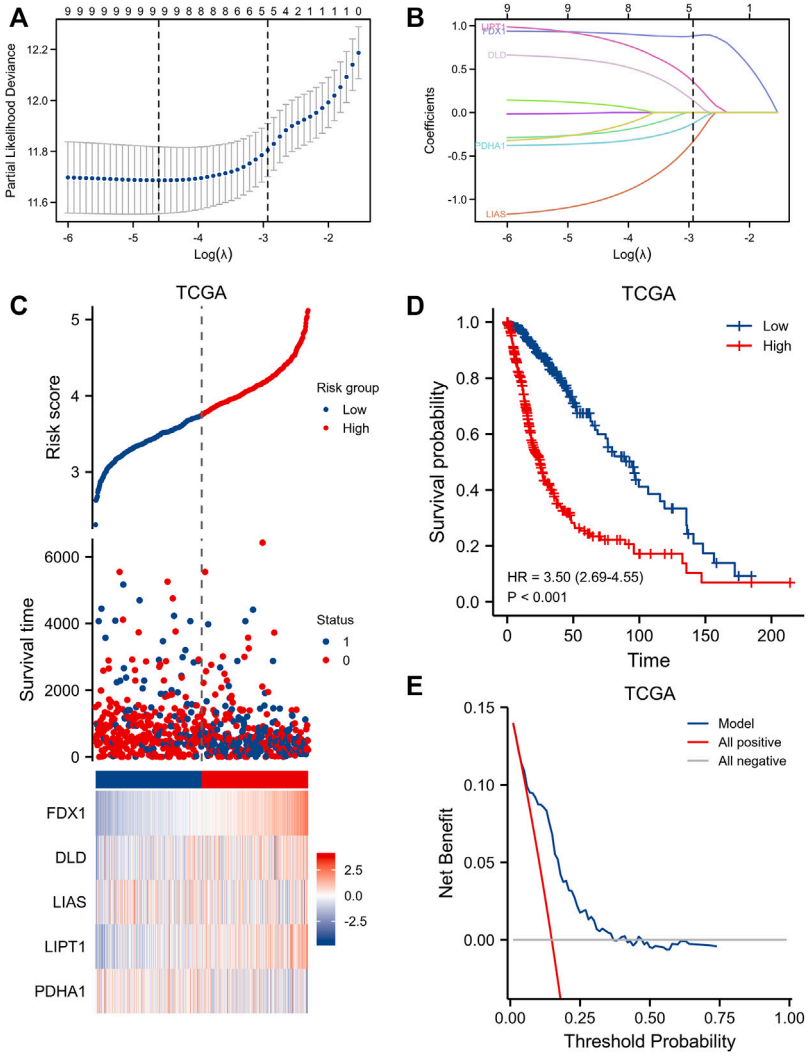


FIGURE 4
Clinical relevance of CRGs in the gliomas patients of TCGA. **(A,B)** LASSO regression analysis and partial likelihood deviance of the CRGs. **(C)** distribution of risk score, survival status and the expression of prognostic gliomas, **(D)** Kaplan–Meier plot of the CRGs signature and overall survival, **(E)** Decision curve analysis of CRGs signature for predicting survival status.

3 Results

3.1 CRG expression levels in glioma patients

Eleven genes (CDKN2A, DLD, DLAT, FDX1, GLS, LIAS, LIPT1, MTF1, PDHA1, and PDHB) are closely associated with cuproptosis. We found that only the expression level of GLS was significantly decreased in gliomas (compared to normal tissues); the levels of all other CRGs increased significantly (Figure 1A). We sought correlations between the levels of different CRGs (Figure 1B). We performed receiver operating characteristic curve (ROC) analyses to assess whether various CRG levels aided glioma detection; the areas under the curves (AUCs) were >0.9 for CDKN2A, FDX1, DLD, and PDHB (Figure 1C). Levels of FDX1, LIAS, LIPT1, DLD, DLAT, MTF1, and PDHA1 were associated with the World Health Organization (WHO) clinical grade (Figure 2A); levels of CDKN2A, DLAT, LIAS, LIPT1 FDX1, DLD, and PDHA1 with were associated with IDH status (Figure 2B); levels of FDX1, LIPT1, MTF1, PDHA1, DLD, LIAS, and PDHB were associated with 1p19q co-deletion (Figure 2C); and levels of CDKN2A, DLD, DLAT, LIAS, LIPT1, FDX1 and PDHA1 levels were associated with age (Figure 2D).

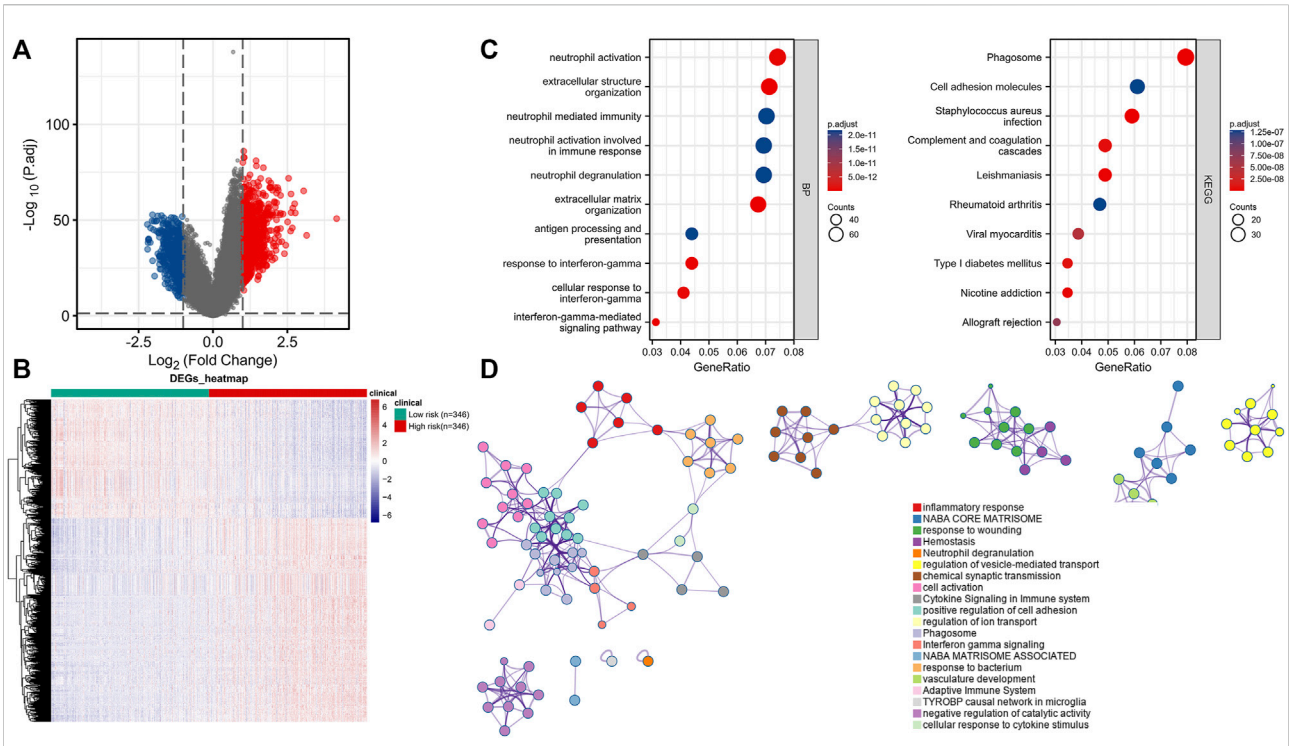


FIGURE 5
Functional enrichment analysis for CRGs signature in gliomas patients. **(A)** Volcano Plot of differentially expressed genes (DEGs) screened based on CRGs signature. **(B)** Gene set enrichment analysis of Gene Ontology biological process (GO-BP). **(C)** Gene set enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG). **(D)** Network and bar chart of enrichment analysis results drawn by Metascape.

3.2 The survival analyses of CRGs in gliomas patients

Patients of the TCGA database expressing high levels of FDX1 (hazard ratio [HR] = 2.92 [95% confidence interval, CI] 2.26–3.77, log-rank $p < 0.001$); DLD (HR = 1.75 [1.37–2.23], log-rank $p < 0.001$); DLAT (HR = 1.57 [1.23–2.01], log-rank $p < 0.001$); and LIPT1 [HR = 1.98 (1.54–2.53), log-rank $p < 0.001$]; and low levels of CDKN2A [HR = 0.64 (0.50–0.81), log-rank $p < 0.001$]; LIAS [HR = 0.73 (0.57–0.92), log-rank $p = 0.009$]; and PDHA1 [HR = 0.67 (0.52–0.85), log-rank $p = 0.001$] experienced poorer OS than others (Figure 3A). Poor DSS was associated with high-level expression of FDX1 [HR = 3.15 (2.39–4.14), log-rank $p < 0.001$]; DLD [HR = 1.74 (1.34–2.26), log-rank $p < 0.001$]; DLAT [HR = 1.60 (1.24–2.07), log-rank $p < 0.001$]; and LIPT1 [HR = 2.18 (1.67–2.83), log-rank $p < 0.001$]; and low-level expression of CDKN2A [HR = 0.63 (0.49–0.81), log-rank $p < 0.001$]; LIAS [HR = 0.76 (0.59–0.98), log-rank $p = 0.036$]; PDHA1 [HR = 0.65 (0.50–0.83), log-rank $p = 0.001$]; and PDHB [HR = 0.77 (0.60–0.99), log-rank $p = 0.045$] (Figure 3B). PFI was affected by the levels of CDKN2A [HR = 0.76 (0.61–0.93), log-rank $p = 0.009$]; FDX1 [HR = 2.59 (2.08–3.24), log-rank $p < 0.001$]; DLD [HR = 1.65 (1.33–2.05), log-rank $p < 0.001$]; DLAT [HR = 1.46 (1.18–1.81), log-rank $p < 0.001$]; LIAS [HR = 0.76 (0.62–0.94), log-rank $p = 0.012$]; LIPT1 [HR = 0.51 (1.22–1.87), log-rank $p < 0.001$]; MTF1 [HR = 1.28 (1.03–1.58), log-rank $p < 0.001$]; PDHA1 [HR = 0.72 (0.58–0.89), log-rank $p = 0.002$]; and PDHB [HR = 0.76 (0.61–0.94), log-rank $p = 0.012$] (Figure 3C).

3.3 Construction of a CRG-derived prognostic gene signature

LASSO Cox's regression generated a five-gene signature (Figures 4A, B): Risk Score = (0.882340) *FDX1 expression + (0.141089) *DLD expression + (−0.333875) *LIAS expression + (0.356469) *LIPT1 expression + (−0.123851) *PDHA1 expression. We found a significant correlation between a high risk score and poor OS [HR = 3.50 (2.69–4.55), log-rank $p < 0.001$] (Figures 4C,D). Decision curve analysis in terms of OS prediction showed that the signature performed well [C-index 0.758 (0.743–0.772)] (Figure 4E). The CGGA database was utilized for external validation; we retrieved all data on glioma patients. As for TCGA patients, CGGA individuals with higher risk scores evidenced significantly shorter OS [HR = 1.68 (1.38–2.05), log-rank $p < 0.001$] (Supplementary Figure S1). The clinical information of TCGA and CGGA datasets were provided in Supplementary Tables S1, S2.

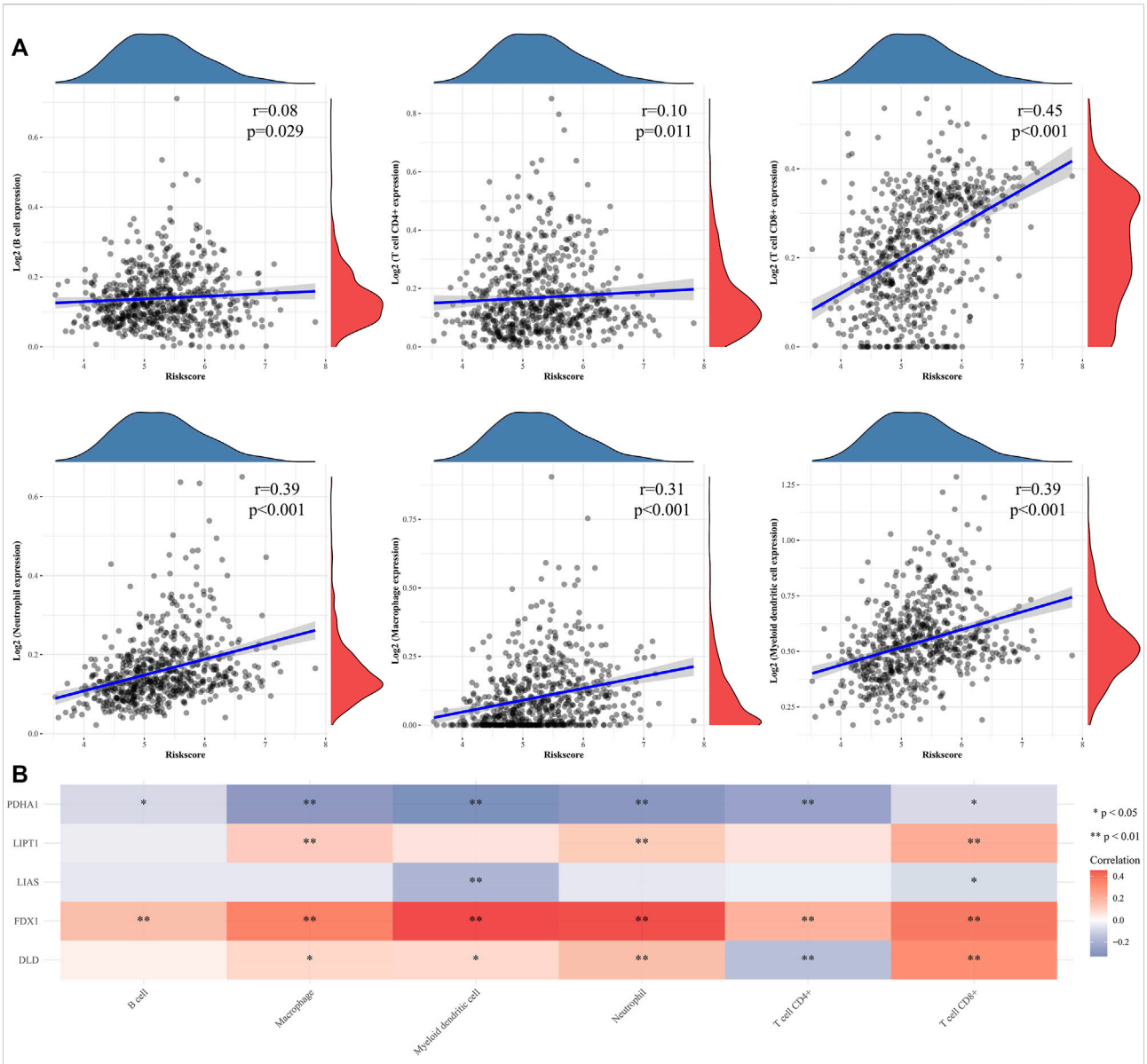


FIGURE 6
Immune infiltration analyse of CRGs signature in gliomas patients. **(A)** The correlations between CRGs signature risk score and immune cell was analysed with Spearman. **(B)** A heatmap of the correlation between CRGs signature genes and immune score. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

3.4 Functional enrichment and immune cell infiltration associated with high and low signature risk scores

We measured the differentially expressed (DE) mRNA levels of gliomas with high and low CRG signature risk scores to explore whether cuproptosis might trigger glioma progression. A volcano plot (Figure 5A) and a heat-map (Figure 5B) identified 1213 DE mRNAs, the expression of 767 of which were increased and 446 decreased (Supplementary Table S3). We sought gene enrichments in terms of biological processes (BP), cellular components (CC), molecular function (MF), and of the criteria of the KEGG (Supplementary Table S4). Neutrophil activation, extracellular structure organization, neutrophil mediated immunity, neutrophil activation involved in immune response, neutrophil degranulation and extracellular matrix organization were all significantly activated on GO-BP enrichment analysis. Phagosome, cell adhesion molecules, *staphylococcus aureus* infection, complement and coagulation cascades and leishmaniasis were enriched on KEGG. The results are shown in Figure 5C. Enrichment results drawn by Metascape were similar (Figure 5D). Immune cell infiltration analysis revealed that the CRG signatures

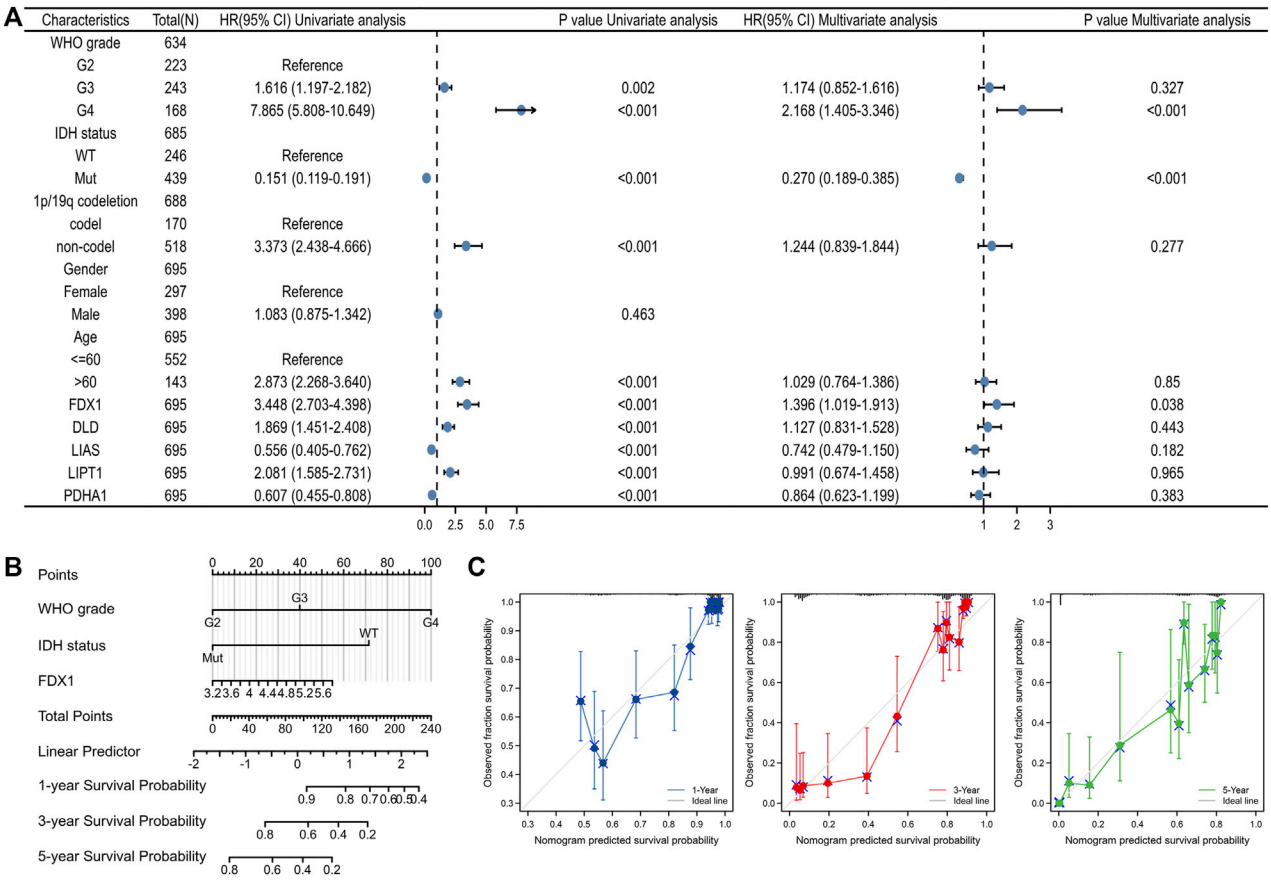


FIGURE 7
Construction of a predictive nomogram. **(A)** The univariate and multivariate Cox regression analysis for clinical parameters of gliomas and five CRGs. **(B)** Nomogram to predict the 1, 3, and 5 years overall survival rate of gliomas patients. **(C)** The calibration curve of the nomogram.

were associated with all six immune cell types, but principally $CD8^+$ T cells (Figure 6A). Figure 6B shows the correlations between the expression levels of CRG signature genes and immune cell activities.

3.5 The nomogram development and construction for gliomas patients

Univariate Cox's regression revealed that a high WHO grade; wild-type IDH status; 1p19q non-co-deletion; age >60 years; high-level expression of FDX1, DLD, and LIPT1; and low-level expression of LIAS and PDHA1 were all associated with poor OS. Multivariate Cox's HR analysis revealed that the WHO grade, IDH status, and the FDX1 mRNA expression level independently predicted OS (Figure 7A). We created an OS nomogram to integrate the FDX1 level with other prognostic factors (the WHO and IDH data) (Figure 7B). We drew a calibration curve to evaluate the contribution of the FDX1 level to nomogram performance; the C-index was 0.834 (0.823–0.846) (Figure 7C). The association between the FDX1 level and the clinicopathological parameters of TCGA patients with gliomas are shown in Table 1. Logistic regression analysis of the FDX1 data revealed strong associations with clinical characteristics including the WHO grade [OR = 3.634 (2.579–5.161), $p < 0.001$]; IDH status [OR = 5.945 (4.20–8.509), $p < 0.001$]; 1p/19q co-deletion [OR = 10.231 (6.492–16.801), $p < 0.001$]; and age (OR = 2.016 (1.384–2.961), $p < 0.001$) (Table 2). Thus, increased FDX1 expression was associated with poor prognosis.

3.6 Validation of the effects of FDX1 levels as revealed by the CGGA and HPA databases

Compared to the low-level FDX1 group of the CGGA database, the high-level FDX1 group evidenced poorer OS [HR = 1.47 (1.12–1.80), log-rank $p < 0.001$) (Figure 8A). The FDX1 level was significantly correlated with the WHO grade ($p < 0.001$) (Figure 8B), IDH status ($p < 0.01$) (Figure 8C), and 1p19q co-deletion status ($p < 0.001$) (Figure 8D). Immunohistochemically, FDX1 staining was positive in gliomas but negative in normal tissues (Figure 8E).

TABLE 1 Association of FDX1 expression and clinicopathological parameters in patients with gliomas.

Characteristic	Low expression of FDX1	High expression of FDX1	*p*
n	348	348	
WHO grade, *n* (%)			<0.001
G2	155 (24.4%)	69 (10.9%)	
G3	130 (20.5%)	113 (17.8%)	
G4	27 (4.3%)	141 (22.2%)	
IDH status, *n* (%)			<0.001
WT	59 (8.6%)	187 (27.3%)	
Mut	287 (41.8%)	153 (22.3%)	
1p/19q codeletion, *n* (%)			<0.001
Codel	148 (21.5%)	23 (3.3%)	
non-codel	200 (29%)	318 (46.2%)	
Gender, *n* (%)			0.818
Female	147 (21.1%)	151 (21.7%)	
Male	201 (28.9%)	197 (28.3%)	
Age, *n* (%)			<0.001
≤60	296 (42.5%)	257 (36.9%)	
>60	52 (7.5%)	91 (13.1%)	
OS event, *n* (%)			<0.001
Alive	262 (37.6%)	162 (23.3%)	
Dead	86 (12.4%)	186 (26.7%)	
DSS event, *n* (%)			<0.001
Alive	266 (39.4%)	165 (24.4%)	
Dead	74 (11%)	170 (25.2%)	
PFI event, *n* (%)			<0.001
Alive	224 (32.2%)	126 (18.1%)	
Dead	124 (17.8%)	222 (31.9%)	
Age, median (IQR)	42 (33, 54)	50 (36, 62)	<0.001

TABLE 2 Logistic regression analysis of FDX1 expression.

Characteristics	Total (N)	Odds ratio (OR)	*p*-Value
WHO grade (G3&G4 vs. G2)	635	3.634 (2.579–5.161)	<0.001
IDH status (WT vs. Mut)	686	5.945 (4.20.-8.509)	<0.001
1p/19q codeletion (non-codel vs. codel)	689	10.231 (6.492–16.801)	<0.001
Age (>60 vs. ≤ 60)	696	2.016 (1.384–2.961)	<0.001
Gender (Male vs. Female)	696	0.954 (0.706–1.288)	0.759

3.7 Correlation between FDX1 expression levels and immune cell infiltration

The correlations between FDX1 expression levels and the numbers of immune cells of 24 different types were assessed. Figure 9A shows the results. The numbers of macrophages, eosinophils, and Th2 cells were significantly positively correlated with the FDX1 expression levels whereas the numbers of pDCs, NK CD56 bright cells, and Treg cells were significantly negatively correlated (Figure 9).

4 Discussion

Gliomas are the most common malignant tumors of the central nervous system, progress rapidly, and are associated

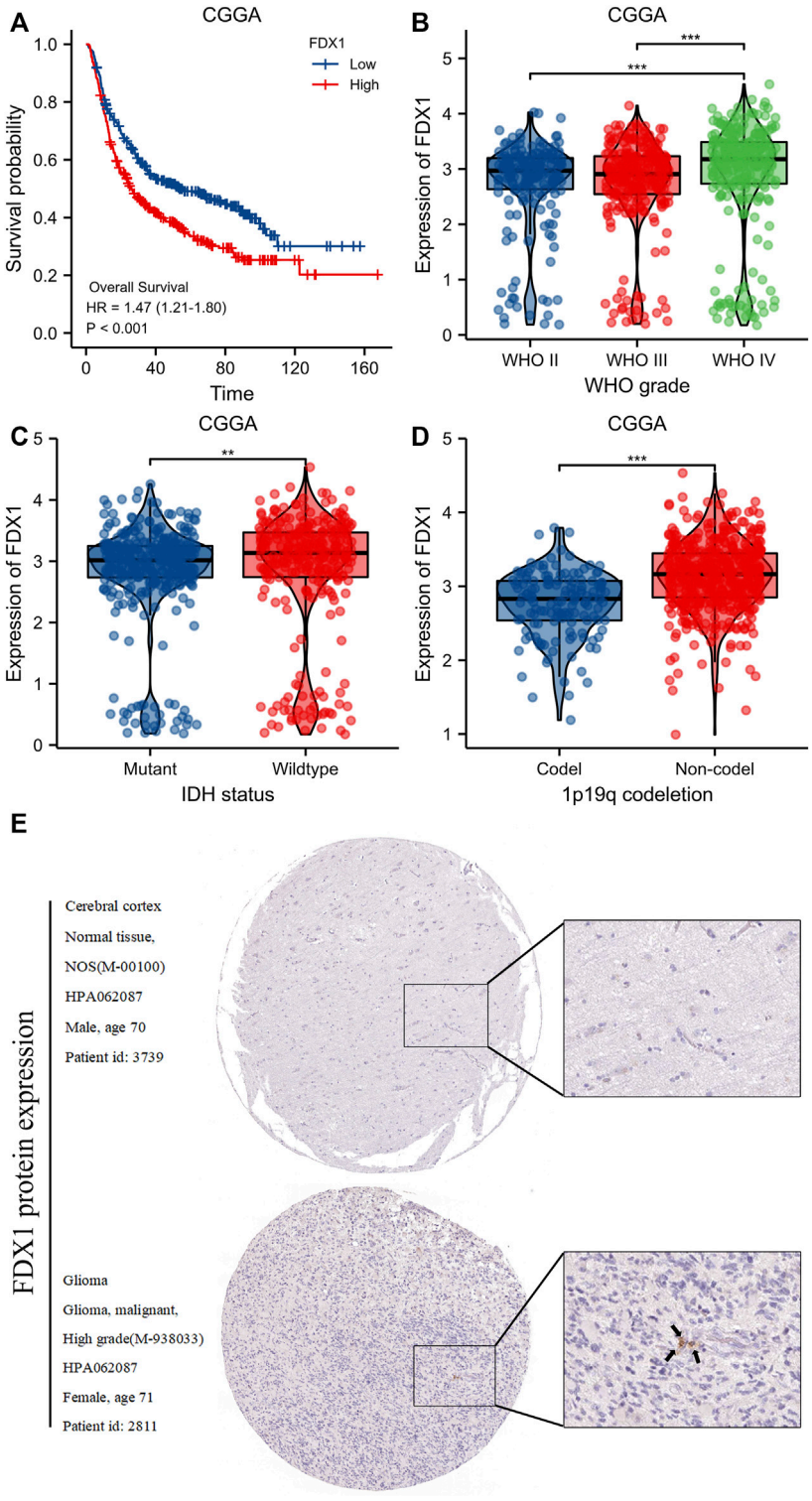


FIGURE 8
Validation of FDX1 expression from CGGA and HPA database. **(A)**Kaplan-Meier survival curve analysis of overall survival (OS) showed that high FDX1 expression correlated to poor prognosis of gliomas patients from CGGA database. **(B–D)** The association of FDX1 expression with WHO grade **(B)**, IDH status **(C)** and 1p/19q codeletion **(D)** in gliomas from CGGA database. **(E)** FDX1 protein expression in gliomas tissues determined using HPA. ***$p < 0.001$, ns, no statistical difference.

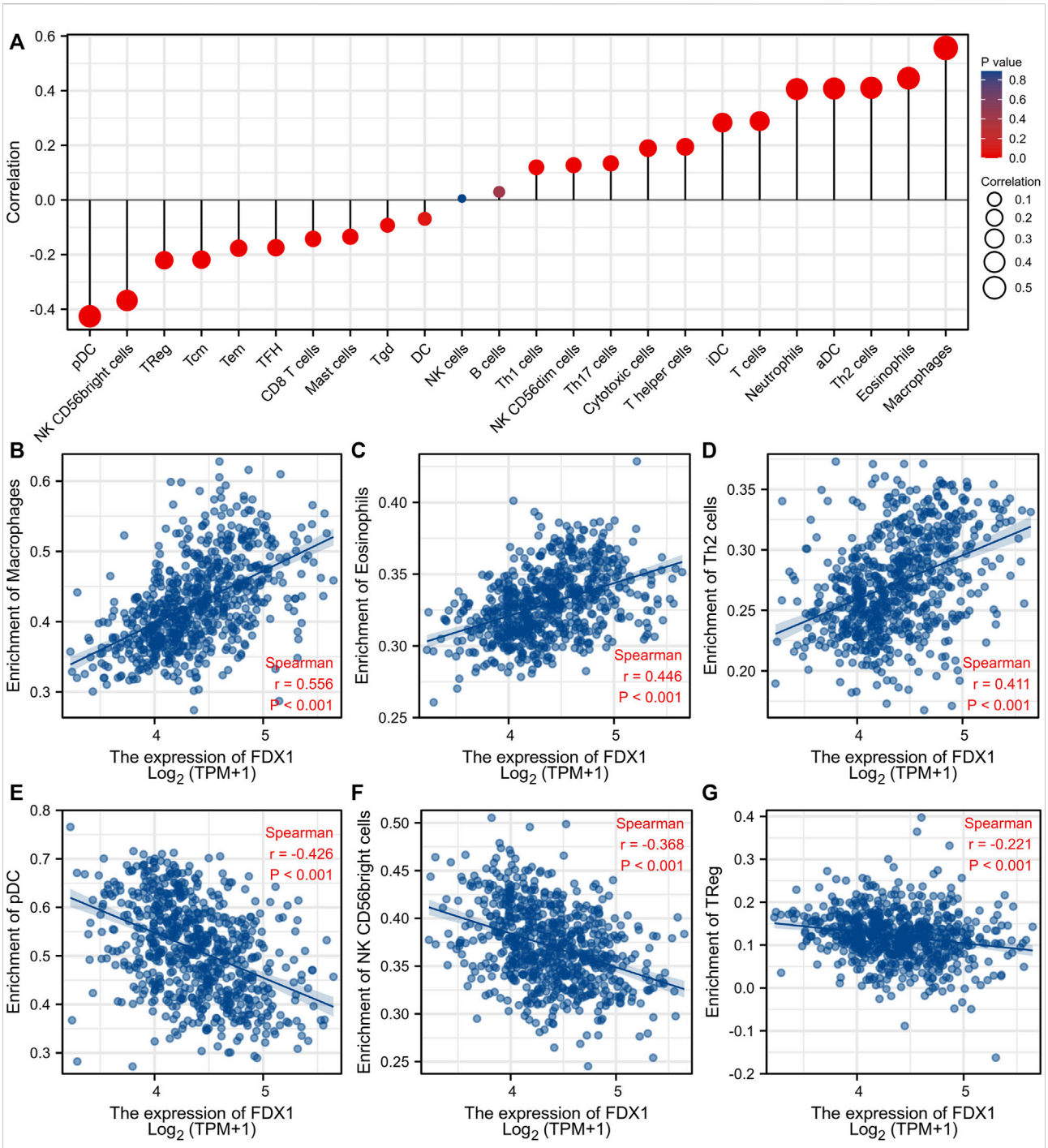


FIGURE 9
Association analysis of FDX1 expression and immune infiltration in gliomas patients. **(A)** The association between FDX1 expression and 24 tumor-infiltrating lymphocytes. **(B–D)** The positive correlation of FDX1 expression with immune infiltration level of Macrophages **(B)**, Eosinophils **(C)**, and Th2 cells**(D)**. **(E–G)** The negative correlation of FDX1 expression with immune infiltration level of pDC cells **(E)**, NK CD56 bright cells **(F)**, and TReg **(G)**.

with poor prognoses and high mortality, imposing significant costs on families and society (Lah et al., 2020; Guo et al., 2021). Early diagnosis, and timely and effective treatment, are critical. Improvements are essential (Muller et al., 2020). Copper binds to the acylated fatty acids of the TCA cycle, triggering toxic protein stress; this is cuproptosis (Tsvetkov

et al., 2022). However, any role for cuproptosis in glioma progression remains poorly understood. We explored the glioma expression levels of 10 CRGs and their prognostic significances. We present a novel, cuproptosis-based, prognostic scoring system. We found that FDX1 might be a valuable therapeutic target.

The expression levels of all 10 CRGs differed between glioma and normal tissues. In addition to the levels of GLS, MTF1, and PDHB, those of the other seven CRGs significantly affected the OS, DSS, and PFI of glioma patients; cuproptosis may play a major role in disease progression. The signature risk scores showed that cuproptosis was highly correlated with tumor immune cell infiltration; cuproptosis may modulate such infiltration.

The "cuproptosis" concept is rather new; few studies on the effects of cuproptosis genes on glioma development have been published. Wang et al. (2022b) subjected CRG data to cluster analysis, and studied the possible roles of genomic mutations and immune cell infiltration in great depth; that study was very valuable. Chen et al. (2022) found that CRG levels were highly correlated with glioma aggressiveness and immune infiltration, and screened potential therapeutic agents, opening a new path toward glioma treatment. Ye et al. (2022) developed a valuable copper death prognostic model for WHO grade 2/3 glioma patients; drug-sensitivity was important. Ouyang et al. (2022) presented models based on lncRNA and CRG levels; cuproptosis-associated glutaminase gene expression reflected the prognosis of glioma patients. The prognostic model of Yan et al. (2022) used the levels of cuproptosis-associated lncRNAs to identify appropriate immune checkpoint blockade (ICB) therapies for LGG patients. We employed LASSO Cox's regression analysis when constructing our prognostic model; the risk score was based on the levels of five CRGs (FDX1, DLD, LIAS, LIPT1, and PDHA1). Our model well-predicted the OS of glioma patients, not only those of the TCGA database but also those of the CGGA database. Our model yields risk scores (numbers), facilitating generalization. In addition, we found that FDX1 status was important in terms of glioma progression.

FDX1 encodes an iron-sulfur protein that reduces Cu^{2+} to Cu^{1+} and regulates protein acylation during cuproptosis (Dorsam and Fahrer, 2016; Weger et al., 2018). DLD is a key enzyme of many dehydrogenase and glycine decarboxylase complexes that maintain cellular and mitochondrial homeostasis (Duarte et al., 2021). LIPT1 and LIAS are members of the lipoic acid metabolic pathway that regulates mitochondrial energy metabolism (Tort et al., 2014; Cronan, 2020). PDHA 1 (a key component of the pyruvate dehydrogenase complex) controls pyruvate entry into the TCA cycle (Echeverri et al., 2021). Univariate Cox's regression confirmed that the FDX1, DLD, LIAS, LIPT1, and PDHA1 levels were all associated with OS, but only the FDX1 level independently predicted survival.

The soluble iron-sulfur protein ferredoxin 1 (FDX1) plays major roles in many metabolic pathways, for example, by transferring electrons during mitochondrial redox reactions (Sawyer and Winkler, 2017). Wang et al. (2021) found that FDX1 was involved in mitochondrial steroid metabolism, and played a role in the development of polycystic ovary syndrome. FDX1 facilitates copper-dependent cell death; an analysis of FDX1 action shed light on how cells respond to proteotoxic stress (Tsvetkov et al., 2019). FDX1 increases ATP production and plays key roles in glucose metabolism, fatty acid oxidation, and amino acid metabolism in lung adenocarcinoma cells (Zhang et al., 2021).

Any prognostic utility of glioma FDX1 status remains unclear. We found that FDX1 status is strongly correlated with glioma prognosis, the WHO glioma grade, IDH status, and 1p19q co-deletion status. The FDX1 level independently predicted glioma prognosis. Immune cells play important roles in tumor creation and development (Chasov et al., 2020). We explored the connections between FDX1 levels and immune cell populations. Higher FDX1 expression strongly enhanced tumor infiltration of macrophages, eosinophils, and Th2 cells but reduced infiltration of pDCs, NK CD56 bright cells, and Treg cells. Thus, FDX1 may regulate immune infiltration into the glioma microenvironment.

Our work had several limitations. Most data were mined from public databases and validated only *in vitro*; some of our findings require further validation. We interrogated only a few databases. Prospective studies with larger sample sizes are needed.

5 Conclusion

We explored the roles played by CRGs in glioma development. Our prognostic risk score based on the expression signature is strongly predictive of OS. The FDX1 level was independently prognostic; targeting of FDX1 expression might serve as a cuproptosis-specific therapeutic strategy for glioma prevention 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 authors.

Author contributions

Conceptualization, HZ and HO; methodology, HZ and JT; software, QW; validation, HO, XP, and HZ; formal analysis, XP; investigation, HZ; resources, QW; data curation, HZ; writing–original draft preparation, HZ; writing–review and editing, ML; visualization, JT and QW; supervision, YZ; project administration, ML; funding acquisition, YZ and ML. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the natural science foundation of Jiangxi Province (No. 20212BAB206029), the National Natural Science Foundation of China (NSFC) (No. 81860225, 82260248), the key research and development plan of Jiangxi Province (No. 20203BBG73060), and young talents research and cultivation fundation of the First Affiliated Hospital of Nanchang University (No. YFYPY202038).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.992995/full#supplementary-material

References

Babak, M. V., and Ahn, D. (2021). Modulation of intracellular copper levels as the mechanism of action of anticancer copper complexes: Clinical relevance. *Biomedicines* 9, 852. doi:10.3390/biomedicines9080852

Blockhuys, S., Celauro, E., Hildesjo, C., Feizi, A., Stal, O., Fierro-Gonzalez, J. C., et al. (2017). Defining the human copper proteome and analysis of its expression variation in cancers. *Metallomics* 9, 112–123. doi:10.1039/c6mt00202a

Brito, C., Azevedo, A., Esteves, S., Marques, A. R., Martins, C., Costa, I., et al. (2019). Clinical insights gained by refining the 2016 who classification of diffuse gliomas with: Egfr amplification, tert mutations, pten deletion and mgmt methylation. *Bmc Cancer* 19, 968. doi:10.1186/s12885-019-6177-0

Chasov, V., Mirgayazova, R., Zmievskaya, E., Khadiullina, R., Valiullina, A., Stephenson, C. J., et al. (2020). Key players in the mutant p53 team: Small molecules, gene editing, immunotherapy. *Front. Oncol.* 10, 1460. doi:10.3389/fonc.2020.01460

Chen, B., Zhou, X., Yang, L., Zhou, H., Meng, M., Zhang, L., et al. (2022). A cuproptosis activation scoring model predicts neoplasm-immunity interactions and personalized treatments in glioma. *Comput. Biol. Med.* 148, 105924. doi:10.1016/j.compbiomed.2022.105924

Chen, R., Smith-Cohn, M., Cohen, A. L., and Colman, H. (2017). Glioma subclassifications and their clinical significance. *Neurotherapeutics* 14, 284–297. doi:10.1007/s13311-017-0519-x

Cronan, J. E. (2020). Progress in the enzymology of the mitochondrial diseases of lipoic acid requiring enzymes. *Front. Genet.* 11, 510. doi:10.3389/fgene.2020.00510

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

Dorsam, B., and Fahrer, J. (2016). The disulfide compound alpha-lipoic acid and its derivatives: A novel class of anticancer agents targeting mitochondria. *Cancer Lett.* 371, 12–19. doi:10.1016/j.canlet.2015.11.019

Duarte, I. F., Caio, J., Moedas, M. F., Rodrigues, L. A., Leandro, A. P., Rivera, I. A., et al. (2021). Dihydrolipoamide dehydrogenase, pyruvate oxidation, and acetylation-dependent mechanisms intersecting drug iatrogenesis. *Cell. Mol. Life Sci.* 78, 7451–7468. doi:10.1007/s00018-021-03996-3

Echeverri, R. N., Mohan, V., Wu, J., Scott, S., Kreamer, M., Benej, M., et al. (2021). Dynamic regulation of mitochondrial pyruvate metabolism is necessary for orthotopic pancreatic tumor growth. *Cancer Metab.* 9, 39. doi:10.1186/s40170-021-00275-4

Ge, E. J., Bush, A. I., Casini, A., Cobine, P. A., Cross, J. R., DeNicola, G. M., et al. (2022). Connecting copper and cancer: From transition metal signalling to metalloplasia. *Nat. Rev. Cancer* 22, 102–113. doi:10.1038/s41568-021-00417-2

Gilbert, M. R., Dignam, J. J., Armstrong, T. S., Wefel, J. S., Blumenthal, D. T., Vogelbaum, M. A., et al. (2014). A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N. Engl. J. Med.* 370, 699–708. doi:10.1056/NEJMoa1308573

Gramatzki, D., Dehler, S., Rushing, E. J., Zaugg, K., Hofer, S., Yonekawa, Y., et al. (2016). Glioblastoma in the canton of zurich, Switzerland revisited: 2005 to 2009. *Cancer* 122, 2206–2215. doi:10.1002/cncr.30023

Guo, X., Wang, T., Huang, G., Li, R., Da, C. C., Li, H., et al. (2021). Rediscovering potential molecular targets for glioma therapy through the analysis of the cell of origin, microenvironment and metabolism. *Curr. Cancer Drug Targets* 21, 558–574. doi:10.2174/1568009621666210504091722

Hatori, Y., Yan, Y., Schmidt, K., Furukawa, E., Hasan, N. M., Yang, N., et al. (2016). Neuronal differentiation is associated with a redox-regulated increase of copper flow to the secretory pathway. *Nat. Commun.* 7, 10640. doi:10.1038/ncomms10640

Ishida, S., Andreux, P., Poitry-Yamate, C., Auwerx, J., and Hanahan, D. (2013). Bioavailable copper modulates oxidative phosphorylation and growth of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19507–19512. doi:10.1073/pnas.1318431110

Lah, T. T., Novak, M., and Breznik, B. (2020). Brain malignancies: Glioblastoma and brain metastases. *Semin. Cancer Biol.* 60, 262–273. doi:10.1016/j.semcancer.2019.10.010

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

Muller, B. J., Kulasinghe, A., Chua, B., Day, B. W., and Punyadeera, C. (2020). Circulating biomarkers in patients with glioblastoma. *Br. J. Cancer* 122, 295–305. doi:10.1038/s41416-019-0603-6

Ouyang, Z., Zhang, H., Lin, W., Su, J., and Wang, X. (2022). Bioinformatic profiling identifies the glutaminase to be a potential novel cuproptosis-related biomarker for glioma. *Front. Cell Dev. Biol.* 10, 982439. doi:10.3389/fcell.2022.982439

Quinones, A., and Le, A. (2018). The multifaceted metabolism of glioblastoma. *Adv. Exp. Med. Biol.* 1063, 59–72. doi:10.1007/978-3-319-77736-8_4

Reifenberger, G., Wirsching, H. G., Knobbe-Thomsen, C. B., and Weller, M. (2017). Advances in the molecular genetics of gliomas - implications for classification and therapy. *Nat. Rev. Clin. Oncol.* 14, 434–452. doi:10.1038/nrclinonc.2016.204

Rouse, C., Gittleman, H., Ostrom, Q. T., Kruchko, C., and Barnholtz-Sloan, J. S. (2010). Years of potential life lost for brain and cns tumors relative to other cancers in adults in the United States. *Neuro. Oncol.* 18, 70–77. doi:10.1093/neuonc/nov249

Sawyer, A., and Winkler, M. (2017). Evolution of chlamydomonas reinhardtii ferredoxins and their interactions with [fefe]-hydrogenases. *Photosynth. Res.* 134, 307–316. doi:10.1007/s11120-017-0409-4

Tort, F., Ferrer-Cortes, X., Thio, M., Navarro-Sastre, A., Matalonga, L., Quintana, E., et al. (2014). Mutations in the lipoyltransferase lipt1 gene cause a fatal disease associated with a specific lipoylation defect of the 2-ketoacid dehydrogenase complexes. *Hum. Mol. Genet.* 23, 1907–1915. doi:10.1093/hmg/ddt585

Tsvetkov, P., Coy, S., Petrova, B., Dreishpoon, M., Verma, A., Abdusamad, M., et al. (2022). Copper induces cell death by targeting lipoylated tca cycle proteins. *Science* 375, 1254–1261. doi:10.1126/science.abf0529

Tsvetkov, P., Detappe, A., Cai, K., Keys, H. R., Brune, Z., Ying, W., et al. (2019). Author correction: Mitochondrial metabolism promotes adaptation to proteotoxic stress. *Nat. Chem. Biol.* 15, 757. doi:10.1038/s41589-019-0315-5

Wang, W., Lu, Z., Wang, M., Liu, Z., Wu, B., Yang, C., et al. (2022). The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma. *Front. Immunol.* 13, 998236. doi:10.3389/fimmu.2022.998236

Wang, Y., Zhang, L., and Zhou, F. (2022). Cuproptosis: A new form of programmed cell death. *Cell. Mol. Immunol.* 19, 867–868. doi:10.1038/s41423-022-00866-1

Wang, Z., Dong, H., Yang, L., Yi, P., Wang, Q., and Huang, D. (2021). The role of fdx1 in granulosa cell of polycystic ovary syndrome (pcos). *BMC Endocr. Disord.* 21, 119. doi:10.1186/s12902-021-00775-w

Weger, M., Weger, B. D., Gorling, B., Poschet, G., Yildiz, M., Hell, R., et al. (2018). Glucocorticoid deficiency causes transcriptional and post-transcriptional reprogramming of glutamine metabolism. *Ebiomedicine* 36, 376–389. doi:10.1016/j.ebiom.2018.09.024

Yan, X., Wang, N., Dong, J., Wang, F., Zhang, J., Hu, X., et al. (2022). A cuproptosis-related lncrnas signature for prognosis, chemotherapy, and immune checkpoint blockade therapy of low-grade glioma. *Front. Mol. Biosci.* 9, 966843. doi:10.3389/fmolb.2022.966843

Ye, Z., Zhang, S., Cai, J., Ye, L., Gao, L., Wang, Y., et al. (2022). Development and validation of cuproptosis-associated prognostic signatures in who 2/3 glioma. *Front. Oncol.* 12, 967159. doi:10.3389/fonc.2022.967159

Zhang, Z., Ma, Y., Guo, X., Du, Y., Zhu, Q., Wang, X., et al. (2021). Fdx1 can impact the prognosis and mediate the metabolism of lung adenocarcinoma. *Front. Pharmacol.* 12, 749134. doi:10.3389/fphar.2021.749134

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 05 January 2023
DOI 10.3389/fgene.2022.988909

OPEN ACCESS

EDITED BY
Ziqi Jia,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Feng Jiang,
Fudan University, China
Yutao Wang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Xusheng Wang,
wangxsh27@mail.sysu.edu.cn
Xiaodong Chen,
cxd234@163.com

†These authors have contributed equally to this work

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

RECEIVED 07 July 2022
ACCEPTED 21 December 2022
PUBLISHED 05 January 2023

CITATION
He J, Huang W, Li X, Wang J, Nie Y, Li G, Wang X, Cao H, Chen X and Wang X (2023), A new ferroptosis-related genetic mutation risk model predicts the prognosis of skin cutaneous melanoma.
Front. Genet. 13:988909.
doi: 10.3389/fgene.2022.988909

A new ferroptosis-related genetic mutation risk model predicts the prognosis of skin cutaneous melanoma

Jia He[1,2†], Wenting Huang[1†], Xinxin Li[1], Jingru Wang[2], Yaxing Nie[3], Guiqiang Li[2], Xiaoxiang Wang[2], Huili Cao[1], Xiaodong Chen[2*] and Xusheng Wang[1*]

[1]School of Pharmaceutical Sciences (Shenzhen), Sun Yat-Sen University, Guangzhou, China, [2]Department of Burn Surgery, The First People's Hospital of Foshan, Foshan, China, [3]CAS Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai, China

Background: Ferroptosis is an iron-dependent cell death mode and closely linked to various cancers, including skin cutaneous melanoma (SKCM). Although attempts have been made to construct ferroptosis-related gene (FRG) signatures for predicting the prognosis of SKCM, the prognostic impact of ferroptosis-related genetic mutations in SKCM remains lacking. This study aims to develop a prediction model to explain the relationship between ferroptosis-related genetic mutations and clinical outcomes of SKCM patients and to explore the potential value of ferroptosis in SKCM treatment.

Methods: FRGs which significantly correlated with the prognosis of SKCM were firstly screened based on their single-nucleotide variant (SNV) status by univariate Cox regression analysis. Subsequently, the least absolute shrinkage and selection operator (LASSO) and Cox regressions were performed to construct a new ferroptosis-related genetic mutation risk (FerrGR) model for predicting the prognosis of SKCM. We then illustrate the survival and receiver operating characteristic (ROC) curves to evaluate the predictive power of the FerrGR model. Moreover, independent prognostic factors, genomic and clinical characteristics, immunotherapy, immune infiltration, and sensitive drugs were compared between high—and low—FerrGR groups.

Results: The FerrGR model was developed with a good performance on survival and ROC analysis. It was a robust independent prognostic indicator and followed a nomogram constructed to predict prognostic outcomes for SKCM patients. Besides, FerrGR combined with tumor mutational burden (TMB) or MSI (microsatellite instability) was considered as a combined biomarker for immunotherapy response. The high FerrGR group patients were associated with an inhibitory immune microenvironment. Furthermore, potential drugs target to high FerrGR samples were predicted.

Abbreviations: AUC, areas under the curve; CNVs, copy number variations; FerrGR, ferroptosis-related genetic mutation risk; GSVA, gene set variation analysis; ICIs, immune checkpoint inhibitors; LASSO, least absolute shrinkage and selection operator; MMR, mismatch repair deficiency; MSI, microsatellite instability; OS, overall survival; PD-L1, programmed cell-death receptor 1 ligand; ROC, receiver operating characteristic; SNV, single-nucleotide variant; TMB, tumor mutational burden; TIME, tumor immunosuppressive environment; TME, tumor microenvironment.

Conclusion: The FerrGR model is valuable to predict prognosis and immunotherapy in SKCM patients. It offers a novel therapeutic option for SKCM.

KEYWORDS

skin cutaneous melanoma, ferroptosis, genetic mutation, single nucleotide variant, prognosis, tumor immunity

Introduction

Skin cutaneous melanoma (SKCM), which is the most aggressive skin cancer, takes up for more than 75% mortality rate of skin-related cancers. Although patients with localized and regional cutaneous melanoma have a 5-year relative survival of 98% and 64% respectively. Once metastasized through the body, the 5-years survival rate falls to 23% (Rebecca et al., 2020). Generally, surgical resection is considered the first choice for patients with early-stage disease. Moreover, some advanced melanoma is insensitive to radiotherapy and chemotherapy as for its high aggressiveness (Ping et al., 2022). Therefore, several therapeutic agents including kinase inhibitors and immune checkpoint inhibitors (ICIs) were developed (Leonardi et al., 2018; Leonardi et al., 2020). Nowadays, Immunotherapy and targeted therapy have shown promising results in clinical trials and become the backbone of systemic treatment (Pelster and Amaria, 2019; Ribas et al., 2019). Despite the rapid development of these therapeutic approaches, limitations emerged since SKCM is heterogeneous cancer. Patients with the same stage and treatments may have a different prognosis and treatment response (Ackerman et al., 2014; Hassel et al., 2016; Simeone et al., 2017). Therefore, it is crucial to identify a prognostic predictive biomarker to inform clinical prognosis and treatment response.

Ferroptosis which was discovered in recent years is a novel form of programmed cell death and is characterized by a large amount of iron accumulation and lipid peroxidation (Li et al., 2020). It differs from other forms of cell death such as apoptosis, pyroptosis, necroptosis, and autophagy in morphology, biochemistry, and genetics (Gao et al., 2016). The main mechanism of ferroptosis is phospholipid peroxidation, which relies on the transition metal iron, reactive oxygen species (ROS), and phospholipids. In addition, nutrients, intra/intercellular signaling, and environmental stresses contribute to ferroptosis by regulating cellular metabolism and ROS levels (Jiang et al., 2021). Increasing evidence has indicated that ferroptosis was closely associated with the tumorigenesis and progression of cancers (Li et al., 2020). Many tumor suppressors show susceptibility to ferroptosis. Hence, regulating the antitumor activity of these tumor suppressors could be explored as an anticancer therapy (Jiang et al., 2021). Furthermore, Erastin, Sulfasalazine, Sorafenib, and other small molecule ferroptosis inducers used in the clinical treatment of cancer showed promising outcomes of anti-tumor effect (Liang et al., 2019; Xu G et al., 2021). Recent studies investigated that the differentiation status of melanoma cells was correlated with the susceptibility to ferroptosis. Ferroptosis inducers could decrease the number of dedifferentiated melanoma cells and prevent their immunosuppressive actions (Rebecca et al., 2020; Ping et al., 2022) (Gagliardi et al., 2020; Talty and Bosenberg, 2022). Apart from ferroptosis inducers, some miRNAs and genes associated with ferroptosis are involved in the development of SKCM. A previous study reported that miR-137 acts as a negative regulator of ferroptosis by directly targeting glutamine transporter SLC1A5 in melanoma cells (Luo et al., 2018). Additionally, miR-9 suppressed Erastin- and RSL3-induced ferroptosis by targeting glutamic-oxaloacetic transaminase GOT1 in melanoma cells (Zhang et al., 2018). Inhibiting mitochondrial complex I induced autophagosome formation, mitophagy, a cytosolic ROS increase and ultimately lead to necroptosis/ferroptosis in melanoma cells (Basit et al., 2017). Besides, evidence suggested that GPX4, VDAC2/3, NEDD4, AKRs, and SLC7A11 are involved in the resistance to ferroptosis in melanoma (Talty and Bosenberg, 2022). Ferroptosis has been a new hope for SKCM therapeutics. Nevertheless, the roles of ferroptosis-related genes in prognostic prediction and tumor microenvironment (TME) remain unclear.

Recent studies have consistently revealed biomarkers such as tumor mutation burden (TMB), neoantigen load (NAL), programmed cell-death receptor 1 ligand (PD-L1) expression, and lactate dehydrogenase (LDH) to predict therapeutic benefit in SKCM (Jiang J et al., 2020). Unfortunately, there still existed several limitations to their clinical application, including the undefined cut-off value, intra/intratumor heterogeneity, unsatisfactory predictive power, and relatively high cost (Jiang J et al., 2020; Bai et al., 2020). This highlights more effective and clinically actionable biomarkers are required to be identified.

Genetic mutations are heritable changes in the nucleotide sequence of DNA that resulted from both inherited and environmental factors. The mutator phenotype hypothesis suggests that the capacity to divide, invade, and metastasize of cancer cells results from genetic mutations that maintain the stability of genes in normal cells. Mutations in genetic stability genes initiate mutations by causing mutations in other genes that govern genetic stability. Next, some of the resulting mutated cells expand and achieve clonal dominance (Loeb et al., 2003). Notably, targeted therapy based on the specific genetic background has made a great progress. For example, BRAF mutations were discovered in nearly half of metastatic SKCM. Patients with BRAF mutations showed improved progression-free survival by treatment with two BRAF inhibitors vemurafenib and dabrafenib (Hauschild et al., 2012; Jin et al., 2019). However, the mutations in cancers affect drug sensitivity and drive drug resistance. Therefore, the outcomes of targeted therapy are largely dependent upon the mutation profile of tumors in patients.

In this study, we performed comprehensive analysis utilizing data downloaded from TCGA and GEO databases, along with FRGs identified in previous studies to determine potential ferroptosis-related prognostic genes of SKCM in accordance with SNV mutational status. Subsequently, we developed and evaluated a ferroptosis-related genetic mutation risk (FerrGR) model for predicting prognosis and assessing multiple roles of ferroptosis-related genetic mutations in the TME of SKCM. In addition, an integrated prognostic nomogram was established by combining the risk model and clinicopathological features to ameliorate the prognostic assessment of SKCM patients. We also characterized the distinctive immune landscape and genetic and epigenetic signature associated

with the FerrGR model. Besides, potential drugs were predicted in the light of the FerrGR score. Overall, the FerrGR model might provide an effective prediction tool and help guide clinical decisions on therapy for SKCM.

Materials and methods

Data collection

All datasets used in this study were publicly available. RNA-seq transcriptome data, somatic mutations, SNVs, copy number variations (CNVs), methylation, clinical characteristics, and survival information were downloaded from The Cancer Genome Atlas (TCGA) database (http://www.cgga.org.cn/) and the Gene-Expression Omnibus (GEO) database (GSE91061). TMB data of Pan-Cancer was received from the GDC database. 63 immune checkpoint marker genes were obtained from the literature (Hu et al., 2020). A total of 299 ferroptosis-related genes were obtained from the FerrDb database (http://www.zhounan.org/ferrdb/) and a literature search (Liang et al., 2020; Zhuo et al., 2020; Hong et al., 2021; Tang et al., 2021). Among them, the TCGA-SKCM cohort contains 286 ferroptosis-related genes which were selected for further analysis (Supplementary Table S1).

Identification of the prognostic FRGs

SNV mutations of FRGs in the TCGA-SKCM cohort were counted by the "maftools" R package. The heatmap of FRGs was drawn by the "ComplexHeatmap" R package. Tumor patients in TCGA-SKCM cohorts were classified as the mutation and the wild-type based on the presence or absence of SNV mutations in FRGs. Thereafter, the prognostic value of FRGs was determined by univariate Cox regression analysis using the R package "survival" where $p < 0.1$ was considered statistically significant. "Forestplot" R package was used to plot the forest map of prognostic FRGs. "ggpubr" R package was used to plot the sample proportion pie chart of prognostic ferroptosis related-genes mutation/wild-type samples. "Survminer" and "Survival" R packages were used to plot the survival curve of mutation/wild-type patients.

Establishment of a ferroptosis-related genetic mutation risk (FerrGR) model

TCGA-SKCM mutational cohorts were divided into training and validation cohorts with the ratio of the training: validation = 7:3. The prognostic risk characteristics were assessed using the "glmnet" and "survival" R package based on the LASSO method in the training cohort. The FerrGR score was calculated according to the SNV mutational status (SNV mutation was equivalent to 1, while wild-type status was 0) of the key FRGs and the corresponding regression coefficient. The computational formula was as follows:

$$FerrGR\ score = \sum LASSO\ regression\ coefficient \times SNV\ mutational\ value\ of\ key\ gene\ (0\ or\ 1)$$

The "forestplotdrug-sensitive" R package was used to draw the forest map of the key genes included in the model and their coefficients in the model.

Validation of the FerrGR model

The patients in the training cohort were divided into High—and Low—FerrGR groups according to the optimal threshold obtained by the "surminer" R package. Then, the SNV type and frequency of key genes in the training cohort were counted with the package "maftools". In addition, the heatmap of FRGs was drawn by the package "ComplexHeatmap", while the survival curves of the two subgroups were created by the package "survminer" and "survival". Subsequently, the package "pROC" was used to calculate and draw the ROC curve of FerrGR, TMB, and MSI in the validation cohort.

Construction of the predictive nomogram based on the FerrGR model

The clinical characteristics including sample type, tumor stage, gender, the value of Clark's level, BMI, TNM-staging, and TCGA molecular typing in different subgroups of the FerrGR model were calculated by the R package "ggpubr". The FerrGR scores and the clinical characteristics were inputted into univariate and multivariate Cox analysis to validate whether the FerrGR score was an independent risk factor for SKCM. After that, a nomogram was constructed by "regplot" and "rms" packages for predicting the progression of SKCM patients.

Multiomics characteristics analysis

The different landscape of SNVs, amplification and deletion of FRGs between high—and low—FerrGR groups was identified by the chi-square test. In addition, the differential expression and genomic methylation of FRGs between subgroups were analyzed with the "limma" package.

FerrGR model for immunotherapy

Data of SNV mutations from the dataset GSE90161 was scored by the FerrGR model, and then the patients were divided into high - and low - FerrGR groups based on the median value. After grouped, the heatmap of immune checkpoint genes and the survival curve of different groups were performed by "complexheatmap", "survminer" and "survival" R packages, respectively. Besides, immunotherapeutic response PD (progressive disease)/SD (stable disease) and CR (complete response)/PR (partial response) was assessed by "ggstatsplot" package.

Survival analysis on basis of the FerrGR model combined with TMB or MSI

We got the MSI status of patients from the TCGA-SKCM dataset by the "PreMSIm" package. Subsequently, patients were grouped into high—and low—MSI groups based on the median value. In combination with the FerrGR model, the patients were split into three groups: the first group's scores in MSI and FerrGR model were both high, the second group's scores were both low, and the third group's scores were single high. The prognostic survival curve of these

three groups was then analyzed and plotted by package "survminer" and "survival". The prognostic survival analysis by the FerrGR model combined with TMB was done in the same way.

Tumor microenvironment analysis

Here, we used package "estimate" to calculate stromal and immune scores for predicting the level of infiltrating stromal and immune cells, and the tumor purity was also inferred in TCGA-SKCM cohort patients. The differences in the clinical characteristics, FerrGR score, stromal and immune scores, and tumor purity between high—and low—FerrGR groups were then statistically analyzed, where the *t*-test and chi-square test were used for continuous and categorical variables respectively. Subsequently, the infiltration of immune cells in the TCGA-SKCM cohort was estimated using the cibersort algorithm. The R package "ggpubr" was performed to count the differential expression of immune checkpoints between the two subgroups. The gene set variation analysis (GSVA) was conducted to calculate the scores of enrichments in immune pathways by complying with the "GSEABase" and "GSVA" R package.

Potential sensitive drug prediction

The drug sensitive information and corresponding expression were downloaded from the PRISM Repurposing 19Q4 dataset (https://depmap.org/portal/download/all/) and Cancer Therapeutics Response Portal v2.1 (https://ocg.cancer.gov/programs/ctd2/data-portal/#). In addition, SNVs and samples' information of CCLE cell lines was obtained (https://depmap.org/portal/download/all/). Next, SKCM cell lines were divided into high—and low—FerrGR groups by calculated FerrGR score. Drug sensitivity of cell lines was qualified as an AUC value, and a lower AUC value suggested higher drug sensitivity. We then used the package "corrr" for exploring the correlations between the FerrGR score and AUC/IC50.

Statistical analysis

The R software (version: 4. 0. 2) was utilized to conduct all the statistical analyses in this article. All *p* values of statistical data were based on two-sided statistical tests, and data with $p < 0.05$ was considered to be statistically significant (except for the univariate Cox proportional hazards regression model, where $p < 0.1$ was considered to be statistically significant).

Results

Identification of prognosis-related FRGs in the TCGA-SKCM cohort

The flowchart of the present research is shown in Supplementary Figure S1. A total of 463 SKCM patients from the TCGA-SKCM cohort were included in this study. The detailed clinical characteristics of these patients were summarized (Supplementary Table S2). We firstly identified the SNV landscape of 286 FRGs in SKCM patients. SNVs were discovered in most FRGs and the 30 top-ranked FRGs were present in the heatmap of Figure 1A. Of note, the top 2 highest ranked FGRs were NRAS and CFTR, which had 29 and 19 percent SNV mutation regions respectively. Subsequently, 24 prognosis-related genes were screened from all 286 FRGs by the univariate Cox regression analysis of overall survival (OS) ($p < 0.1$), shown in the forest plot (Figure 1B). According to the value of hazard ratio (HR), ATP6V1G2, SRC, IL6, CEBPG, and NGB were considered the genes with the highest risk. To examine the prognostic significance of these screened risk FRGs, DNA Damage Inducible Transcript 3 (DDIT3), one of the risk FRGs, was performed as an example. 3.37% of SKCM patients were observed to carry DDIT3 mutations and these patients significantly had worse OS than patients without DDIT3 mutations by Kaplan-Meier survival analysis (Figures 1C,D).

Construction and validation of the ferroptosis-related genetic mutation risk (FerrGR) model

To prevent the risk of over-fitting, the LASSO Cox regression analysis was performed to establish a prognostic prediction model based on whether patients carrying SNV mutations in the above screened 24 FRGs or not. As a result, 19 key genes (TP63, CDKN2A, MTOR, EGFR, BRD4, PLIN4, GCLC, HELLS, MAPK9, FH, PHKG2, DDIT3, SLC11A2, SRC, CISD2, PLIN2, IL6, HSD17B11, ATP6V1G2) were filtered out by the minimum value of lambda (λ) (Supplementary Figure S2). The coefficients of these genes were shown in Figure 2A. The risk score was calculated with the following formula: 0.295704560557983 × SNV mutational value of TP63 + (−0.457905201754129) × SNV mutational value of CDKN2A + (−0.440669922200648) × SNV mutational value of MTOR + 0.370753348528944 × SNV mutational value of EGFR + (−0.655423174928644) × SNV mutational value of BRD4 + (−0.257562121098952) × SNV mutational value of PLIN4 + 0.790692660435278 × SNV mutational value of GCLC + 0.485910362202744 × SNV mutational value of HELLS + 0.70467787865497 × SNV mutational value of MAPK9 + 0.851337406732927 × SNV mutational value of FH + 0.947077530588054 × SNV mutational value of PHKG2 + 1.45060826783863 × SNV mutational value of DDIT3 + 2.28402264491982 × SNV mutational value of SLC11A2 + 2.06527868435253 × SNV mutational value of SRC + 1.55256390491241 × SNV mutational value of CISD2 + 5.80583718370437 × SNV mutational value of PLIN2 + 2.7431230119215 × SNV mutational value of IL6 + 1.60058988692493 × SNV mutational value of HSD17B11 + 0.0133531074086934 × SNV mutational value of ATP6V1G2.

The patients in the training cohort were classified into the high ferroptosis-related genetic mutation risk (high FerrGR) group and low ferroptosis-related genetic mutation risk (low FerrGR) group by the median risk score as a cut-off value, which was calculated as 0.2467727. The SNV landscape of 19 key FRGs in the TCGA training cohort was further figured out based on the two subgroups (Figure 2B). The Kaplan-Meier analysis indicated that patients in the high FerrGR group had significantly worse OS than those in the low FerrGR group (Figure 2C).

To test the reliability of the FerrGR model, the same formula as the training cohort was performed to calculate risk scores for the patients in the validation cohort. The patients were then allocated into the high

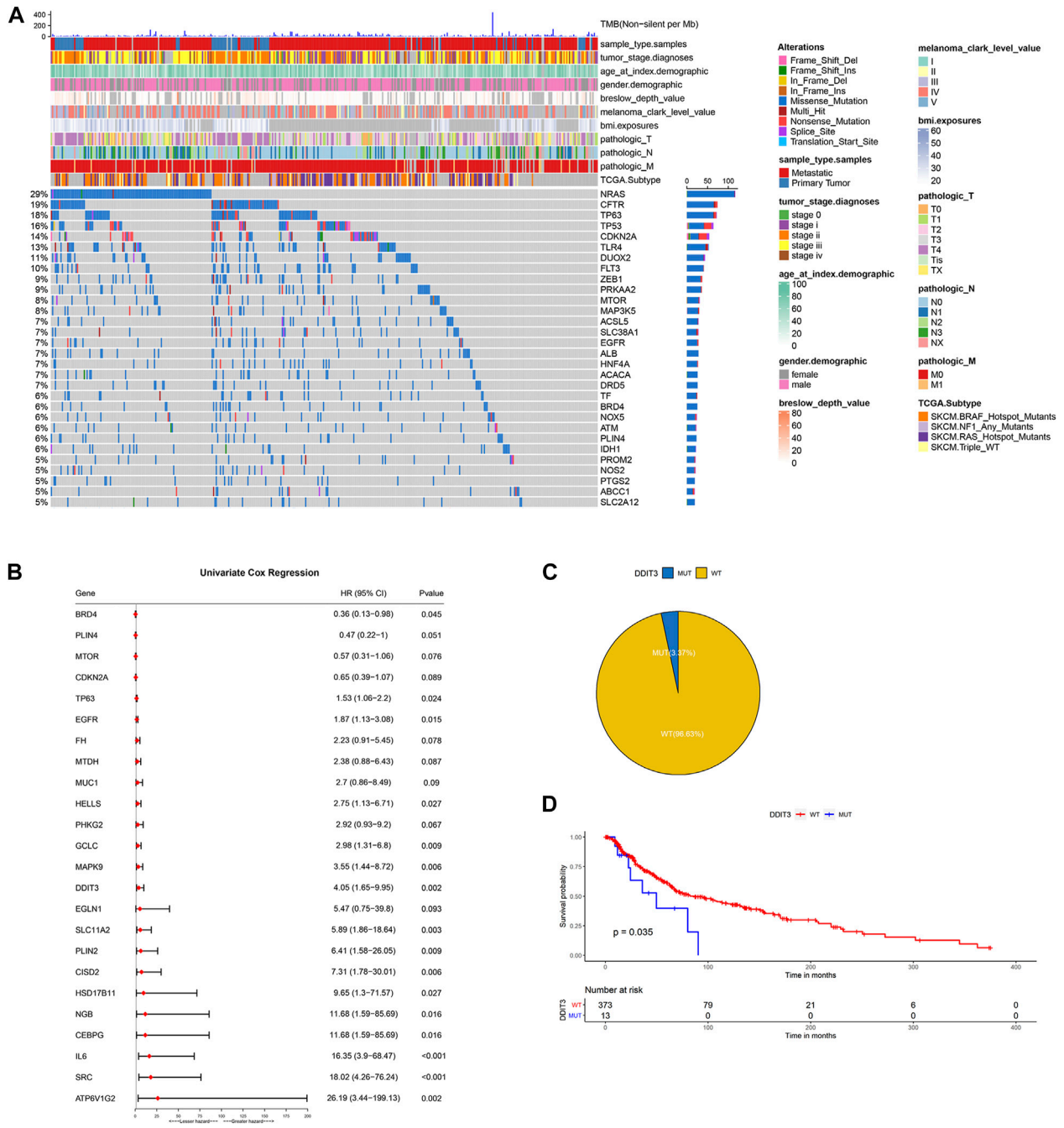


FIGURE 1
Identification of prognosis-related key FRGs in SKCM. **(A)** Heatmap to show the SNV landscape of the top 30 FGRs with the most frequent SNV mutations in the TCGA-SKCM cohort. **(B)** Forest plots showing the results of the univariate Cox regression analysis between FGRs and prognosis. **(C)** Pie charts depicting the proportions of wild-type and DDIT3-mutant patients. **(D)** Kaplan-Meier survival analysis of the wild-type and DDIT3-mutant patients.

FerrGR group and low FerrGR group by the same cut-off value. The SNV landscape of 19 key FRGs for patients in the validation cohort was shown in Figure 2D. Similar to the training cohort, The high FerrGR group exhibited a poorer survival outcome when compared to the low FerrGR group (Figure 2E).

Subsequently, we used the ROC curve to evaluate the prediction efficacy of the model by calculating the areas under the curve (AUC). The AUCs of the FerrGR model for one-year survival time were 0.643 in the training cohort and 0.721 in the validation cohort respectively; Besides, the FerrGR model showed the best prognostic power compared with TMB and MSI (Figures 2F,G).

Correlations between the FerrGR score and clinicopathological factors

To further explore the roles of the FerrGR model in the SKCM development, the correlations between the FerrGR score and

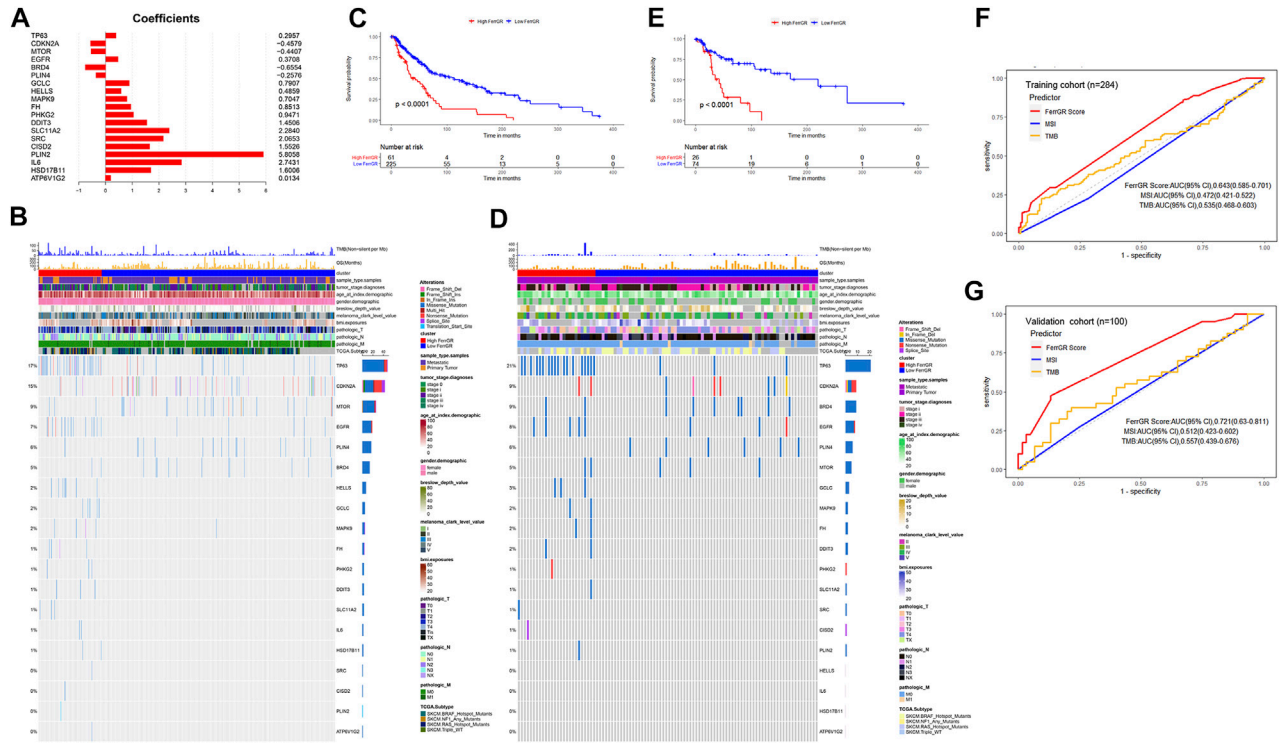


FIGURE 2
Construction and prognostic analysis of the ferroptosis-related genetic mutation risk (FerrGR)model in the training cohort and validation cohort. **(A)** Lasso coefficient spectrum of 19 key FRGs in the FerrGR model **(B,D)** SNV heatmap and clinicopathologic features of 19 key FRGs in the **(B)** training cohort and **(D)** validation cohort. **(C,E)** Kaplan-Meier curves of the FerrGR model for SKCM patients with different risk groups in the **(C)** training cohort and **(D)** validation cohort. **(F,G)** ROC analysis of the FerrGR model compared with TMB and MSI.

clinicopathological factors were studied. Our results showed that the FerrGR score was independent with sample type ($p = 0.19$, Figure 3A), gender ($p = 0.66$, Figure 3C), T stage ($p = 0.13$, Figure 3E) and N stage ($p = 0.93$, Figure 3F). Further, there may be some correlation between the FerrGR score and tumor stage ($p = 0.082$, Figure 3B). The FerrGR score in stage II patients was higher than in other stages. Furthermore, The FerrGR score in stage M0 patients was higher than in the M1 stage ($p = 0.065$, Figure 3G). In particular, the FerrGR score was significantly among the values of Clark levels ($p = 0.012$, Figure 3D), and the signature was associated with TCGA subtypes ($p = .017$, Figure 3H). The FerrGR score in Clark level III patients was the highest. Besides, patients with NF1 mutations have a higher score than patients in other TCGA subtypes.

Independent prognostic factors analysis and nomogram prediction model construction

To evaluate whether the risk score was a suitable independent prognostic indicator, univariate and multivariate Cox regression analysis were performed among the clinical characteristics and risk scores in the TCGA cohort. The univariate Univariate Cox regression revealed that clinical parameters, including primary tumor, T4 stage, N2 stage, N3 stage, NF1 mutated subtype, RAS mutated subtype, triple wild type, low FerrGR score, age ≥ 60, Breslow depth value >4.5, Breslow depth value = (3–4.5) were significantly associated with OS (Figure 4A). Through multivariate Cox regression, N2 stage, N3 stage, low FerrGR score, age ≥ 60, and Breslow depth value >4.5 were independent predictors of SKCM (Figure 4B).

What's more, a nomogram was created based on the values of multiple variables to predict the probability of specific clinical outcomes or events. We constructed the nomogram with the following factors: Breslow depth value, age, FerrGR score, and N stage. In the nomogram, columnar height represents the distribution and number of SKCM patients (Figure 4C). Testing of the proportional hazards hypothesis demonstrated the individual and global variables satisfied the requirement of the hypothesis (Figure 4D). Additionally, the calibration curve for the 1-, 2-, and 3-year survival rates displayed good agreement between the prediction and the investigation (Figure 4E).

Mutation landscape of FRGs between the high FerrGR group and low FerrGR group

Further, the SNV mutation profiles of FRGs in 284 SKCM patients were utilized to explore the different landscape of SNVs in high—and low—FerrRG group patients. Among these patients, 60 belonged to the high FerrRG group and 100% had SNV alterations, while 224 were classified into the low FerrRG group and 136 (60.71%) carried SNV mutations in FRGs. We then

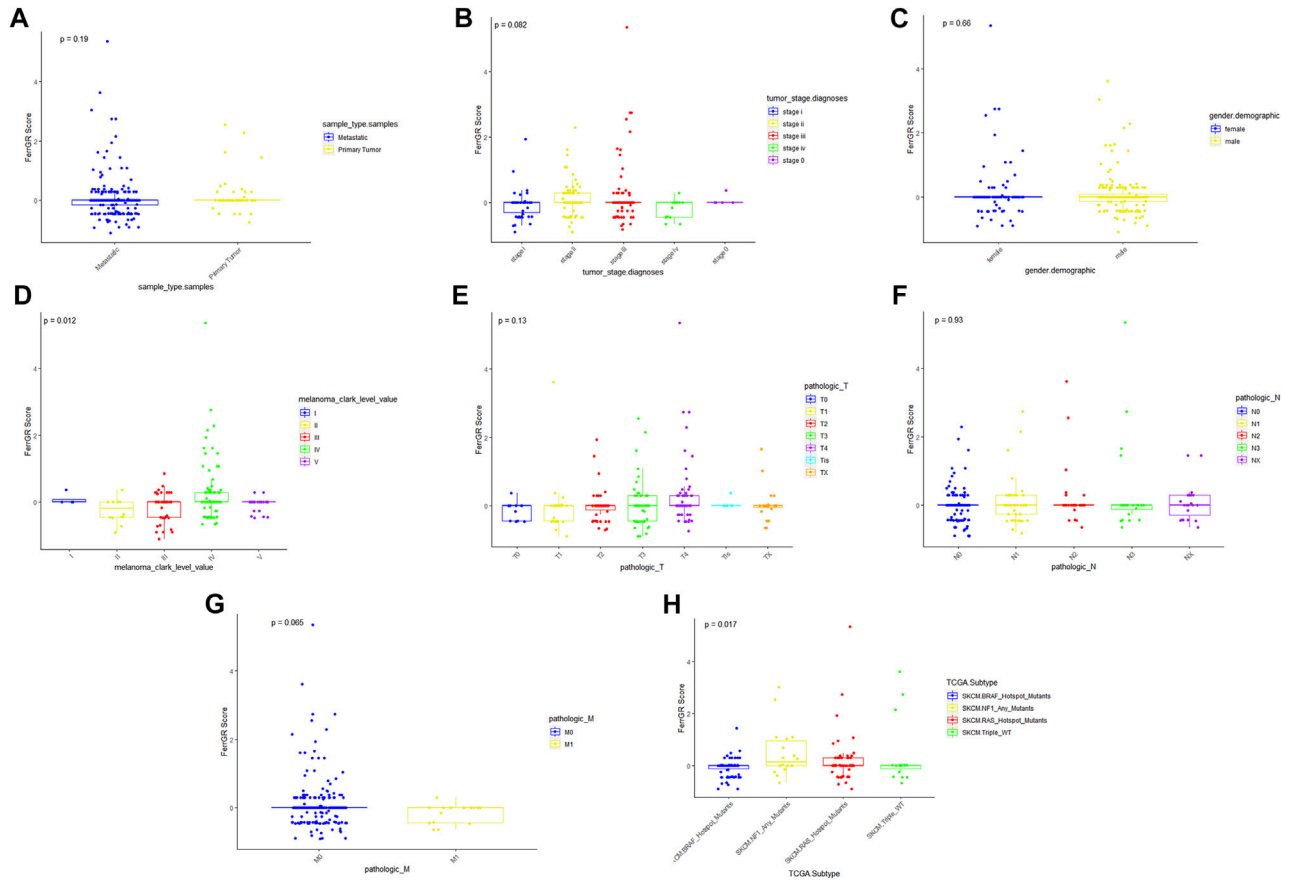


FIGURE 3
Relationships between the FerrGR score and clinicopathological features. The boxplots showed whether the FerrGR score was correlated with pathological features in SKCM patients, including **(A)** sample type, **(B)** tumor stage, **(C)** gender, **(D)** Clark level, **(E)** T stage, **(F)** N stage, **(G)** M stage, and **(H)** TCGA subtype.

collected SNV mutation information in each sample of both groups and presented the top 30 FRGs in Figures 5A,B, respectively. We revealed TP63 (55%), NRAS (42%), CFTR (27%), EGFR (22%), and FLT3 (17%) were the top 5 FRGs with highest mutation frequencies in the high FerrGR group, and NRAS (26%), CFTR (17%), CDKN2A (17%), FLT3 (8%) and TP63 (7%) were top 5 in the low FerrGR group. Notably, missense mutation was the largest fraction of mutation types in both groups.

Next, the differential expression of the top 30 FRGs between the high FerrGR group and the low FerrGR group was exhibited (Figure 5C). We found that CASP8, ATP6V1G1, RRM2, ENPP2, and TFAP2C were the top five differentially expressed genes. CNV analysis then showed RELA and NOX4 were the two FRGs with significantly different CNVs ($p < 0.1$) between high - and Low - FerrGR groups. RELA and NOX4 in the high FerrGR group possessed more widespread CNV deletion (Figures 5D,E). However, there was no significant difference in the CNV status of other FRGs between the two groups (Supplementary Table S3). DNA methylation is an important consideration in the pathogenesis of cancer (McMahon et al., 2017). Therefore, the heat map summarized the 30 most significant FRGs-associated DNA methylation sites between two groups (Figure 5F).

FerrGR-based prognostic stratification of SKCM patients with immunotherapy

Immunotherapy is an innovative treatment strategy for cancers. In particular, immune checkpoint blockade (ICB) therapy has made great progress in immunotherapy for cancer patients (Havel et al., 2019). Hence, we firstly determined the differences in the expression levels of 61 immune checkpoints between the high FerrGR group and low FerrGR group of the GSE91061 dataset (Figure 6A). We then revealed that there was no significant difference in patient OS between these two groups (Figure 6B). Subsequently, the response to immunotherapy was studied and found that no significant difference in immunotherapy responses between the high FerrGR group ($n = 18$) the and low FerrGR group ($n = 80$), implying that the FerrGR model may not be a direct biomarker of immunotherapy (Figure 6C). Thus, we further investigated the joint utility of FerrGR combined with TMB or MSI for patient stratification and prediction of clinical outcomes. The FerrGR-high/TMB-high and FerrGR-high/MSI-high (both high) subgroups had a remarkably poorer survival outcome compared with the subgroups where both were low or single was high (Figures 6D,E). These results demonstrated that a combination of FerrGR and TMB/MSI served as a combined biomarker with better predictive value for favorable ICIs benefit.

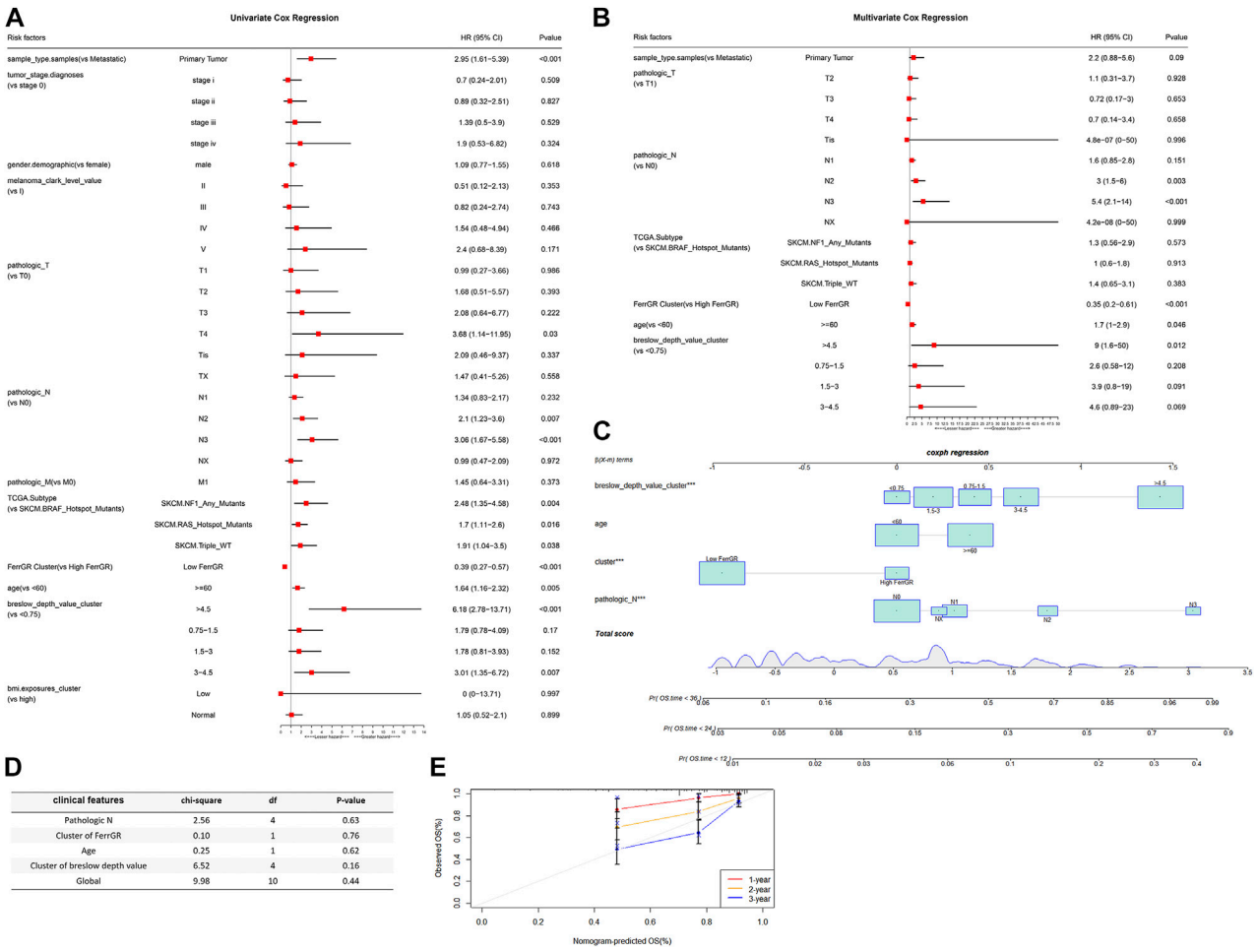


FIGURE 4
The connection between FerrGR score and conventional clinical characteristics. **(A,B)** Univariate and multivariate regression analysis of FerrGR score and clinical characteristics in prognostic value showed FerrGR score had excellent prognostic independence, **(C)** Prognostic nomogram for SKCM patients with factors, including Breslow depth value, age, FerrGR score and N stage, **(D)** Test for the proportional hazards hypothesis, **(E)** Calibration maps for predicting patient survival at 1, 2, and 3 years. The x-axis and y-axis represent the expected and actual survival rates of the nomogram.

Identification of the relationship between the FerrGR model and tumor immune microenvironment

To better study how the FerrGR model and the immune microenvironment interact, we firstly evaluated the different distribution of clinicopathological features between two FerrGR group patients of the TCGA cohort, and revealed that patients in the high FerrGR group had a lower immune score, higher tumor grade, and higher tumor purity than in low FerrGR group (Figure 7A). The distribution patterns of 22 immune cells between two groups were next calculated by the CIBERSORT algorithm. The comprehensive comparisons with the FerrGR score showed that B cells naive and T cells regulatory (Tregs) were enriched in the high FerrGR group obviously, while the patients in the low FerrGR group had a higher level of T cells CD4 memory activated (Figure 7B). It is known that immune checkpoint genes usually make an immunosuppressive effect in tumorigenesis and immune evasion. Therefore, the expression levels of immune checkpoint genes in high—and low—FerrGR groups were compared, and the results indicated that the expression levels of common immune checkpoint genes, including CD274 (PD-L1), CD80, CD86 and PDCD1LG2 (PD-L2), in the low FerrGR group were all higher than those in the high FerrGR group (Figures 7C–F). However, there was no remarkable differential expression of CTLA4 between these two groups (Figure 7G). Subsequently, GSEA was performed to determine the biological functions and signal transduction pathway associated with the FerrGR score. The results showed that the FerrGR score was negatively correlated with inflammatory response, interferon-alpha response, interferon-gamma response, antigen processing and presentation, and the JAK-STAT signaling pathway, respectively (Figure 7H). These findings revealed that SKCM patients with high FerrGR scores prefer to form a suppressive immune microenvironment by increasing suppressive immune infiltration cells and upregulating immune checkpoint genes.

Potential sensitive drugs for SKCM according to the FerrGR model

According to the data on drug sensitivity and expression, 1,311 and 481 potential sensitive compounds were figured out from

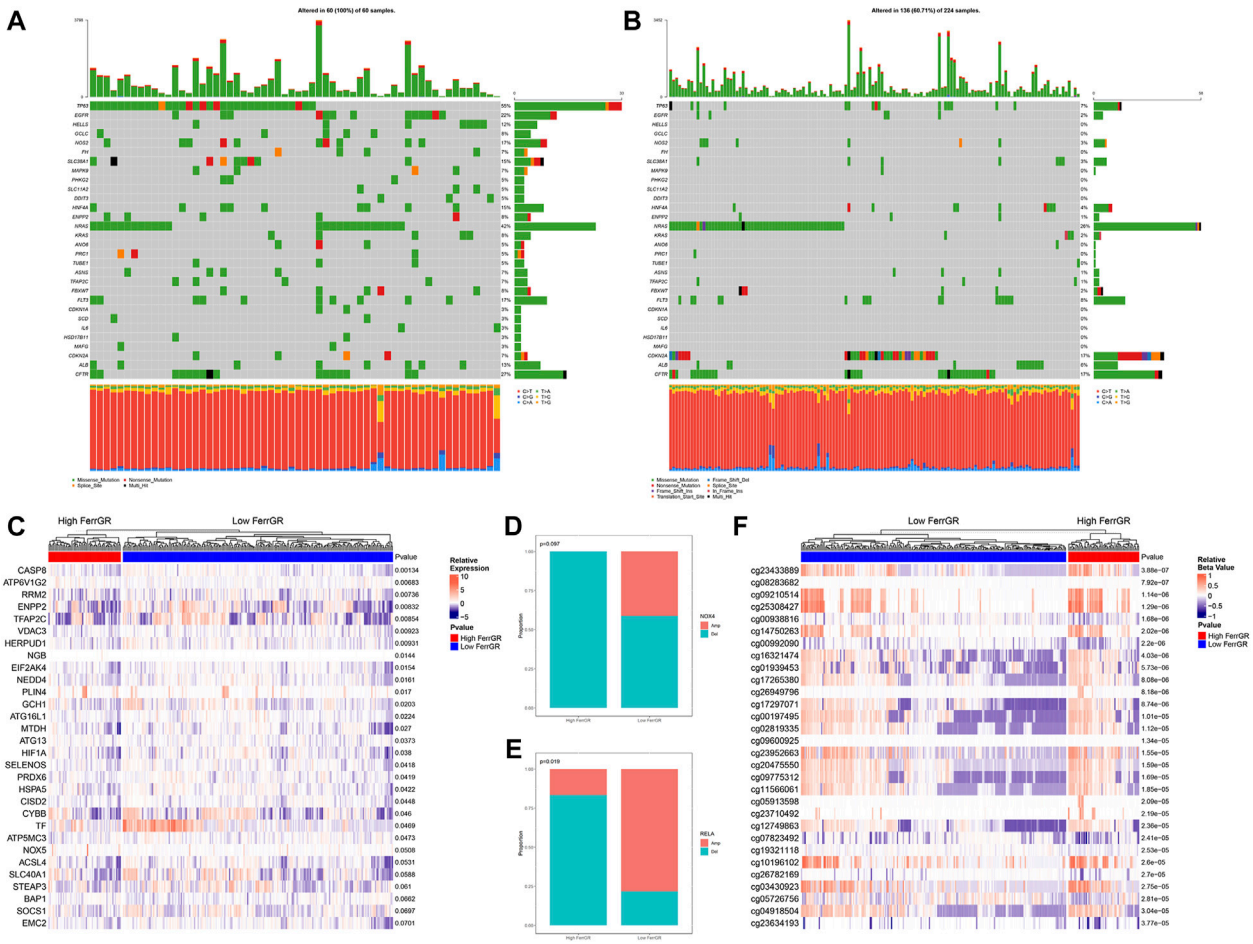


FIGURE 5
Analysis of ferroptosis-related genomic variation in the high FerrRG group and low FerrRG group **(A,B)** waterfall plots represent mutation information of FRGs in each sample of the high FerrGR group and low FerrGR group SKCM patients, **(C)** Heatmap of top 30 differentially expression FRGs between the high- and the low-FerrGR groups, **(D,E)** The CNV mutation proportion of **(D)** RELA, and **(E)** NOX4 between groups, Amp: Gene amplification, Del: gene deletion **(F)** Heatmap showed the methylation sites of FRGs with top 30 significantly different methylation levels between groups.

the PRISM and CTRP database respectively, and 152 overlapped compounds were filtered out (Figure 8A). It was accepted that values of AUC and IC50 represented the sensitivity of the cells to drugs and were negatively correlated with the sensitivity. We then identified the top differential AUC value and IC50 value between high—and low—FerrGR group samples, and determined a threshold to select potential compounds. The Spearman's correlation >0.2 was set as the threshold. The AUC values of pevonedistat, crystal-violet, bardoxolone-methyl, BNTX from the PRISM database, and cerulenin, HBX-41108 from the CTRP database exhibited significant correlations with the FerrGR score. Apart from BNTX showing a positive correlation with the FerrGR score, the other five selected compounds had negative correlations (Figure 8B). Besides, differential distribution of the AUC value of six potential compounds in high—and low—FerrGR groups was depicted (Figure 8C, D). Similarly, pevonedistat, crystal-violet, bardoxolone-methyl, and BNTX were identified from the PRISM database based on their IC50 values, and no potential compounds were found in the CTRP database. BNTX had a positive correlation with the FerrGR score, while pevonedistat, crystal-violet, and bardoxolone-methyl had negative correlations (Figure 8E). The differential distribution of the IC50 value of these four potential compounds in high—and low—FerrGR groups was exhibited in Figure 8F. Therefore, those compounds may be novel options for SKCM treatments in the future.

Discussion

SKCM is highly heterogeneous in the genetic, epigenetic, and gene expression with high metastases and death threats (Grzywa et al., 2017; Hendrix et al., 2017). Understanding the rapid progression of this heterogeneity makes possible the molecular classification and individualized treatment of SKCM. Ferroptosis has gained the interest of numerous researchers due to its unique cell death mechanism and its potential therapeutic prospects in cancers (Jiang et al., 2021). Current studies have constructed several prognostic prediction models for SKCM based on the expression of FRGs. Zeng et al. developed a prognostic model depending on the expression of two FRGs (ALOX5, CHAC1), and differences in the underlying diseases of SKCM did not effect on the expression features of these two genes (Zeng et al., 2021). Additionally, studies showed five-, six-, eight-, nine- and ten- FRG predictive models according to

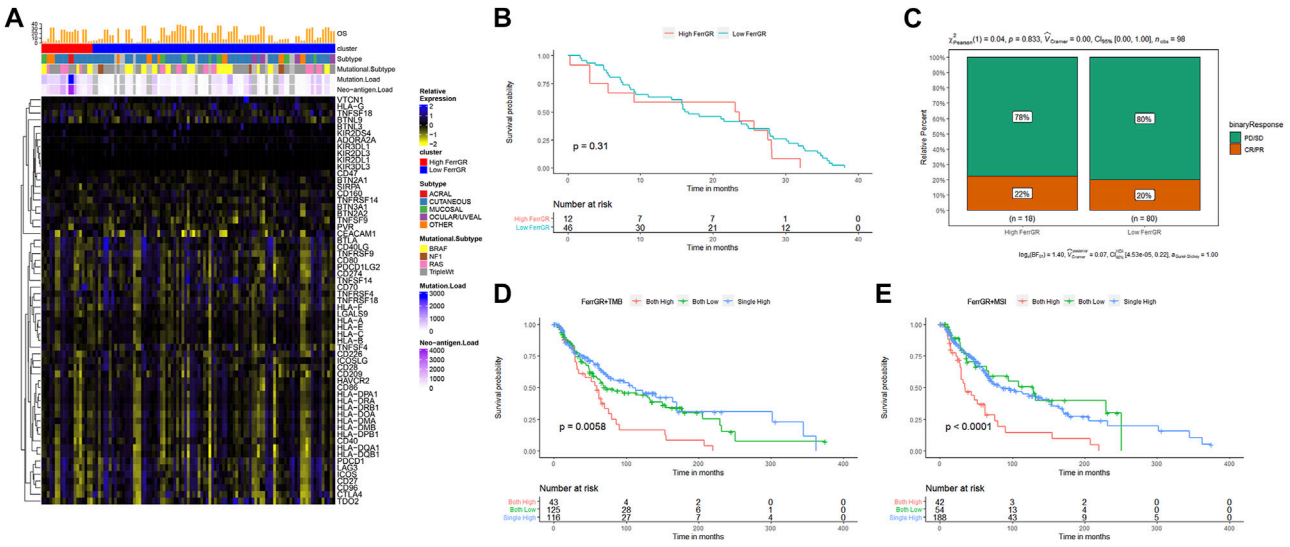


FIGURE 6
Evaluation of immunotherapy in SKCM with FerrGR model. **(A)** Heatmap displaying the expression levels of 61 immune checkpoint genes in the high FerrGR group and low FerrGR group. **(B)** Kaplan-Meier curves of the high-FerrGR and low-FerrGR group patients in the GSE91061 dataset. **(C)** The proportion of immune response to immunotherapy of high- and low-FerrGR groups in the TCGA cohort. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease. **(D)** Kaplan - Meier survival analysis of OS among patients within each of the three indicated subgroups (Both high: FerrGR-high/TMB -high; Both low: FerrGR-low/TMB-low; Single high: FerrGR-high/TMB-low or FerrGR-low/TMB-high). **(E)** Kaplan-Meier survival analysis of OS among patients within each of the three indicated subgroups (Both high: FerrGR-high/MSI -high; Both low: FerrGR-low/MSI-low; Single high: FerrGR-high or MSI-high).

the RNA sequencing data have been constructed (Xu Z et al., 2021; Xu and Chen, 2021; Chen et al., 2022; Ping et al., 2022; Yue et al., 2022). These models forecasted the melanoma patients' prognosis and exhibited the close relationship between immune function and FRGs. However, the role of ferroptosis in SKCM patients, especially the mechanism of the association and interaction between ferroptosis-related genetic mutations and the clinical outcomes is still unclear. After performing a series of bioinformatics analyses, we found that SNVs of FRGs are an indicator of prognosis and TME status in SKCM patients, which may be of great significance for future research. Therefore, based on the SNV landscape, we systematically identified FRGs with prognostic ability to establish a robust and accurate ferroptosis-associated genetic mutation risk model to predict prognosis in SKCM patients and illustrate the relationship between ferroptosis-related SNVs and the TME.

In this study, the TCGA-SKCM cohort was used to perform univariate Cox regression combined with the previously reported and identified 24 FRGs that were correlated with SKCM prognosis. Subsequently, the LASSO algorithm was used to reduce dimensionality and construct a 19-gene signature prognostic model (FerrGR model). We verified the effectiveness of this model in the training cohort and the validation cohort. The FerrGR score of each sample is calculated on basis of whether the sample has SNV mutations in the 19 key genes or not. Then, patients in the training cohort and validation cohort were classified into the high FerrGR group and the low FerrGR group. The results showed that it is an independent, effective and robust prognostic model in both cohorts where the prognosis was worse in the high FerrGR group. In addition, aiming at the characteristics of high heterogeneity in SKCM patients, we established and validated a nomogram based on FerrGR score and clinicopathological indications that can predict 1-, 2-, and 3-year OS for individual SKCM patients specifically.

Nowadays, high-throughput sequencing technologies allow us to detect numerous genes which are significantly related to melanoma prognosis through comprehensive analyses and establish multiple biomarkers. Therefore, recent studies explored novel favorable prognostic genes, such as aging-related genes, metabolic genes and pyroptosis-related genes, to predict prognosis and immune response for SKCM (Ju et al., 2021; Guo et al., 2022; Zeng et al., 2022). However, most of these studies did not systematic and in-depth analysis of the genetic mutations of these prognostic genes. SNVs are somatic point mutations found in cancer tissues and enriched in cancer driver genes and cellular pathways which are essential for tumorigenesis. Tumorigenesis is an evolutionary process of accumulation of somatic mutations (driver mutations), which promotes a selective growth advantage for cancer cells (He et al., 2014). Several pathogenic CNVs in special genes have been reported in the beginning and development of breast cancer subtypes, including BRCA1, MTUS1, and hTERT, suggesting that CNVs also play a unique role in breast cancer (Frank et al., 2007; Silva et al., 2014). Here, we found that the proportion of SNV mutations in ferroptosis-related genes associated with SKCM was 82.17% (235/286), among which the SNV mutation frequency of NRAS (20%), CFTR (19%), and TP63 (18%) ranked the top three. SNV mutations in FGRs might play an important role in the development and progression of SKCM. Meanwhile, the SNV mutation frequency of FGRs in the high FerrGR group was higher than in the low FerrGR group, especially TP63 and NRAS. TP63 mutations, which are present in the majority of cancers, are associated with poorer clinical outcomes in SKCM (Matin et al., 2013; Monti et al., 2017), which is consistent with our findings. The relationship between TP53 and NRAS mutational status and SKCM survival was substantially more pronounced. NRAS mutations are discovered in 15% of SKCM cases and more likely to have an aggressive tumor (Kelleher and McArthur, 2012; Muñoz-Couselo et al., 2017). In addition to SNV mutations, CNV mutations were also assessed in this study. However, the notable differential CNV

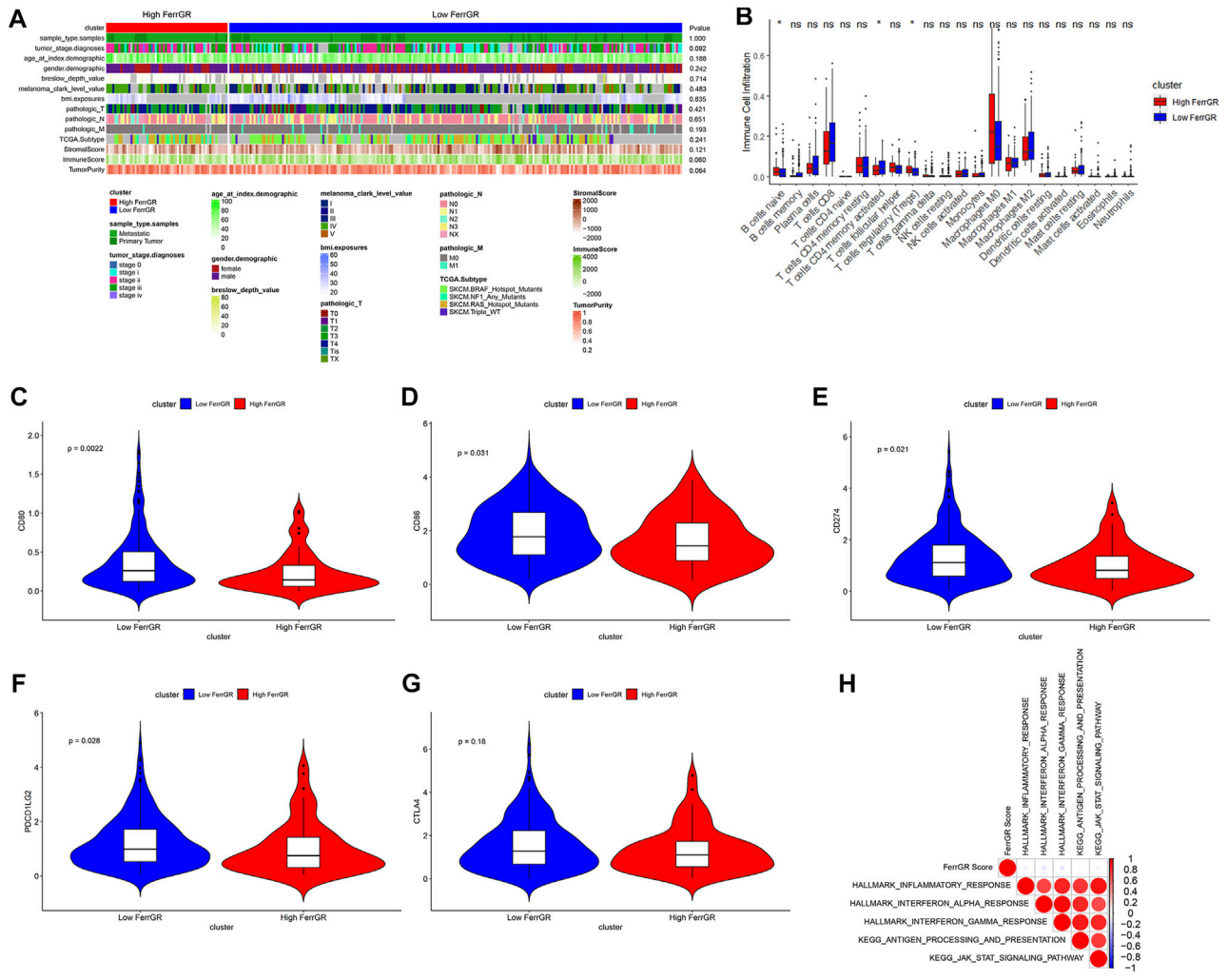


FIGURE 7
Tumour immune microenvironment analysis of FerrGR model. **(A)** The heatmap showed the correlations between FerrGR score and clinicopathological features. **(B)** Boxplot exhibited the distribution patterns of immune cell infiltration between two FerrGR groups. **(C–G)** violin plots visualizing the expression of immune checkpoint genes in the high—and Low—FerrGR groups, **(C)** CD80, **(D)** CD86, **(E)** CD274, **(F)** PDCD1LG2, **(G)** CTLA4 **(H)** Gene set enrichment analysis (GSEA) results depicting enrichment of immune-related pathways based on FerrGR score.

variants between high—and low—FerrGR groups were only discovered in RELA and NOX4. It is consistent with literature reports that SKCM mainly had SNV mutations but rarely CNVs (Liu et al., 2020). We also discovered that two FerrGR groups have their unique methylation levels. Overall, these results indicated that there were differences in expression and variation of FRGs between the two FerrGR groups. Besides, SNVs are likely to be the reason why the FRGs are ferroptosis resistance genes.

ICIs, such as anti-PD-1 and anti-PD-L1 antibodies, exert their effects by releasing the braking effect of the anti-tumor response immune system. Despite the breakthrough in ICIs therapy, it did not work as well in all patients. Therefore, there is a strong interest in finding biomarkers that can identify good responses to ICIs treatment. Several studies showed the risk of TMB and neoepitopes had a close correlation with immunotherapy. High TMB contributed to low survival outcomes and lower SKCM immune infiltrates (Jiang F et al., 2020). In addition, tumors having a high mutation load tend to respond to PD-1 immunotherapy more quickly and have a better prognosis (Samstein et al., 2019). There are still some patients with high TMB who do not respond and *vice versa*. The important reasons are that TMB only focuses on the number of mutations, and the current TMB calculation method gives the same weight to each gene mutation, which is not precise enough to define the overall pattern of anti-tumor immune response (Sha et al., 2020). MSI is often caused by a mismatch repair deficiency (MMR), and patients with MMR have extremely high rates of TMB (Nakayama et al., 1983). These tumors also have a significant response to immunotherapy (Le et al., 2015; Le et al., 2017). However, MMR is uncommon in melanoma (Tomlinson et al., 1996; Richetta et al., 1997; Birindelli et al., 2000). Unfortunately, the model in this study cannot be used as an independent prognostic factor for immunotherapy. Therefore, we explored the effect of the FerrGR model in combination with TMB or MSI for patient stratification and prediction of clinical outcomes, and found that patients with high-FerrGR/high-TMB, and patients with high-FerrGR/high-MSI, had the worst outcomes. Although other literature suggests that SKCM patients with high TMB or high MSI

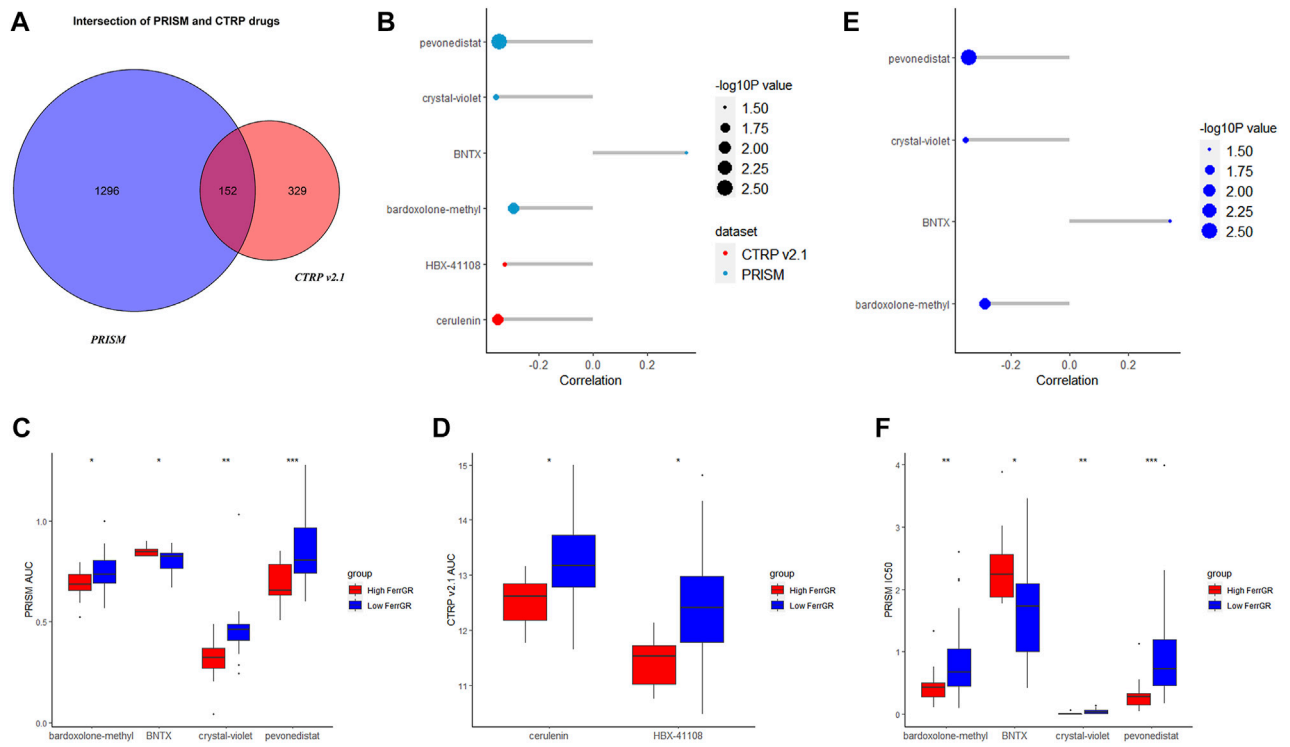


FIGURE 8
Potential targeted drugs prediction on basis of the FerrGR model. **(A)** The Venn chart showed the number of drugs in the PRISM dataset and CTRP v 2.1 databases. **(B)** The correlation between the AUC value of potential drugs and the FerrGR score. **(C,D)** The distribution of the AUC value of each potential drug from **(C)** RISM dataset and **(D)** CTRP v 2.1 database according to the FerrGR model. **(E)** The correlation between the IC50 value of potential drugs and the FerrGR score. **(F)** The distribution of the IC50 value of each potential drug on basis of the FerrGR model.

have a better prognosis (Sha et al., 2020), our results showed that these patients had a worse prognosis when the FerrGR score was high at the same time. Therefore, we hypothesized that the FerrGR score was a co-predictor with TMB and MSI, and that the FerrGR model could enhance the predictive ability of TMB and MSI.

The TME is made up of tumor cells and non-tumor cells that play a vital role in tumor growth and progression (Lim and South, 2014). Immune cells and stromal cells are two major types of non-tumor components in the TME. According to recent research, tumor progression can be caused by imbalances between tumor progression and the host immune response (Galon et al., 2013). Of note, ferroptosis can promote tumor growth by driving the polarization of macrophages in the TME (Dai et al., 2020). Moreover, hypoxia-inducible factor (HIF) pathways are a positive trigger for ferroptosis in clear-cell carcinoma (CCC) (Zou et al., 2019). Based on the ESTIMATE algorithm, we analyzed the relationship between the two groups and clinical features and assessed immune cell infiltration in TME. Compared with the low FerrGR group, the high FerrGR group had a higher immune score and higher tumor purity. Generally, immune scores increased significantly with the malignant progression of SKCM (Ning et al., 2021), while tumor purity decreased at higher grades in our analysis. Immune cells constitute a comfortable environment for tumor growth, suggesting that the poor prognosis of patients in the high FerrGR group is due to the tumor immunosuppressive environment (TIME). TIME is the immunosuppressive part of TME, which consists of immunosuppressive cells and immunosuppressive cytokines. Ferroptosis-related genes with a higher frequency of SNV mutations in the high FerrGR group may be associated with increased infiltration of immunosuppressive cells in SKCM. Tregs regulate innate and adaptive immune cells and maintain self-tolerance (Sakaguchi et al., 2008). A high proportion of Tregs is associated with tumor progression, poor survival in many solid tumors, including SKCM (Gerber et al., 2014), and poor clinical outcomes in SKCM patients treated with immunotherapy (Cesana et al., 2006). Here, our studies supported that a high proportion of Tregs in the high FerrGR group existed antitumor immune responses mediated by T cells. Conversely, the proportion of T cells CD4 activated in the low FerrGR group contributed more to immune response than in the high FerrGR group, according to our findings. Interestingly, we also found that almost all immune checkpoint genes, including PD-L1 and PD-L2, were upregulated in the high FerrGR group. Collectively, these results may be a sign of immune escape in the high FerrGR group patients.

Six potential targeted drugs, including pevonedistat, crystal-violet, bardoxolone-methyl, BNTX, cerulenin and HBX-41108, for high FerrGR samples were predicted. Pevonedistat (MLN4924) leads to DNA re-replication, cell cycle arrest and death *via* targeting the NEDD8-activating enzyme (NAE). It has anti-tumor activity and supports the clinical benefits observed in recent clinical trials in SKCM patients (Wong et al., 2017; Wood et al., 2020). For immunotherapy, the combination of pevonedistat and anti-PD-L1therapy had a better therapeutic efficacy compared to each agent alone. Pevonedistat attenuated T cell killing through PD-L1 induction,

whereas blockade of PD-L1 successfully potentiated the sensitivity of pevonedistat-treated glioblastoma cancer cells to T cell killing (Zhou et al., 2019). Hence, a combination of pevonedistat with immune checkpoint blockade treatment might be promising combinatorial regimens. Bardoxolone methyl is a novel synthetic triterpenoid and antioxidant inflammation modulator that activates Nrf2 and inhibits NF-κB. It can impair tumor growth and induces radiosensitization of oral squamous cell carcinoma cells (Hermann et al., 2021). But a further examination of its effects in SKCM is required. Cerulenin, a fatty acid synthase inhibitor, can retard the growth of melanoma cells and activates caspase-dependent apoptosis (Ho et al., 2007). Moreover, the anti-tumor immune responses of cytotoxic T cells were potentiated and ovarian tumor growth was inhibited by treatment with cerulenin (Yoon and Lee, 2022). It indicated that cerulenin might have potential applications in cancer immunotherapy. HBX-41108 is a partially-selective ubiquitin-specific proteases (USPs) inhibitor that stabilizes p53 and induces caspase 3 and PARP cleavage in cancer cells. As USPs are therapeutic targets for tumor treatment, HBX-41108 is likely to be an effective drug for SKCM (Pal and Donato, 2014). Therefore, the relationship between these potential targeted drugs and SNV mutations, ferroptosis, SKCM progression and immunotherapy needs further exploration.

Taken together, our results suggested that the FerrGR model based on SNV mutations of 19 key FRGs is a reliable prognostic risk prediction model for predicting the overall survival of SKCM patients. This may help guide treatment strategies for SKCM to improve clinical outcomes and provide theoretical references for explaining the prognosis difference between patients. Nevertheless, our study has several limitations. Firstly, there are not relatively abundant key FRGs in the risk model, which may limit it for clinical application. In addition, there was no significant difference between the two FerrGR groups in the immunotherapeutic response, thus more prospective real-world data should be used to confirm the accuracy and applicability of this model. Besides, further validation of this model in prospective studies of SKCM patients is needed.

Conclusion

In a word, we developed the FerrGR model for predicting the clinical outcomes and guiding the treatment of SKCM. It might have a contribution to distinguish immune and molecular features, stratify SKCM patients benefiting from immunotherapy, predict patient survival, and discover potential targeted drugs Our study provides new insights into genetic mutations of FRGs in SKCM's development and progression, and offers novel ideas for advancing the treatment of SKCM by targeting ferroptosis. However, further research on confirming the prognostic value of the FerrGR model is required.

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

XW and XC designed and organized the manuscript. JH and WH participated in data analysis and manuscript writing. XL, JW, and YN revised the manuscript. GL, XW, and HC contributed to collect data. All authors read and approved the submitted version.

Funding

This work was supported by grants from Regional Joint Fund of Natural Science Foundation of Guangdong Province (No. 2020A1515110432), Shenzhen Science and Technology Innovation Committee (No. JCYJ20200109142444449), Special fund of Foshan Summit plan (Nos. 2020A015 and 2019A006), Special Fund for Science and Technology Innovation Strategy of Guangdong Province (No. 2020A1515011402).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.988909/full#supplementary-material

References

Ackerman, A., Klein, O., McDermott, D. F., Wang, W., Ibrahim, N., Lawrence, D. P., et al. (2014). Outcomes of patients with metastatic melanoma treated with immunotherapy prior to or after BRAF inhibitors. *Cancer* 120 (11), 1695–1701. doi:10.1002/cncr.28620

Bai, X., Wu, D. H., Ma, S. C., Wang, J., Tang, X. R., Kang, S., et al. (2020). Development and validation of a genomic mutation signature to predict response to PD-1 inhibitors in non-squamous NSCLC: A multicohort study. *J. Immunother. Cancer* 8 (1), e000381. doi:10.1136/jitc-2019-000381

Basit, F., van Oppen, L. M., Schöckel, L., Bossenbroek, H. M., van Emst-de Vries, S. E., Hermeling, J. C., et al. (2017). Mitochondrial complex I inhibition triggers a mitophagy-dependent ROS increase leading to necroptosis and ferroptosis in melanoma cells. *Cell Death Dis.* 8 (3), e2716. doi:10.1038/cddis.2017.133

Birindelli, S., Tragni, G., Bartoli, C., Ranzani, G. N., Rilke, F., Pierotti, M. A., et al. (2000). Detection of microsatellite alterations in the spectrum of melanocytic nevi in patients with or without individual or family history of melanoma. *Int. J. Cancer* 86 (2), 255–261. doi:10.1002/(sici)1097-0215(20000415)86:2<255:aid-ijc16>3.0.co;2-l

Cesana, G. C., DeRaffele, G., Cohen, S., Moroziewicz, D., Mitcham, J., Stoutenburg, J., et al. (2006). Characterization of CD4+CD25+ regulatory T cells in patients treated with high-dose interleukin-2 for metastatic melanoma or renal cell carcinoma. *J. Clin. Oncol.* 24 (7), 1169–1177. doi:10.1200/JCO.2005.03.6830

Chen, Y., Guo, L., Zhou, Z., An, R., and Wang, J. (2022). Identification and validation of a prognostic model for melanoma patients with 9 ferroptosis-related gene signature. *Bmc Genomics* 23, 245. doi:10.1186/s12864-022-08475-y

Dai, E., Han, L., Liu, J., Xie, Y., Kroemer, G., Klionsky, D. J., et al. (2020). Autophagy-dependent ferroptosis drives tumor-associated macrophage polarization via release and uptake of oncogenic KRAS protein. *Autophagy* 16 (11), 2069–2083. doi:10.1080/15548627.2020.1714209

Frank, B., Bermejo, J. L., Hemminki, K., Sutter, C., Wappenschmidt, B., Meindl, A., et al. (2007). Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. *Carcinogenesis* 28 (7), 1442–1445. doi:10.1093/carcin/bgm033

Gagliardi, M., Saverio, V., Monzani, R., Ferrari, E., Piacentini, M., and Corazzari, M. (2020). Ferroptosis: A new unexpected chance to treat metastatic melanoma? *Cell Cycle* 19 (19), 2411–2425. doi:10.1080/15384101.2020.1806426

Galon, J., Angell, H. K., Bedognetti, D., and Marincola, F. M. (2013). The continuum of cancer immunosurveillance: Prognostic, predictive, and mechanistic signatures. *Immunity* 39 (1), 11–26. doi:10.1016/j.immuni.2013.07.008

Gao, M., Monian, P., Pan, Q., Zhang, W., Xiang, J., and Jiang, X. (2016). Ferroptosis is an autophagic cell death process. *Cell Res.* 26 (9), 1021–1032. doi:10.1038/cr.2016.95

Gerber, A. L., Münst, A., Schlapbach, C., Shafighi, M., Kiermeir, D., Hüsler, R., et al. (2014). High expression of FOXP3 in primary melanoma is associated with tumour progression. *Br. J. Dermatol* 170 (1), 103–109. doi:10.1111/bjd.12641

Grzywa, T. M., Paskal, W., and Włodarski, P. K. (2017). Intratumor and intertumor heterogeneity in melanoma. *Transl. Oncol.* 10 (6), 956–975. doi:10.1016/j.tranon.2017.09.007

Guo, X., Yu, X., Li, F., Xia, Q., Ren, H., Chen, Z., et al. (2022). Identification of survival-related metabolic genes and a novel gene signature predicting the overall survival for patients with uveal melanoma. *Ophthalmic Res.* 65, 516–528. doi:10.1159/000524505

Hassel, J. C., Lee, S. B., Meiss, F., Meier, F., Dimitrakopoulou-Strauss, A., Jäger, D., et al. (2016). Vemurafenib and ipilimumab: A promising combination? Results of a case series. *Oncoimmunology* 5 (4), e1101207. doi:10.1080/2162402X.2015.1101207

Hauschild, A., Grob, J., Demidov, L. V., Jouary, T., Gutzmer, R., Millward, M., et al. (2012). Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial. *Lancet* 380, 358–365. doi:10.1016/S0140-6736(12)60868-X

Havel, J. J., Chowell, D., and Chan, T. A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* 19, 133–150. doi:10.1038/s41568-019-0116-x

He, Q., He, Q., Liu, X., Wei, Y., Shen, S., Hu, X., et al. (2014). Genome-wide prediction of cancer driver genes based on SNP and cancer SNV data. *Am. J. Cancer Res.* 4 (4), 394–410.

Hendrix, M. J. C., Seftor, E. A., Margaryan, N. V., and Seftor, R. E. B. (2017). "Heterogeneity and plasticity of melanoma: Challenges of current therapies," in *Cutaneous melanoma: Etiology and therapy*. Editors W. H. Ward and J. M. Farma (Brisbane (AU): Codon Publications).

Hermann, C., Lang, S., Popp, T., Hafner, S., Steinritz, D., Rump, A., et al. (2021). Bardoxolone-methyl (CDDO-Me) impairs tumor growth and induces radiosensitization of oral squamous cell carcinoma cells. (CDDO-Me) impairs tumor growth and induces radiosensitization of oral squamous cell carcinoma cells. *Front. Pharmacol.* 11, 607580. doi:10.3389/fphar.2020.607580

Ho, T., Ho, Y., Wong, W., Chi-Ming Chiu, L., Wong, Y., and Eng-Choon Ooi, V. (2007). Fatty acid synthase inhibitors cerulenin and C75 retard growth and induce caspase-dependent apoptosis in human melanoma A-375 cells. *Biomed. Pharmacother.* 61, 578–587. doi:10.1016/j.biopha.2007.08.020

Hong, Z., Tang, P., Liu, B., Ran, C., Yuan, C., Zhang, Y., et al. (2021). ferroptosis-related genes for overall survival prediction in patients with colorectal cancer can be inhibited by gallic acid. *Int. J. Biol. Sci.* 17, 942–956. doi:10.7150/ijbs.57164

Hu, F., Liu, C., Liu, L., Zhang, Q., and Guo, A. (2020). Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. *Brief. Bioinform* 22, bbaa176. doi:10.1093/bib/bbaa176

Jiang, F., Wu, C., Wang, M., Wei, K., Zhou, G., and Wang, J. (2020). Multi-omics analysis of tumor mutation burden combined with immune infiltrates in melanoma. *Clin. Chim. Acta* 511, 306–318. doi:10.1016/j.cca.2020.10.030

Jiang, J., Ding, Y., Wu, M., Chen, Y., Lyu, X., Lu, J., et al. (2020). Integrated genomic analysis identifies a genetic mutation model predicting response to immune checkpoint inhibitors in melanoma. *Cancer Med.* 9 (22), 8498–8518. doi:10.1002/cam4.3481

Jiang, X., Stockwell, B. R., and Conrad, M. (2021). Ferroptosis: Mechanisms, biology and role in disease. *Nat. Rev. Mol. Cell Biol.* 22 (4), 266–282. doi:10.1038/s41580-020-00324-8

Jin, J., Wu, X., Yin, J., Li, M., Shen, J., Li, J., et al. (2019). Identification of genetic mutations in cancer: Challenge and opportunity in the new era of targeted therapy. Therapy. *Front. Oncol.* 9, 263. doi:10.3389/fonc.2019.00263

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

Kelleher, F. C., and McArthur, G. A. (2012). Targeting NRAS in melanoma. *Cancer J.* 18 (2), 132–136. doi:10.1097/PPO.0b013e31824ba4df

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

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

Leonardi, G. C., Candido, S., Falzone, L., Spandidos, D. A., and Libra, M. (2020). Cutaneous melanoma and the immunotherapy revolution (Review). *Int. J. Oncol.* 57 (3), 609–618. doi:10.3892/ijo.2020.5088

Leonardi, G. C., Falzone, L., Salemi, R., Zanghì, A., Spandidos, D. A., McCubrey, J. A., et al. (2018). Cutaneous melanoma: From pathogenesis to therapy (Review). *Int. J. Oncol.* 52 (4), 1071–1080. doi:10.3892/ijo.2018.4287

Li, J., Cao, F., Yin, H. L., Huang, Z. J., Lin, Z. T., Mao, N., et al. (2020). Ferroptosis: Past, present, and future. *Cell Death Dis.* 11 (2), 88. doi:10.1038/s41419-020-2298-2

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

Liang, J., Wang, D., Lin, H., Chen, X., Yang, H., Zheng, Y., et al. (2020). A novel ferroptosis-related gene signature for overall survival prediction in patients with hepatocellular carcinoma. Signature for overall survival prediction in patients with hepatocellular carcinoma. *Int. J. Biol. Sci.* 16, 2430–2441. doi:10.7150/ijbs.45050

Lim, Y. Z., and South, A. P. (2014). Tumour-stroma crosstalk in the development of squamous cell carcinoma. *Int. J. Biochem. Cell Biol.* 53, 450–458. doi:10.1016/j.biocel.2014.06.012

Liu, H., Zhang, B., and Sun, Z. (2020). Spectrum of EGFR aberrations and potential clinical implications: Insights from integrative pan-cancer analysis. *Cancer Commun. (Lond)* 40 (1), 43–59. doi:10.1002/cac2.12005

Loeb, L. A., Loeb, K. R., and Anderson, J. P. (2003). Multiple mutations and cancer. *PNAS* 100 (3), 776–781. doi:10.1073/pnas.0334858100

Luo, M., Wu, L., Zhang, K., Wang, H., Zhang, T., Gutierrez, L., et al. (2018). miR-137 regulates ferroptosis by targeting glutamine transporter SLC1A5 in melanoma. *Cell Death Differ.* 25 (8), 1457–1472. doi:10.1038/s41418-017-0053-8

Matin, R. N., Chikh, A., Chong, S. L., Mesher, D., Graf, M., Sanza, P., et al. (2013). p63 is an alternative p53 repressor in melanoma that confers chemoresistance and a poor prognosis. *J. Exp. Med.* 210 (3), 581–603. doi:10.1084/jem.20121439

McMahon, K. W., Karunasena, E., and Ahuja, N. (2017). The roles of DNA methylation in the stages of cancer. *Cancer J.* 23, 257–261. doi:10.1097/PPO.0000000000000279

Monti, P., Ghiorzo, P., Menichini, P., Foggetti, G., Queirolo, P., Izzotti, A., et al. (2017). TP63 mutations are frequent in cutaneous melanoma, support UV etiology, but their role in melanomagenesis is unclear. *Oncol. Rep.* 38 (4), 1985–1994. doi:10.3892/or.2017.5903

Muñoz-Couselo, E., Adelantado, E. Z., Ortiz, C., García, J. S., and Perez-Garcia, J. (2017). NRAS-Mutant melanoma: Current challenges and future prospect. *Onco Targets Ther.* 10, 3941–3947. doi:10.2147/OTT.S117121

Nakayama, T., Yasuda, K., Kamura, K., Yamashiro, Y., Hama, T., Shimazaki, J., et al. (1983). The continence mechanism. 1. Functional urethral length necessary for continence. *Nihon Hinyokika Gakkai Zasshi* 74 (3), 384–389. doi:10.5980/jpnjurol1928.74.3_384

Ning, X., Li, R., Zhang, B., Wang, Y., Zhou, Z., Ji, Z., et al. (2021). Immune score indicator for the survival of melanoma patients based on tumor microenvironment. *Int. J. Gen. Med.* 14, 10397–10416. doi:10.2147/IJGM.S336105

Pal, A., and Donato, N. J. (2014). Ubiquitin-specific proteases as therapeutic targets for the treatment of breast cancer. *Breast Cancer Res.* 16, 461. doi:10.1186/s13058-014-0461-3

Pelster, M. S., and Amaria, R. N. (2019). Combined targeted therapy and immunotherapy in melanoma: A review of the impact on the tumor microenvironment and outcomes of early clinical trials. *Ther. Adv. Med. Oncol.* 11, 1758835919830826. doi:10.1177/1758835919830826

Ping, S., Wang, S., Zhao, Y., He, J., Li, G., Li, D., et al. (2022). Identification and validation of a ferroptosis-related gene signature for predicting survival in skin cutaneous melanoma. *Cancer Med.* 11, 3529–3541. doi:10.1002/cam4.4706

Rebecca, V. W., Somasundaram, R., and Herlyn, M. (2020). Pre-clinical modeling of cutaneous melanoma. *Nat. Commun.* 11 (1), 2858. doi:10.1038/s41467-020-15546-9

Ribas, A., Lawrence, D., Atkinson, V., Agarwal, S., Miller, W. H., Jr., Carlino, M. S., et al. (2019). Combined BRAF and MEK inhibition with PD-1 blockade immunotherapy in BRAF-mutant melanoma. *Nat. Med.* 25 (6), 936–940. doi:10.1038/s41591-019-0476-5

Richetta, A., Silipo, V., Calvieri, S., Frati, L., Ottini, L., Cama, A., et al. (1997). Microsatellite instability in primary and metastatic melanoma. *J. Invest. Dermatol* 109 (1), 119–120. doi:10.1111/1523-1747.ep12276783

Sakaguchi, S., Yamaguchi, T., Nomura, T., and Ono, M. (2008). Regulatory T cells and immune tolerance. *Cell* 133 (5), 775–787. doi:10.1016/j.cell.2008.05.009

Samstein, R. M., Lee, C. H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* 51 (2), 202–206. doi:10.1038/s41588-018-0312-8

Sha, D., Jin, Z., Budczies, J., Kluck, K., Stenzinger, A., and Sinicrope, F. A. (2020). Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* 10 (12), 1808–1825. doi:10.1158/2159-8290.CD-20-0522

Silva, F. C., Lisboa, B. C., Figueiredo, M. C., Torrezan, G. T., Santos, E. M., Krepischi, A. C., et al. (2014). Hereditary breast and ovarian cancer: Assessment of point mutations and copy number variations in Brazilian patients. *BMC Med. Genet.* 15, 55. doi:10.1186/1471-2350-15-55

Simeone, E., Grimaldi, A. M., Festino, L., Giannarelli, D., Vanella, V., Palla, M., et al. (2017). Correlation between previous treatment with BRAF inhibitors and clinical

response to pembrolizumab in patients with advanced melanoma. *Oncoimmunology* 6 (3), e1283462. doi:10.1080/2162402X.2017.1283462

Talty, R., and Bosenberg, M. (2022). The role of ferroptosis in melanoma. *Pigment. Cell Melanoma Res.* 35 (1), 18–25. doi:10.1111/pcmr.13009

Tang, Y., Li, C., Zhang, Y., and Wu, Z. (2021). ferroptosis-related long non-coding rna signature predicts the prognosis of head and neck squamous cell carcinoma. *Int. J. Biol. Sci.* 17, 702–711. doi:10.7150/ijbs.55552

Tomlinson, I. P., Beck, N. E., and Bodmer, W. F. (1996). Allele loss on chromosome 11q and microsatellite instability in malignant melanoma. *Eur. J. Cancer* 32a (10), 1797–1802. doi:10.1016/0959-8049(96)00198-0

Wong, K. M., Micel, L. N., Selby, H. M., Tan, A. C., Pitts, T. M., Bagby, S. M., et al. (2017). Targeting the protein ubiquitination machinery in melanoma by the NEDD8-activating enzyme inhibitor pevonedistat (MLN4924). *Invest. New Drug* 35, 11–25. doi:10.1007/s10637-016-0398-8

Wood, E. A., Lu, Z., Jia, S., Assumpção, A. L. F. V., Van Hesteren, M. A., Huelsmeyer, M. K., et al. (2020). Pevonedistat targeted therapy inhibits canine melanoma cell growth through induction of DNA re- replication and senescence. *Vet. Comp. Oncol.* 18, 269–280.

Xu, C., and Chen, H. (2021). A ferroptosis-related gene model predicts prognosis and immune microenvironment for cutaneous melanoma. *Front. Genet.* 12, 697043. doi:10.3389/fgene.2021.697043

Xu, G., Wang, H., Li, X., Huang, R., and Luo, L. (2021). Recent progress on targeting ferroptosis for cancer therapy. *Biochem. Pharmacol.* 190, 114584. doi:10.1016/j.bcp.2021.114584

Xu, Z., Xie, Y., Mao, Y., Huang, J., Mei, X., Song, J., et al. (2021). Ferroptosis-related gene signature predicts the prognosis of skin cutaneous melanoma and response to immunotherapy. *Front. Genet.* 12, 758981. doi:10.3389/fgene.2021.758981

Yoon, H., and Lee, S. (2022). Fatty acid metabolism in ovarian cancer: Therapeutic implications. *Int. J. Mol. Sci.* 23, 2170. doi:10.3390/ijms23042170

Yue, Z., Sun, J., and Shi, L. (2022). Construction and validation of a 6-ferroptosis related gene signature for prognosis and immune landscape prediction in melanoma. *Front. Genet.* 13, 887542. doi:10.3389/fgene.2022.887542

Zeng, N., Guo, C., Wang, Y., Li, L., Chen, X., Gao, S., et al. (2022). Characterization of aging-related genes to predict prognosis and evaluate the tumor immune microenvironment in malignant melanoma. *J. Oncol.*, 1271378. doi:10.1155/2022/1271378

Zeng, N., Ma, L., Cheng, Y., Xia, Q., Li, Y., Chen, Y., et al. (2021). Construction of a ferroptosis-related gene signature for predicting survival and immune microenvironment in melanoma patients. *Int. J. General Med.* 14, 6423–6438. doi:10.2147/IJGM.S327348

Zhang, K., Wu, L., Zhang, P., Luo, M., Du, J., Gao, T., et al. (2018). miR-9 regulates ferroptosis by targeting glutamic-oxaloacetic transaminase GOT1 in melanoma. *Mol. Carcinog.* 57 (11), 1566–1576. doi:10.1002/mc.22878

Zhou, S., Zhao, X., Yang, Z., Yang, R., Chen, C., Zhao, K., et al. (2019). Neddylation inhibition upregulates PD-L1 expression and enhances the efficacy of immune checkpoint blockade in glioblastoma. *Int. J. Cancer* 145, 763–774. doi:10.1002/ijc.32379

Zhuo, S., Chen, Z., Yang, Y., Zhang, J., Tang, J., and Yang, K. (2020). Clinical and biological significances of a ferroptosis-related gene signature in glioma. Ferroptosis-related gene signature in glioma. *Front. Oncol.* 10, 590861. doi:10.3389/fonc.2020.590861

Zou, Y., Palte, M. J., Deik, A. A., Li, H., Eaton, J. K., Wang, W., et al. (2019). A GPX4-dependent cancer cell state underlies the clear-cell morphology and confers sensitivity to ferroptosis. *Nat. Commun.* 10 (1), 1617. doi:10.1038/s41467-019-09277-9

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fgene.2022.975404

OPEN ACCESS

EDITED BY
Zhiyuan Zhang,
Fudan University, China

REVIEWED BY
Nan Zhou,
Guangzhou Medical University, China
Qizhou Lian,
The University of Hong Kong, Hong Kong SAR, China

*CORRESPONDENCE
Zhongtao Liu,
158202088@csu.edu.cn
Yu Wen,
wenyu2861@csu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 22 June 2022
ACCEPTED 21 December 2022
PUBLISHED 09 January 2023

CITATION
Zhou H, Wang Y, Zhang Z, Xiong L, Liu Z and Wen Y (2023), A novel prognostic gene set for colon adenocarcinoma relative to the tumor microenvironment, chemotherapy, and immune therapy. *Front. Genet.* 13:975404. doi: 10.3389/fgene.2022.975404

A novel prognostic gene set for colon adenocarcinoma relative to the tumor microenvironment, chemotherapy, and immune therapy

Hui Zhou†, Yongxiang Wang†, Zijian Zhang, Li Xiong, Zhongtao Liu* and Yu Wen*

Department of General Surgery, Second Xiangya Hospital, Central South University, Changsha, China

Background: Colon adenocarcinoma (COAD) is a common aggressive malignant tumor. Heterogeneity in tumorigenesis and therapy response leads to an unsatisfactory overall survival of colon adenocarcinoma patients. Our study aimed to identify tools for a better prediction of colon adenocarcinoma prognosis, bolstering the development of a better personalized treatment and management.

Method: We used the least absolute shrinkage and selection operator (LASSO) Cox model to analyze the prognosis-related gene datasets from the Gene Expression Omnibus (GEO) database and verified them using The Cancer Genome Atlas (TCGA) database. The area under the curve (AUC) was calculated using the receiver operating characteristic (ROC) curve to evaluate the predictive ability of the risk score model. Gene Set Enrichment Analysis (GSEA) was used to identify the significantly enriched and depleted biological processes. The tumor immune dysfunction and exclusion (TIDE) algorithm was taken to explore the relationship between the risk score and immunotherapy. The observations collectively helped us construct a nomogram to predict prognosis. Finally, the correlation between drug sensitivity and prognostic gene sets was conducted based on the Cancer Therapeutics Response Portal (CTRP) analyses.

Results: We constructed a scoring model to assess the significance of the prognosis risk-related gene signatures, which was relative to common tumor characteristics and tumor mutational burdens. Patients with a high-risk score had higher tumor stage and poor prognosis ($p< 0.05$). Moreover, the expressions of these genes were in correlation with changes in the tumor microenvironment (TME). The risk score is an independent prognostic factor for COAD ($p< 0.05$). The accuracy of the novel nomogram model with a risk score and TNM-stage prediction prognosis in the predicting prognosis was higher than that of the TNM stage. Further analysis showed that a high-risk score was associated with tumor immune rejection. Patients with a low-risk score have a better prognosis with chemotherapy than those with a high-risk score. Compared to patients in the high-risk group, patients in the low-risk group had a significant survival advantage after receiving chemotherapy. In addition, the prognostic gene sets aid the assessment of drug sensitivity.

Conclusion: This study establishes a new prognostic model to better predict the clinical outcome and TME characteristics of colon adenocarcinoma. We believe, our

model also serves as a useful clinical tool to strengthen the functioning of chemotherapy, immunotherapy, and targeted drugs.

KEYWORDS

TMB, immune, prognosis, chemotherapy, drug sensitivity, colon adenocarcinoma

Introduction

Colon adenocarcinoma (COAD) is the main pathological type of colon cancer and the second leading cause of cancer deaths worldwide (Yoshino et al., 2018; Keum and Giovannucci, 2019). Approximately 900,000 COAD patients die each year from this malignancy due to its late clinical diagnosis (Dekker et al., 2019). Moreover, the incidence and mortality rates of COAD have been continuously growing, owing to the unsatisfactory prognosis of advanced COAD cases. The poor prognosis of COAD may also be due to its tumor recurrence and metastasis, characteristic of the disease. The 5-year and 10-year survival rates of most patients with metastatic COAD are 40% and 20%, respectively (Zhou et al., 2022). Treatment decisions are primarily based on assessing the tumor node metastasis (TNM) staging system (Amin et al., 2017). COAD is a heterogeneous cancer with genetic and clinicopathologic features regulating its occurrence and development (Gu et al., 2020). However, TNM staging fails to reveal its biological heterogeneity (Zhou et al., 2021). Moreover, an accurate prediction of the survival duration of COAD patients is helpful for clinical decision-making, warranting an urgent need to find more precise prognosis-predictive tools.

Currently, the standard treatment modalities for patients with COAD include surgery, adjuvant chemotherapy, and radiotherapy. It is challenging to remove all the cancer cells *via* surgery, causing advanced COAD patients to receive further treatment with adjuvant chemotherapy and radiotherapy (Ganesh et al., 2019). Chemotherapeutic drugs are non-specific and cytotoxic in nature with many side effects to any normal growing and dividing cell of the body. Notably, immunotherapy is one of the novel and current alternative treatments for COAD patients. Immune checkpoint

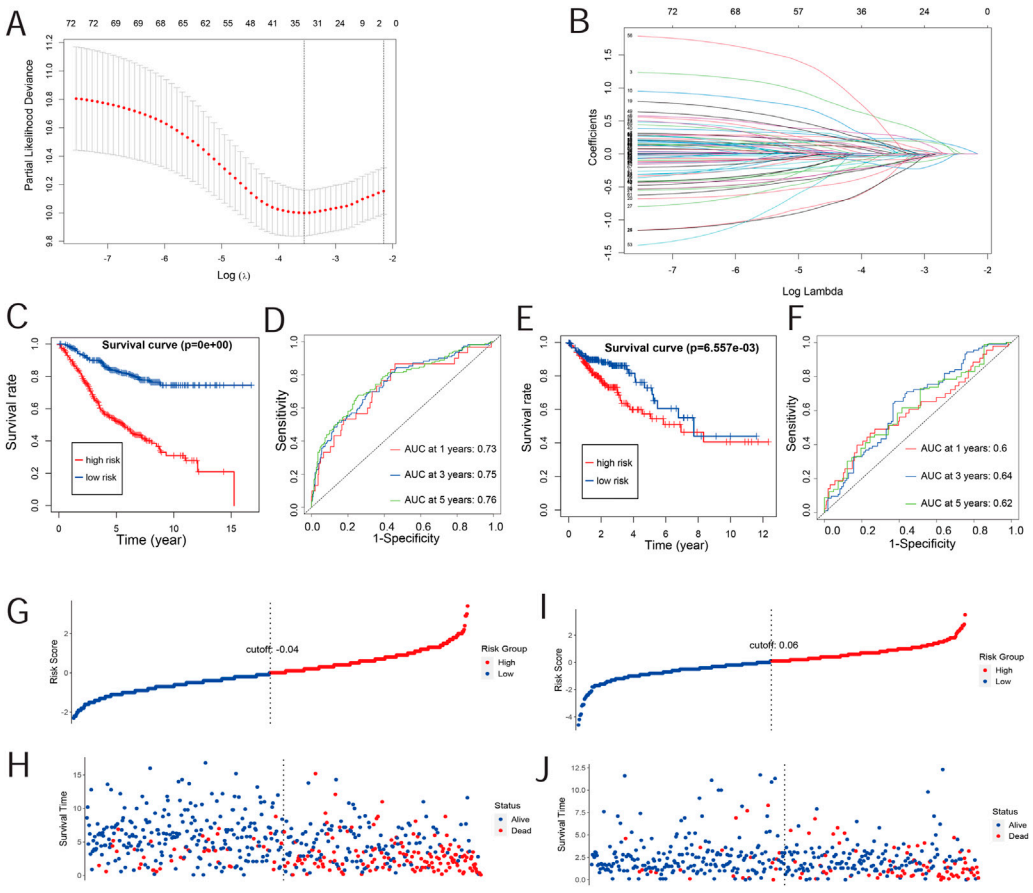


FIGURE 1
Risk-scoring model construction and validation. **(A)** Illustration for the LASSO coefficient spectrum of prognostic genes. **(B)** Adjusted parameters of the LASSO regression model. **(C)** Kaplan–Meier curve analysis of OS in high-risk and low-risk groups based on the GEO database. **(D)** Kaplan–Meier curve analysis of OS in high-risk and low-risk groups based on the TCGA database. **(E)** ROC curve of risk scores in the GEO database. **(F)** ROC curve of risk scores in TCGA database. **(G,H)** Overview of the survival time and distribution of risk scores of the GEO database. **(I,J)** Overview of the survival time and distribution of risk scores of TCGA database.

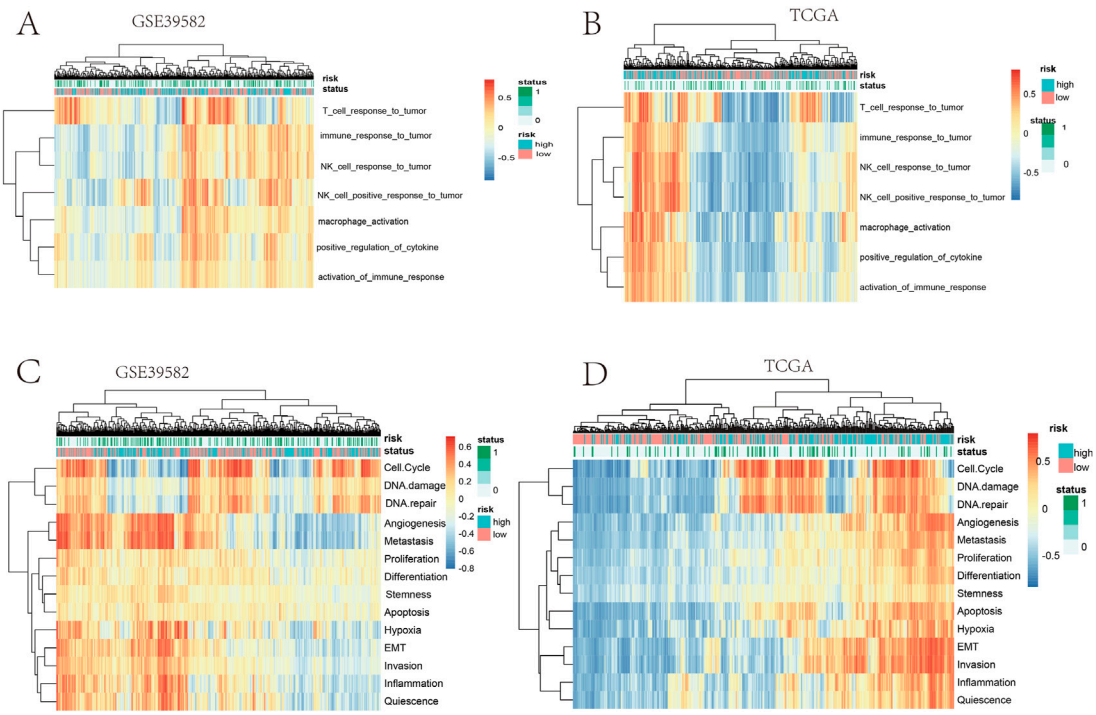


FIGURE 2
(A,B) Heatmap clustering of the risk score and immune signatures in the training cohort. **(C,D)** Heatmap clustering of the risk score and tumor signatures in the validation cohort.

therapy, which received a regulatory approval in 2017, primarily treats severely mutated COAD patients with deficient mismatch repair (dMMR) or high levels of microsatellite instability (MSI-H) (Picard et al., 2020). However, COAD patients, upon receiving adjuvant immunotherapy, may exhibit an immune exclusion response (Fan et al., 2021). Moreover, different chemotherapy drugs elicit variable prognoses for different types of COAD patients. However, choosing a personalized treatment plan still remains challenging and confusing. Hence, the need of the hour is to identify a prognostic model to predict the survival outcomes of COAD patients. The aim was to use this model to clinically guide COAD treatment decisions.

In this study, large data from a cohort of COAD patients from TCGA database were screened for differentially expressed prognostic risk-associated genes. These genes were chosen from the GEO database and verified using TCGA expression data. Herein, we aimed to construct a novel prognostic risk scoring method for COAD that could lead to the administration of a better personalized treatment and management.

Materials and methods

Data collection and preprocessing

The expression profiles were downloaded from two platforms: the GSE39582 dataset from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), and transcriptome profiling (TCGA-COAD-RNAseq) and single-nucleotide variant (TCGA-COAD-SNV) datasets from The Cancer Genome Atlas (TCGA) database (https://www.tcga.org). Single-nucleotide variant (SNV) datasets from TCGA. TCGA-COAD-RNAseq contains 515 samples, including 473 tumor tissue samples and 41 normal solid tissue samples. TCGA-COAD-SNV contains 896 samples, including 448 tumor tissue samples and 448 normal samples. GSE39582 contains 585 samples, containing 566 tumor tissue samples and 19 normal tissue samples. We carried out quantile normalization for expression profiles with the preprocessCore package. Then, we carried out survival analysis and univariate Cox regression analysis for every gene in GSE39582 to obtain the overlap genes as prognostic genes (with the cutoff p-value<0.05) with survival packages. Using the Human Protein Atlas (HPA) database (https://www.proteinatlas.org), by immunohistochemical (IHC) staining, we tested normal intestinal tissue and performed prognosis in COAD organization gene expression differences in the protein level (Asplund et al., 2012).

Construction and external validation of the risk-scoring model

We took GSE39582 as the training dataset to construct a risk-scoring model based on these prognostic genes; the robust prognosis risk-related genes were selected from all prognostic genes *via* a risk score evaluated by the LASSO regression model. To validate the effect of predictive ability of the robust prognosis risk-related genes, the ROC curve was applied to calculate the area under the curve (AUC) on the foundation of the risk score model. The risk-scoring model obtained from GSE39582 was validated by TCGA data *via* Kaplan–Meier survival analysis and ROC curves. Taking the median risk score as the cut-off point, the survival analysis was carried out.

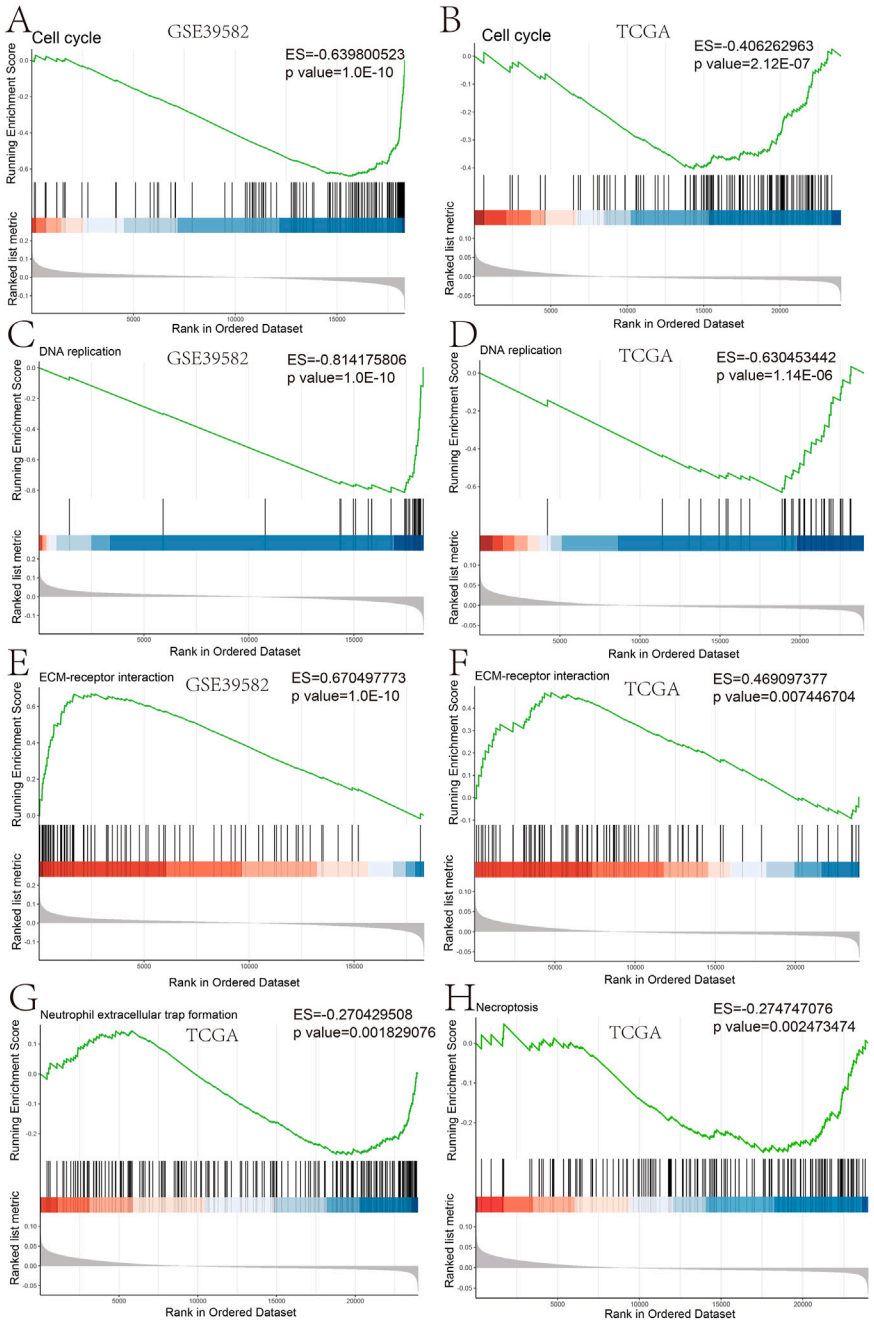


FIGURE 3
GSEA plot of the biological process based on the risk score GSEA analysis in GSE39582 **(A,C,E)** and GSEA analysis in TCGA **(B,D,F–H)**.

Comprehensive analysis about prognostic gene sets

To analyze the biological process based on the risk score group, we carried out Gene Set Enrichment Analysis (GSEA) with the clusterProfiler package. Seven gene sets (GO_ACTIVATION_OF_IMMUNE_RESPONSE, GO_IMMUNE_RESPONSE_TO_TUMOR_CELL, GO_MACROPHAGE_ACTIVATION_INVOLVED_IN_IMMUNE_RESPONSE,GO_NATURAL_KILLER_CELL_MEDIATED_IMMUNE_RESPONSE_TO_TUMOR_CELL, GO_POSITIVE_REGULATION_OF_CYTOKINE_PRODUCTION_INVOLVED_IN_IMMUNE_RESPONSE, GO_POSITIVE_REGULATION_OF_NATURAL_KILLER_CELL_MEDIATED_IMMUNE_RESPONSE_TO_TUMOR_CELL, and GO_T_CELL_MEDIATED_IMMUNE_RESPONSE_TO_TUMOR_CELL) were obtained from GSEA (http://www.gsea-msigdb.org/gsea/index.jsp). Moreover, 14 gene sets (angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, EMT, hypoxia, inflammation, invasion, metastasis, proliferation, quiescence, and stemness) were obtained from CancerSEA (http://biocc.hrbmu.edu.cn/CancerSEA/). We performed a gene set variation analysis about immune signatures and tumor signatures and analyzed the

relationship between the risk score and GSVA score. Also, we analyzed immune infiltration with different tools to know about the status of immune filtration in different risk groups. Combining the clinical information, we explored the difference between high-risk and low-risk groups in the TNM stage and drug reaction. Apart from these, we combined the TCGA-COAD-RNAseq dataset and the TCGA-COAD-SNV dataset to analyze the genetic background behind the two groups with the maftools package.

The clinical value analysis of prognostic gene sets

We investigated the therapeutic value of genes associated with a robust prognostic risk. The tumor immune dysfunction and exclusion (TIDE) algorithm was used to explore the relationship between the risk score and immunotherapy. We also analyzed the relationship between the risk score and chemotherapy. Moreover, we combined the risk score and TNM stage to construct a novel nomogram model with the rms package to improve the model value in predicting prognosis.

Drug sensitivity data analysis

We collected the corresponding mRNA gene expression from the genomics of the Cancer Therapeutics Response Portal (CTRP) and merged the mRNA expression and drug sensitivity data. Pearson correlation analysis was performed to obtain the correlation between mRNA expression and drug IC_{50} values. An FDR-adjusted *p*-value was used in all the analyses (Liu et al., 2018).

Results

Construction and validation of the prognostic model

Survival analyses helped obtain the prognosis-related gene expression profile for COAD patients. Moreover, Cox regression analysis in the GEO dataset, which was verified by TCGA dataset, also helped in the process. The analysis identified a total of 76 prognosis-related genes. In the LASSO regression model, 33 genes were identified as robust prognosis risk-related genes (Figures 1A, B). The 33 genes selected for the model included *ATOH1*, *C4orf47*, *CPA4*, *DNASE1L1*, *ERFE*, *F2RL2*, *FBXO39*, *FZD3*, *HPCAL4*, *ICOS*, *INHBB*, *ITLN1*, *KIF7*, *KLHL26*, *LINC00629*, *LRRC29*, *MMP12*, *MYL6B*, *NPM3*, *PCBD1*, *PLEC*, *POLR2F*, *POU5F1P4*, *PRRX2*, *PTPRU*, *PTTG3P*, *RNF112*, *SERPINB7*, *SLCO1A2*, *TH*, *TMEM39B*, *TRDV3*, and *ZDHHC1*. The detailed characteristics of these prognostic genes in this study are given in Supplementary Table S1. Most prognostic genes were differentially expressed between COAD and normal tissues (Supplementary Figure S1). *ATOH1*, *HPCAL4*, *ITLN1*, *POLR2F*, *RNF112*, *SERPINB7*, *SLCO1A2*, *TH*, and *TMEM39B* were significantly downregulated in COAD tissues ($p < 0.05$) compared to those in normal tissues. *C4orf47*, *CPA4*, *DNASE1L1*, *ERFE*, *F2RL2*, *FBXO39*, *FZD3*, *INHBB*, *KLHL26*, *LRRC29*, *MMP12*, *MYL6B*, *NPM3*, *PCBD1*, *POU5F1P4*, *PRRX2*, *PTPRU*, *PTTG3P*, *TRDV3*, and *ZDHHC1* were significantly upregulated in COAD samples ($p < 0.05$) compared to those in normal tissues. Here, we also used the Human Protein Atlas (HPA) database to validate the expression of these prognostic genes at the protein level (Supplementary Figure S2).

The risk score was evaluated by the coefficient of each of these genes. The formula for the risk score model is as follows: $risk\ score = \sum_{i=1}^{n} coef i^{*} expri$. The median risk score was the cut-off point for dividing the patients. The Kaplan–Meier survival analysis showed that the OS in the low-risk group was significantly higher than that in the high-risk group ($p < 0.0001$, Figure 1C) in the GEO datasets. The ROC curve showed that the AUC values of the 1-, 3-, and 5-year OS were 0.73, 0.75, and 0.76, respectively, in the GEO dataset (Figure 1D). Moreover, TCGA datasets were used for further validation; the OS in the low-risk group was significantly higher than that in the high-risk group ($p < 0.0001$, Figure 1E). The ROC curve indicated that the AUC values of the 1-, 3-, and 5-year OS were 0.6, 0.64, and 0.62, respectively (Figure 1F). Moreover, the low-risk score group has a better outcome of prognosis (Figures 1G, H). Similarly, the risk score group also showed a better outcome (Figures 1I, J) in the validation cohort. The heatmap depicts the expression pattern of prognosis risk-related genes between the high- and low-risk groups in the training and validation cohorts (Supplementary Figure S3). These results collectively indicated that these 33 genes, making up a prognostic gene set, can be used to construct a novel risk model to accurately predict the prognosis of COAD patients.

Roles of the prognostic gene sets in regulating the tumor immune microenvironment and tumor signatures

Furthermore, we studied the relationships between the prognostic gene sets, tumor immune microenvironment, and tumor signatures. The heatmap in Figures 2A, B shows the proportions of tumor-infiltrating natural killer (NK) cells, T cells, neutrophils, and macrophages in the TME. It also indicates that the immune response to the tumor corroborated our prognostic risk score ($p < 0.05$). Moreover, the heatmap in Figures 2C, D shows that the proportions of the cell cycle, DNA damage, DNA repair, angiogenesis, metastasis, proliferation, differentiation, stemness, apoptosis, hypoxia, EMT, invasion, inflammation, and quiescence are significantly related to our prognostic risk score ($p < 0.05$). Also, we carried out a GSEA to analyze the enriched biological processes based on the risk score group. The GSEA showed enrichment of the GO biological processes like cell cycle (ES = −0.406262963; p = 2.12225E-07), DNA replication (ES = −0.630453442; p = 1.14446E-06), ECM–receptor interaction (ES = 0.469097377; p = 0.007446704), neutrophil extracellular trap formation (ES = −0.270429508; p = 0.001829776), and necroptosis (ES = −0.274747076; p = 0.002473474) in TCGA dataset when comparing the high-risk group with the low-risk group (Figure 3). GSEA of the GSE39582 dataset revealed that a higher risk score was closely related to the enrichment of gene sets related to the cell cycle (SE = −0.639800523; p = 1E-10), DNA replication (SE = −0.814175806; p = 1E-10), and ECM–receptor interactions (ES = 0.670497773; p = 1E-10) (Figure 3). The risk score is closely correlated with tumor signatures, including cell cycle, DNA damage, DNA repair, angiogenesis, metastasis, proliferation, differentiation, stemness, apoptosis, hypoxia, EMT, invasion, inflammation, and quiescence (Figure 4). A *p*-value cut-off of <0.05 revealed that the high-risk score group had a higher GSVA score in the aforementioned 14 tumor signatures.

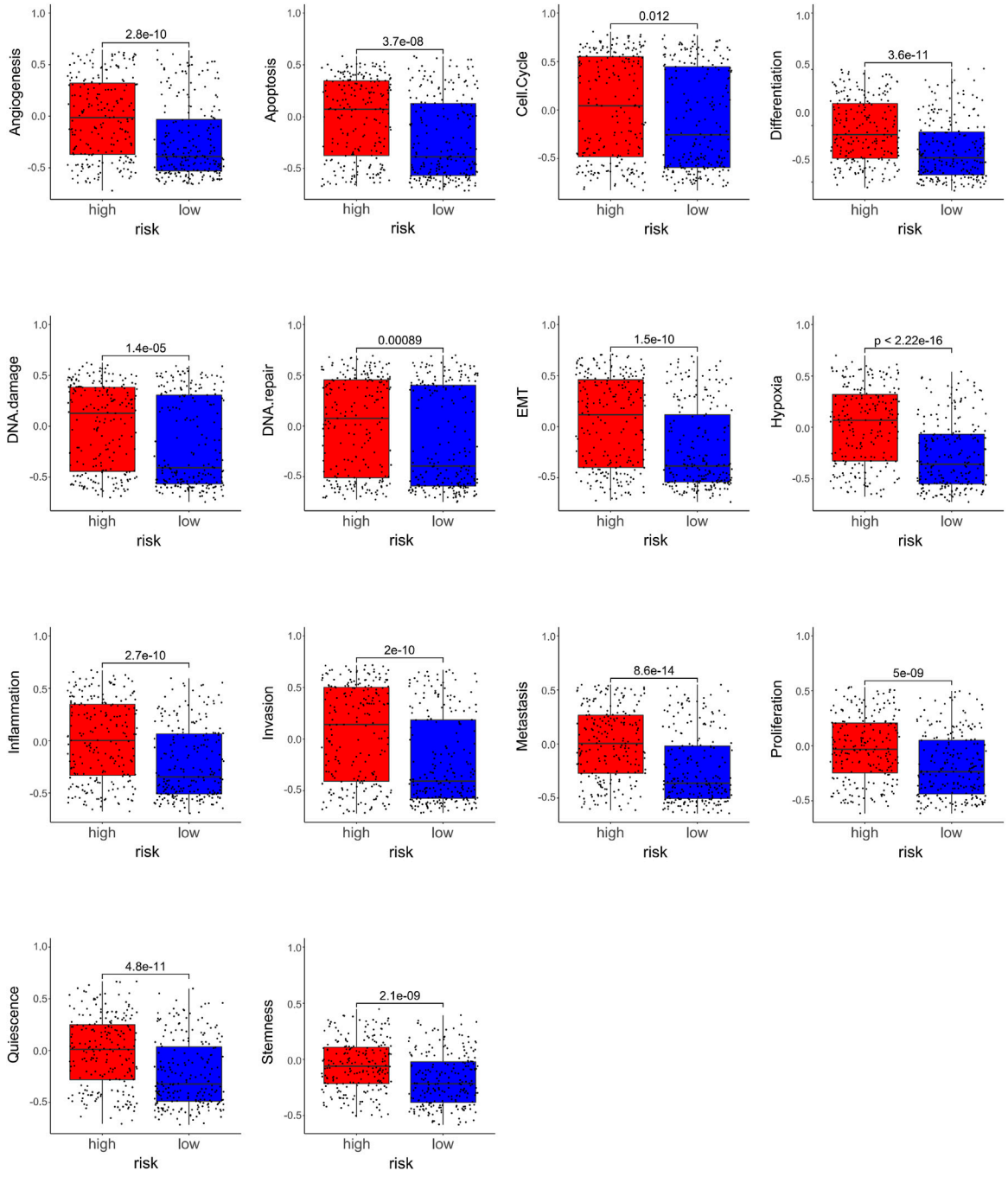


FIGURE 4
Relationships between the compositions of the risk scores and the tumor signatures.

Correlation of a mutation landscape within the prognostic gene sets

We further analyzed the tumor mutational burden between the high- and low-risk groups. Figures 5A, B show how the mutation frequency and mutation spectrum of the mutated genes are higher in the high-risk group. *TTN* was the most significantly mutated gene in the high-risk group, while *APC* was the most significantly mutated gene in the low-risk group. Supplementary Figure S4A shows that the co-occurrence and mutually exclusive mutations were investigated and were observed in the high- and low-risk groups. In the high-risk group, *SYNE1*, *MUC16*, *OBSCN*, and *DNAH5* mutations almost co-occurred with *TTN* mutations ($p < 0.01$), while *ZFHX4* co-mutated with *OBSCN*, *FAT4*, *MUC16*, and *DNAH5* ($p < 0.01$). Moreover, *TP53* and *MUC16* mutations were almost mutually exclusive in the low-risk group ($p < 0.01$), which had a higher tumor mutational burden (TMB) than the low-risk group ($p = 0.041$) (Supplementary Figure S4B). In addition, *BRAF* mutations showed higher scores in the prognostic gene sets than the wild-type mutations ($p = 0.011$) (Supplementary Figure S4C). Moreover, macrophages, NK cells, DC cells, and CD8+T cells were

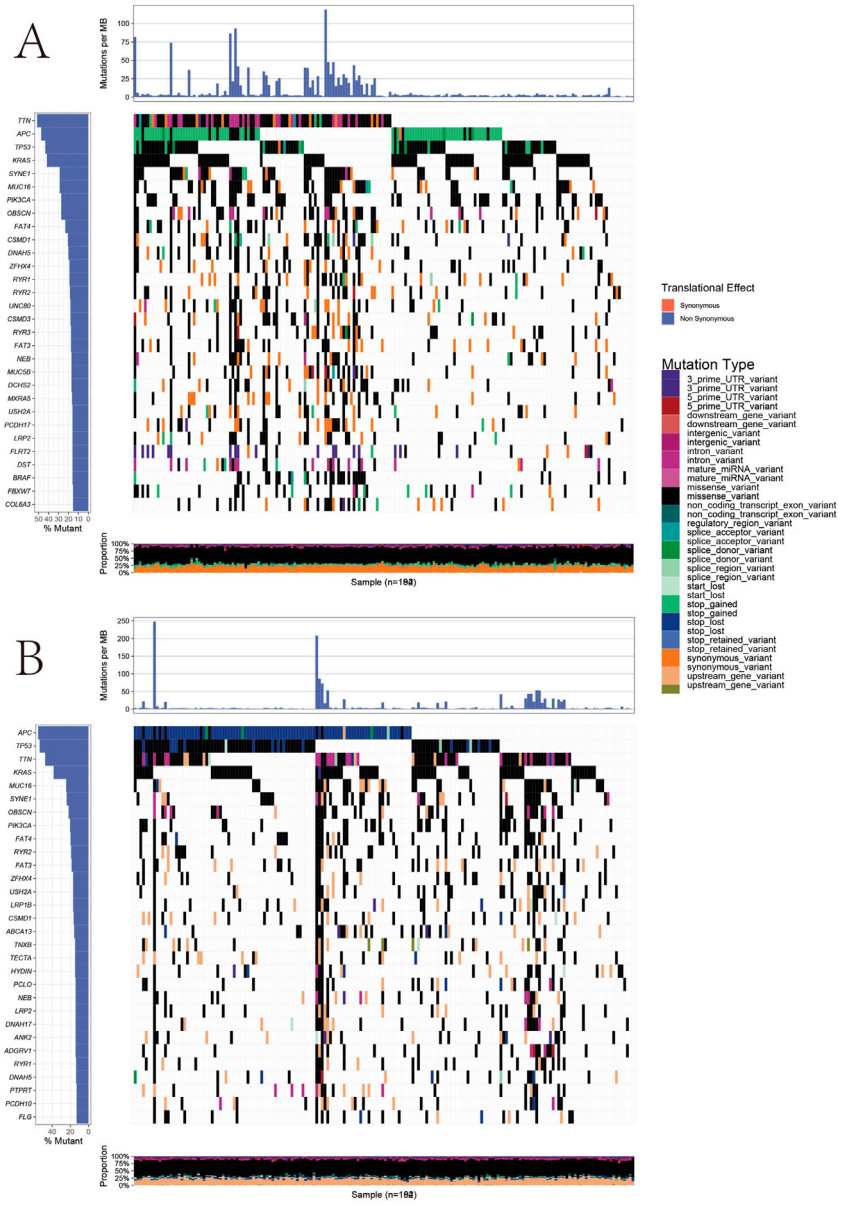


FIGURE 5
Landscape of mutations between high-risk groups and low-risk groups. **(A)** Heatmap illustrates the co-occurrence and mutually exclusive mutations in high-risk groups. **(B)** Heatmap illustrates the co-occurrence and mutually exclusive mutations in low-risk groups.

increased in the mutant type, compared to the wild type, while the natural killer T cells (NKT), neutrophils, and naive $CD8^+$ T cells decreased (Figure 6A). Moreover, the immune cells in the mutant were utterly exhausted. The genome rearrangement-driven copy number variation (CNV) generally refers to an increase or decrease in the copy number of a large genome segment, usually more than 1 KB in length (Lye and Purugganan, 2019). The number of NK cells in the mutant group was significantly reduced compared to the wild-type group (Figure 6B). When the copy number decreases, CD8 T cells, NK cells, and Th1 cells decrease, while NKT cells and CD4 T cells increase.

Furthermore, differences in prognostic gene sets regarding tumor copy number variation and patient prognosis were also investigated. CNVs in *ICOS*, *TRDV3*, *PTPRU*, *PCBD1*, *NPM3*, *LINC00629*, *KLHL26*, *ITLN1*, *INHBB*, *HPCAL4*, *ERFE*, *DNASE1L1*, and *C4orf47* were associated with a disease-free interval (DFI). Patients may change the disease-free survival (DFS) when *PRRX2*, *ZDHHC1*, *PTPRU*, and *LRRC29* have copy number variations. In addition, CNVs in *PRRX2* and *ICOS* change the OS of patients. CNVs in *ICOS*, *ZDHHC1*, *TMEM39B*, *PTPRU*, *POLR2F*, *LRRC29*, *KLHL26*, *HPCAL4*, *F2RL2*, and *ATOH1* genes were associated with a progression-free survival.

Correlation of clinical features with prognostic gene sets

Patients with higher risk scores generally have bigger tumor sizes (T), more tumor nodes (N), and higher tumor node metastasis (M) stages (Figures 7A–C). In TCGA and GSE39582 data, the mean

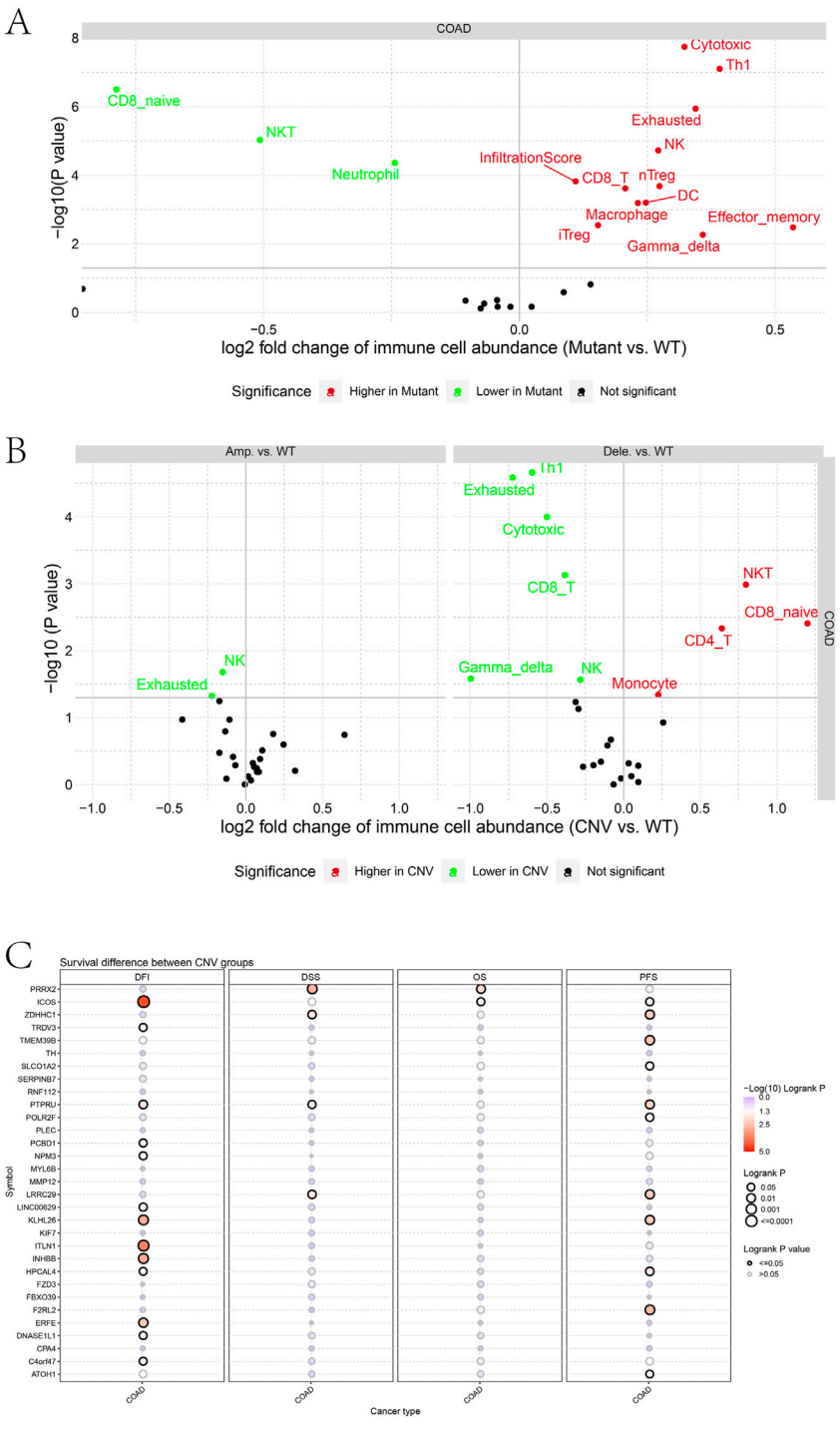


FIGURE 6
(A) Differences in immune cell abundance between mutant and WT groups. **(B)** Differences in immune cell abundance between CNV and WT groups. **(C)** Survival difference between CNV groups.

risk score for patients with T1, T2, T3, and T4 stages increased sequentially (Supplementary Table S2). The risk score is also related to the disease type and tumor location. The risk score is an independent prognostic factor associated with OS ($p = 0.007$), as determined by TCGA analysis (Figure 7D). Moreover, in GSE39582, the risk score also acted as a prognostic indicator of COAD ($p < 0.001$) (Figure 7G). Adenomas and adenocarcinomas had lower risk scores than mucinous and serous neoplasms ($p = 0.0022$) (Figure 7E). The risk score of patients with proximal COAD was higher than distal COAD ($p < 0.05$) (Figure 7F). Together, these results strongly demonstrated the correlation between prognostic gene sets and tumor clinical features.

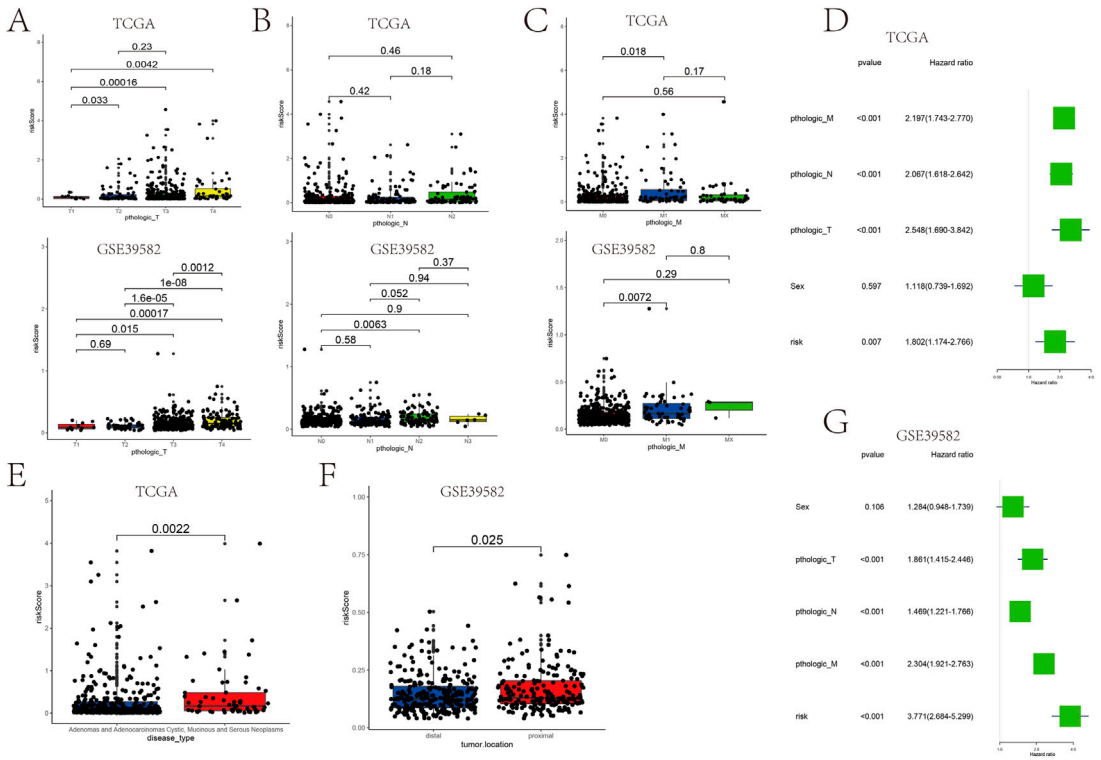


FIGURE 7
Association between the risk score and the patient's clinical characteristics. **(A–C)** Association between the tumor size (T), tumor node (N), tumor metastasis (M), and risk score in TCGA cohort and GSE39582. **(D)** Univariate analysis of factors influencing patient prognosis in TCGA cohort. **(G)** Multifactorial analysis of factors influencing GSE39582 patient prognosis. **(E)** Relationship between the disease type and risk score. **(F)** Relationship between the tumor location and risk score.

The nomogram based on the prognostic gene sets and clinical attributes

Furthermore, a nomogram integrating the genetic risk score (high risk vs. low risk) and TNM stage was constructed to provide quantitative methods to predict a patient's probability of OS to the clinician (Figures 8A, B). The total points were calculated by adding the risk score and TNM-stage points. To evaluate the effect of the nomogram model, we also calculated its C-index. The C-index for the TNM stage with the risk score was higher than that for the TNM stage, indicating that this model is a valuable indicator for prognostic prediction (Figures 8C, E). The calibration curve for predicting a 1-, 3-, and 5-year DFS indicated that the nomogram-predicted survival closely corresponded with actual survival outcomes in GSE39582 (Figures 8D, G, H). In TCGA, the calibration curve for predicting the 1-, 3-, and 5-year DFS indicated that the nomogram-predicted survival closely corresponded with actual survival outcomes (Figures 8F, I, J). These results showed that the prognostic model accurately predicted a patient's OS probability.

Correlation of the prognostic gene sets with adjuvant chemotherapy

The survival time of patients receiving adjuvant chemotherapy was statistically significant in the high-risk group compared to the low-risk group ($p < 0.0001$), as the same in patients without adjuvant chemotherapy (Figure 9A). In addition, adjuvant chemotherapy with 5-fluorouracil, FOLFOX (folinic acid, 5-fluorouracil, and oxaliplatin), FOLFIRI (5-fluorouracil, folinic acid, and irinotecan), or FUFOL (5-fluorouracil and folinic acid) was associated with a better prognosis in both the low-risk groups than in the high-risk group (Figure 9B). The results of multiple comparative analyses of survival curves showed that patients receiving FOLFIRI chemotherapy had the worst prognosis (Figure 9C). Through further analysis, patients receiving FOLFIRI chemotherapy also had the highest risk score (Supplementary Figure S5). These results suggest that the risk score can predict the prognosis of patients treated with chemotherapy.

Correlation of the prognostic gene sets with immunotherapy

The risk score model might reflect the tumor immune microenvironment status in COAD patients, implying that the prognostic gene set also closely correlates with immunotherapy. The risk score positively correlated with the cancer-associated fibroblasts (CAF) ($p < 0.0001$) (Figure 9D). Patients with high-risk scores have a higher probability of immune exclusion than those with low-risk scores ($p = 0.028$) (Figure 9E). Subsequently, we found that the content of neutrophils and macrophages was significantly higher in a high-risk group than that in a low-risk group ($p < 0.05$)

10.3389/fgene.2022.975404

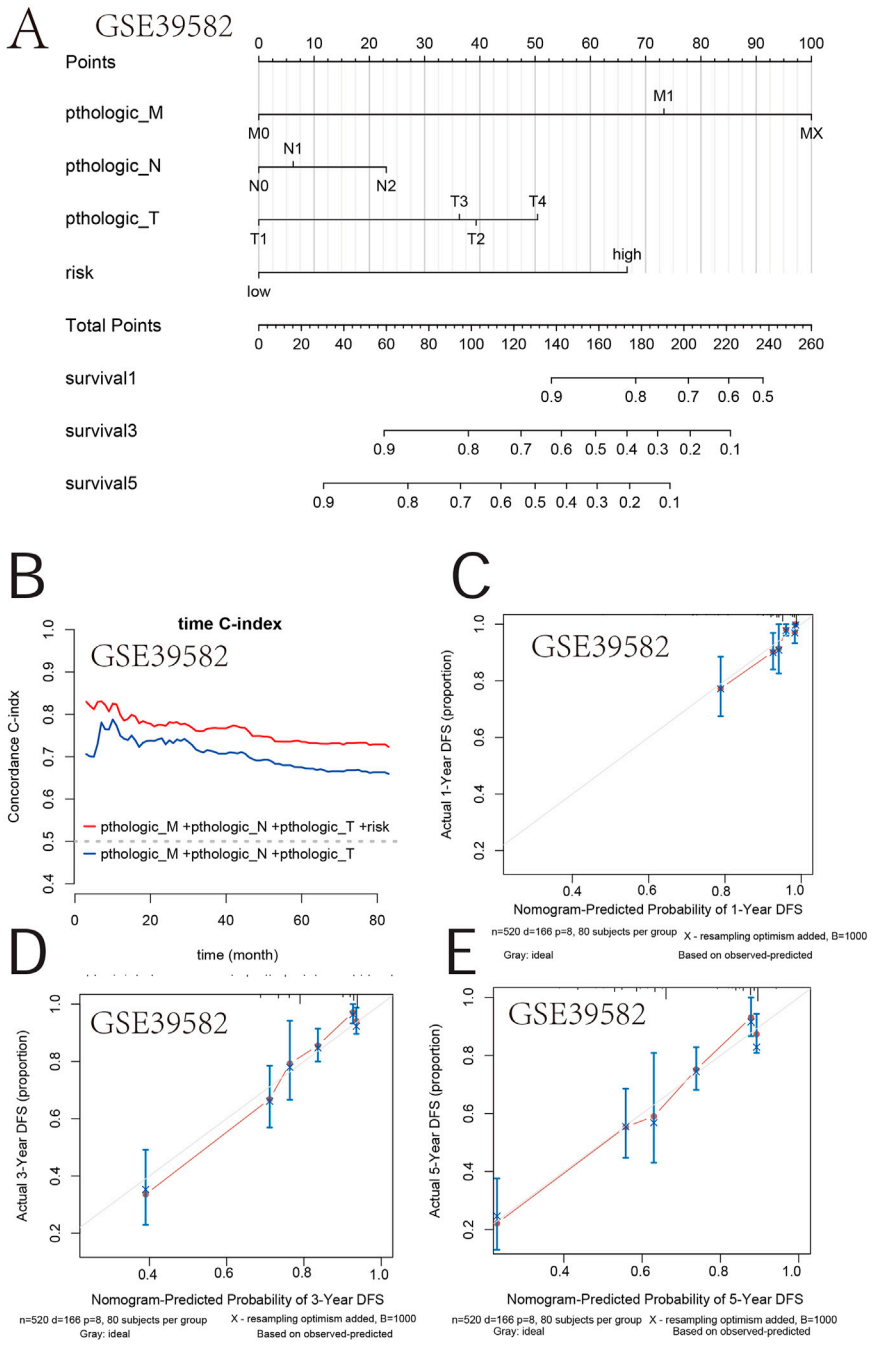


FIGURE 8
Survival nomogram. **(A)** Nomogram for the overall survival was developed in the primary cohort with three prognostic factors: pthologic M, pthologic N, and pthologic T. **(B)** Compared with the TNM and TNM + risk, the novel nomogram exhibited a better powerful capacity for survival prediction. **(C–E)** Nomogram predicting the 1-, 3-, and 5-year overall survival of COAD patients in GSE39582.

(Figure 10). These results collectively suggested that patients with a low-risk score may be better suited to undergo immunotherapy.

Correlation between drug sensitivity and prognostic gene sets

We analyzed the correlation between drug sensitivity and predictive gene sets to further explore the value of prognostic gene sets in a clinical treatment. Most of the genes in the prognostic gene set had correlations between gene expression levels and drug sensitivity (Figure 11). The high expression levels of *POLR2F*, *KLHL26*, *ICOS*, *ITLN1*, *HPCAL4*, *NPM3*, *TMEM39B*, *TH*, *SLCO1A2*, *FZD3*, and *ATOH1* genes were resistant to drugs. The high expression of *PLEC*, *CPA4*, *SERPINB7*, *DNASE1L1*, *KIF7*, *C4orf47*, *F2RL2*, and *PCBD1* genes with an elevated expression was more sensitive to drugs. *PLEC*, *CPA4*, *SERPINB7*, *DNASE1L1*, *C4orf47*, *KIF7*, and *F2RL2* were prognostic genes positively

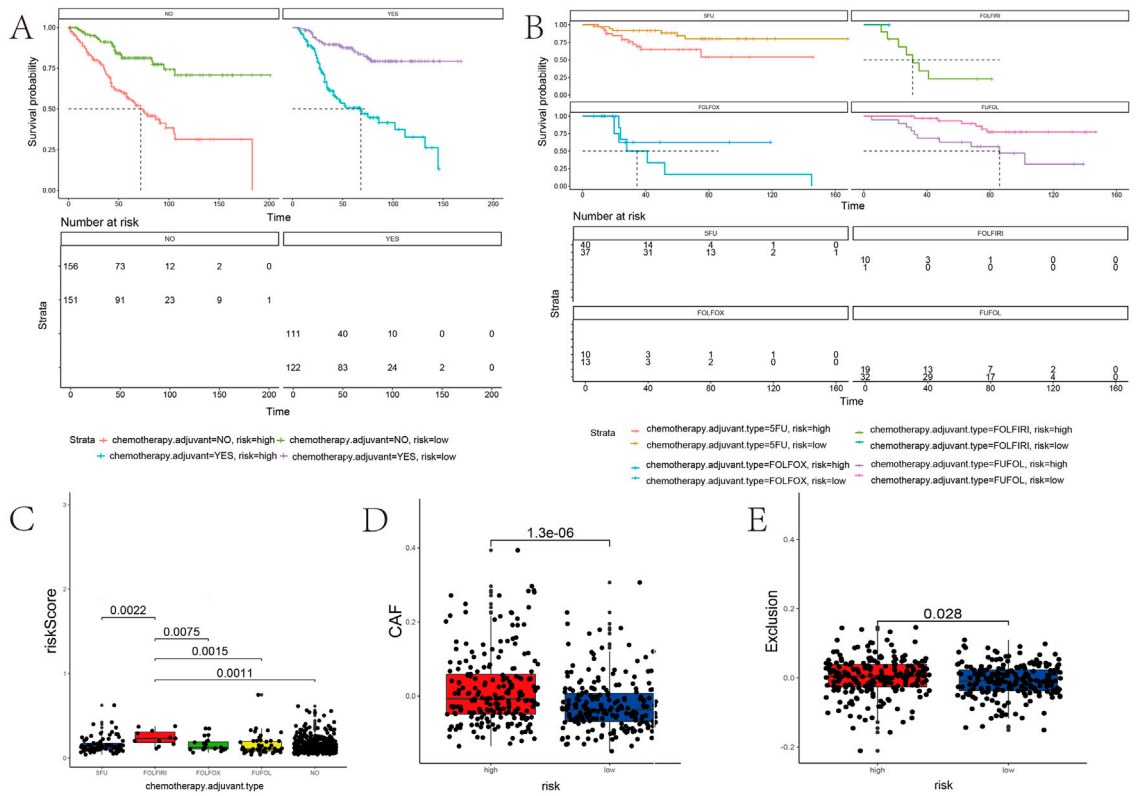


FIGURE 9
Association between the risk score and the adjuvant chemotherapy and immunotherapy response of PD-1. **(A,B)** Kaplan–Meier curves stratified by quartiles of risk scores and the usage of adjuvant chemotherapy. **(C)** Correlation of the risk score with a chemotherapy drug. **(D,E)** Risk score positively correlated with *CAF* and immune exclusion.

associated with the classical antitumor drug fluorouracil. In contrast, *ATOH1*, *FZD3*, *SLCO1A2*, *TH*, *TMEM39B*, and *NPM3* were prognostic genes negatively related to fluorouracil (p <0.001). The expressions of *PLEC*, *CPA4*, and *SERPINB7* positively correlated with belinostat sensitivity (p <0.001). The expressions of *FZD3*, *TMEM39B*, and *NPM3* negatively correlated with narciclasine sensitivity (p <0.001). The higher the expression of *PTPRU*, the lower the drug sensitivity of afatinib and PD153035 is (p <0.001). Thus, the prognostic gene set is a useful clinical tool for guiding drug use.

Discussion

COAD is a common aggressive malignant tumor, with a high mortality rate worldwide (Biller and Schrag, 2021). The etiology and pathology of COAD are highly variable within individuals. For patients with COAD, the current standard treatment includes early surgical resection, following which, patients usually would receive immunotherapy and adjunct chemotherapy, thereby improving the overall survival rate. Even then, there are still many COAD patients who suffer a relapse and would die due to disease recurrence and distant metastasis (Goldstein et al., 2014). So far, the specific underlying molecular pathogenesis of COAD remains largely unclear. Considering COAD's poor prognosis, the need of the hour is to develop a model to predict survival outcomes of COAD patients based on prognosis risk-related gene expression profiling. Currently, COAD patients are diagnosed by the pathophysiological evaluation of prognostic molecular markers (Dekker and Rex, 2018). However, the current biomarkers of COAD are inadequate to predict patients' survival accurately. A single biomarker may not be suitable for the treatment of every patient. Due to individual patient-specific differences, the expression of biomarkers is usually not the same. These biomarkers also fail to predict which patients will benefit from the treatments.

In this study, we used a bioinformatically developed and validated novel prognostic gene set that was significantly associated with OS in COAD patients. A risk score model was also constructed to divide COAD patients into high- and low-risk groups. The Kaplan–Meier survival analysis with the log-rank test and ROC was used to establish the prognostic ability of the model. More importantly, by establishing a validation set, we further verified the reliability of this risk score model.

Moreover, the novel prognostic gene set is closely correlated with pro-tumorigenic signatures, including the cell cycle, DNA damage, DNA repair, angiogenesis, metastasis, proliferation, differentiation, stemness, apoptosis, hypoxia, EMT, invasion, inflammation, and

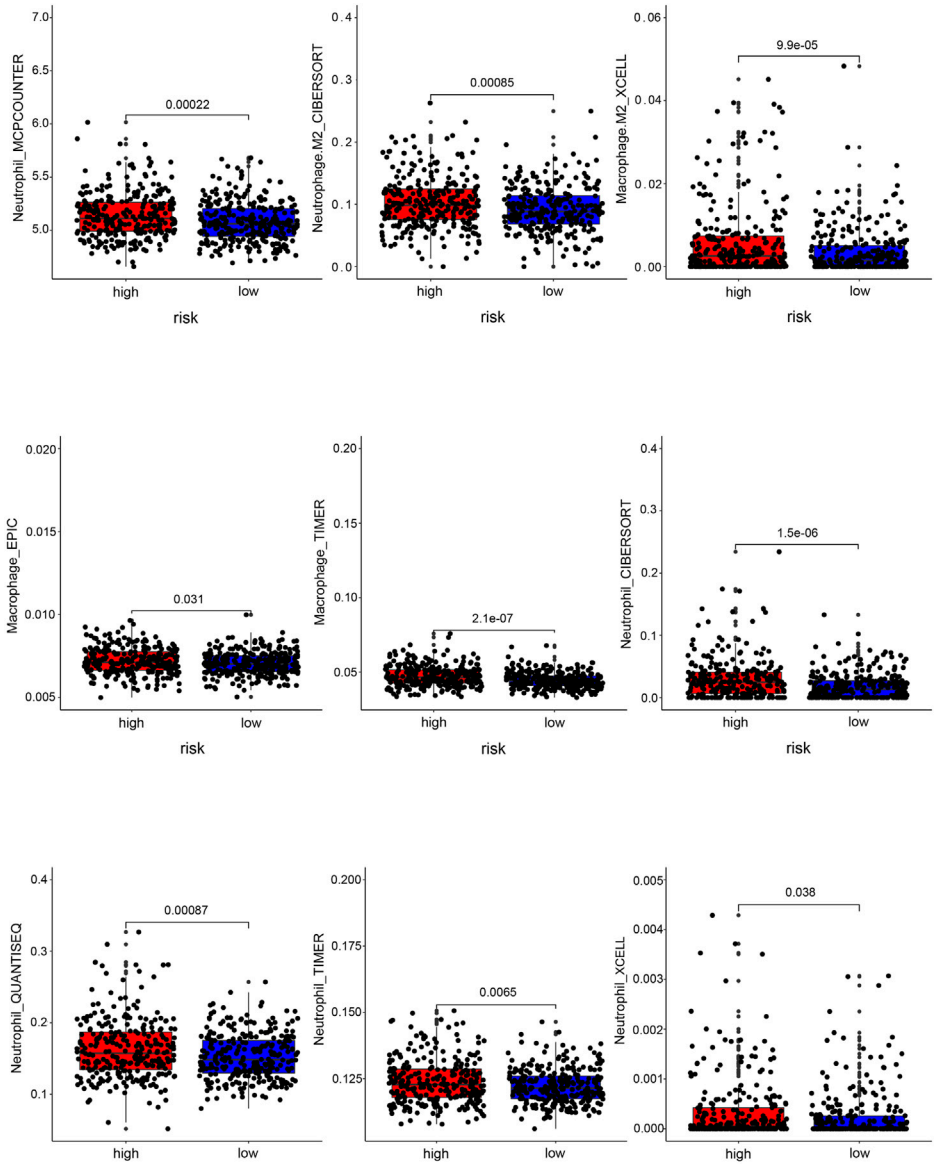


FIGURE 10
Correlation analysis between the risk score and immune cell infiltration. Box plots of immune cells with significant difference in high- and low-risk groups.

quiescence. Many studies showed that these hallmarks of cancer and the immune microenvironment dictate the disease prognosis in COAD. Furthermore, the correlation between the risk score model and gene mutation was also illustrated. The gene mutation probability was significantly higher in a high-risk group than in a low-risk group, which to the best of our knowledge, substantially contributes to cancer progression. Further investigations are necessary to determine the potential functional mechanisms underlying these prognosis risk-related genes. Collectively, our risk score model might be reliable in predicting the prognosis of COAD based on these results.

COAD patients treated with 5-fluorouracil, oxaliplatin, irinotecan, and folinic acid (used sequentially or together upfront) have a better objective response and survival outcome (Wang et al., 2018). 5-Fluorouracil is the most widely used drug and has a low impact on the survival rate (Golfinopoulos et al., 2007). As a result, FOLFOX, FOLFIRI, and FUFOL were in a clinical practice and substantially affected the survival rate (Harada et al., 2019). Higher toxicity renders significant side effects from chemotherapy, warranting careful evaluation before the complete use, limiting it to a small group of patients. Using this scoring model to predict the effect of chemotherapy in COAD patients, the survival advantage in the low-risk group was significant among patients who received chemotherapy. According to our analysis, chemotherapy treatments with FOLFIRI have a higher risk score than 5-fluorouracil, which may deteriorate the survival rate. The scoring model based on prognostic

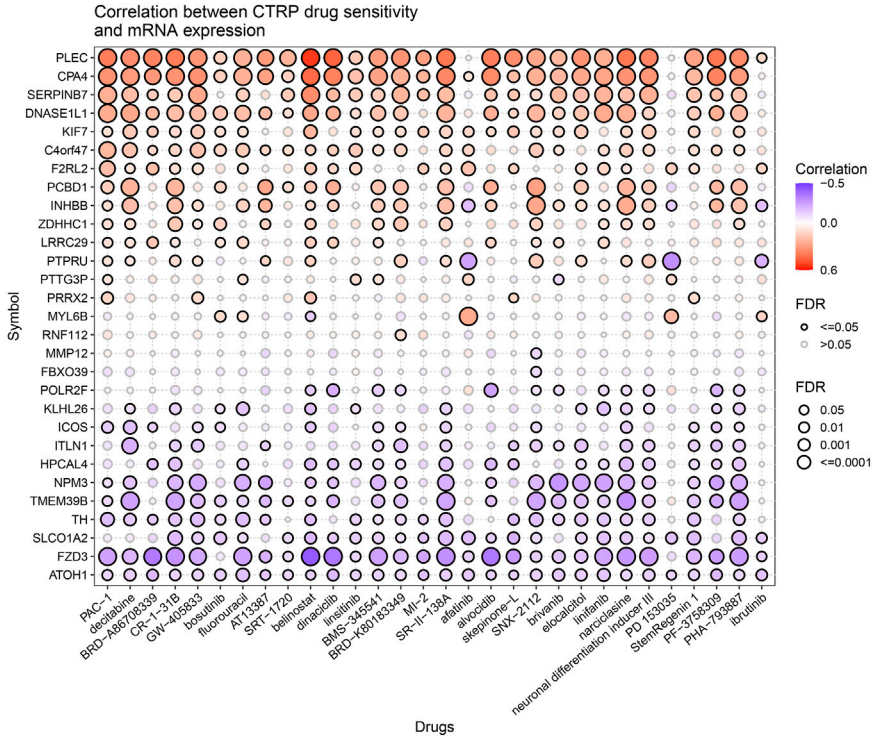


FIGURE 11
Correlation between CTRP drug sensitivity and mRNA expression.

gene sets can efficiently predict the chemotherapy effect of COAD patients.

In addition to adjuvant chemotherapy, immunotherapy is one of the most common treatments for patients with COAD with or without metastasis. The immune microenvironment dictates the efficacy of immune drugs. Hence, patients who use the same therapy during the same phase may have different therapeutic effects. Immune infiltrating cells play essential roles in the progression of COAD (Huang et al., 2019). The prognosis of patients with COAD is mainly related to immunity. Recent studies have demonstrated that a higher density of CD4 naive T cells, regulatory T cells, and M2 macrophages is closely associated with a worse clinical prognosis in many malignant tumors, including COAD (Komohara et al., 2016; Speiser et al., 2016; Labanieh et al., 2018). In contrast, naive B cells, CD8 T cells, and CD4 memory-activated T cells were the protective factors of patients (Yang et al., 2019). Moreover, *CAF* promotes cancer progression by inducing an immunosuppressive tumor microenvironment, rendering resistance to immunotherapy (Miyai et al., 2020; Abuwarwar et al., 2021). Therefore, studying tumor immune infiltration helped analyze the patient's prognosis and develop new cancer diagnosis and treatment methods. The risk score positively correlated with the *CAF* and immune exclusion. Therefore, immunotherapy may be less effective in patients with high-risk scores than in patients with low-risk scores—these risk scores guide immunotherapy decision-making.

Previous studies have also built prognostic models for COAD. Wang et al. developed and validated a novel stem-related prognostic model (AUC = 0.705) for COAD cancer (Wang et al., 2021). Li et al. established a COAD resistance prediction model (AUC = 0.659), which provides therapeutic targets for COAD (Li et al., 2022). The AUC values of our prognostic model at 1, 3, and 5 years in the training set were 0.73, 0.75, and 0.76, respectively, which were all higher than the aforementioned AUC values. In addition, through further verification, it is found that our prognosis model also has a medium accuracy for predicting the survival of COAD patients with 1, 3, and 5 years in the verification set, further confirming the enhanced performance of our prognostic model, hypothesizing that it may likely become a new type of COAD prognostic index. Our study also analyzed the relationship between predictive gene sets and anticancer drug susceptibility, providing novel insights into the search for selecting a more effective anticancer drug therapy and avoiding tumor resistance. Univariate and multivariate independent prognostic analyses showed that the predictive gene set and TNM stage were critical, independent predictors of COAD OS.

Furthermore, we generated a nomogram to quantify the risk assessment and survival probability. Compared to TNM, the nomogram exhibited the highest accuracy and discrimination in OS prediction. In addition, the new prognostic gene set can guide the clinical application of chemotherapy, immunotherapy, and targeted drugs. Therefore, our predictive model may help avoid unnecessary overtreatment of indolent disease and select the best management strategy. However, there are still some limitations to this study. Pure bioinformatics analysis is the main drawback of this study. Second, the interaction between genes in the

prognostic gene set should be investigated better to understand the molecular mechanism of COAD occurrence and progression.

Conclusion

Our study profiled a novel risk score model based on 33 genes for predicting the overall survival in COAD patients. More importantly, the risk score model is significantly associated with the unfavorable clinical outcome of COAD and might monitor its development to provide more effective personalized therapeutic decision-making. A nomogram model might aid in identifying high-risk COAD patients and selecting appropriate clinical follow-up plans accordingly.

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

HZ designed and wrote the manuscript. YW, ZL, ZZ, and LX edited the work. YW reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81970569) and the Natural Science Foundation of Hunan Province (No. 2022JJ40700).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.975404/full#supplementary-material

References

Abuwarwar, M. H., Baker, A. T., Harding, J., Payne, N. L., Nagy, A., Knoblich, K., et al. (2021). *In vitro* suppression of T cell proliferation is a conserved function of primary and immortalized human cancer-associated fibroblasts. *Int. J. Mol. Sci.* 22 (4), 1827. doi:10.3390/ijms22041827

Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., Gershenwald, J. E., Brookland, R. K., et al. (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J. Clin.* 67 (2), 93–99. doi:10.3322/caac.21388

Asplund, A., Edqvist, P. H. D., Schwenk, J. M., and Ponten, F. (2012). Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. *Proteomics* 12 (13), 2067–2077. doi:10.1002/pmic.201100504

Biller, L. H., and Schrag, D. (2021). Diagnosis and treatment of metastatic colorectal cancer: A review. *Jama* 325 (7), 669–685. doi:10.1001/jama.2021.0106

Dekker, E., and Rex, D. K. (2018). Advances in CRC prevention: Screening and surveillance. *Gastroenterology* 154 (7), 1970–1984. doi:10.1053/j.gastro.2018.01.069

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal cancer. *Lancet* 394 (10207), 1467–1480. doi:10.1016/S0140-6736(19)32319-0

Fan, A., Wang, B., Wang, X., Nie, Y., Fan, D., Zhao, X., et al. (2021). Immunotherapy in colorectal cancer: Current achievements and future perspective. *Int. J. Biol. Sci.* 17 (14), 3837–3849. doi:10.7150/ijbs.64077

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in colorectal cancer: Rationale, challenges and potential. *Nat. Rev. Gastroenterol. Hepatol.* 16 (6), 361–375. doi:10.1038/s41575-019-0126-x

Goldstein, J., Tran, B., Ensor, J., Gibbs, P., Wong, H. L., Wong, S. F., et al. (2014). Multicenter retrospective analysis of metastatic colorectal cancer (CRC) with high-level microsatellite instability (MSI-H). *Ann. Oncol.* 25 (5), 1032–1038. doi:10.1093/annonc/mdu100

Golfinopoulos, V., Salanti, G., Pavlidis, N., and Ioannidis, J. P. A. (2007). Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: A meta-analysis. *Lancet Oncol.* 8 (10), 898–911. doi:10.1016/S1470-2045(07)70281-4

Gu, L., Liu, Y., Jiang, C., Sun, L., and Zhou, H. (2020). Identification and clinical validation of metastasis-associated biomarkers based on large-scale samples in colon-adenocarcinoma. *Pharmacol. Res.* 160, 105087. doi:10.1016/j.phrs.2020.105087

Harada, K., Okamoto, W., Mimaki, S., Kawamoto, Y., Bando, H., Yamashita, R., et al. (2019). Comparative sequence analysis of patient-matched primary colorectal cancer, metastatic, and recurrent metastatic tumors after adjuvant FOLFOX chemotherapy. *BMC Cancer* 19 (1), 255. doi:10.1186/s12885-019-5479-6

Huang, R., Zeng, Z., Li, G., Song, D., Yan, P., Yin, H., et al. (2019). The construction and comprehensive analysis of ceRNA networks and tumor-infiltrating immune cells in bone metastatic melanoma. *Front. Genet.* 10, 828. doi:10.3389/fgene.2019.00828

Keum, N., and Giovannucci, E. (2019). Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* 16 (12), 713–732. doi:10.1038/s41575-019-0189-8

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

Labanieh, L., Majzner, R. G., and Mackall, C. L. (2018). Programming CAR-T cells to kill cancer. *Nat. Biomed. Eng.* 2 (6), 377–391. doi:10.1038/s41551-018-0235-9

Li, Z., Chen, J., Zhu, D., Wang, X., Chen, J., Zhang, Y., et al. (2022). Identification of prognostic stemness biomarkers in colon adenocarcinoma drug resistance. *BMC Genom Data* 23 (1), 51. doi:10.1186/s12863-022-01063-9

Liu, C. J., Hu, F. F., Xia, M. X., Han, L., Zhang, Q., and Guo, A. Y. (2018). GSCALite: A web server for gene set cancer analysis. *Bioinformatics* 34 (21), 3771–3772. doi:10.1093/bioinformatics/bty411

Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. *Trends Plant Sci.* 24 (4), 352–365. doi:10.1016/j.tplants.2019.01.003

Miyai, Y., Esaki, N., Takahashi, M., and Enomoto, A. (2020). Cancer-associated fibroblasts that restrain cancer progression: Hypotheses and perspectives. *Cancer Sci.* 111 (4), 1047–1057. doi:10.1111/cas.14346

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

Speiser, D. E., Ho, P. C., and Verdeil, G. (2016). Regulatory circuits of T cell function in cancer. *Nat. Rev. Immunol.* 16 (10), 599–611. doi:10.1038/nri.2016.80

Wang, J., Luo, L., Wang, D., Guo, B., Li, J., Yang, Z., et al. (2018). Combination adjuvant chemotherapy with targeted drugs for treatment of colorectal cancer: A network meta-analysis. *J. Cell Biochem.* 119 (2), 1521–1537. doi:10.1002/jcb.26312

Wang, W., Xu, C., Ren, Y., Wang, S., Liao, C., Fu, X., et al. (2021). A novel cancer stemness-related signature for predicting prognosis in patients with colon adenocarcinoma. *Stem Cells Int.* 2021, 7036059. doi:10.1155/2021/7036059

Yang, S., Liu, T., Cheng, Y., Bai, Y., and Liang, G. (2019). Immune cell infiltration as a biomarker for the diagnosis and prognosis of digestive system cancer. *Cancer Sci.* 110 (12), 3639–3649. doi:10.1111/cas.14216

Yoshino, T., ArnolD, D., TanigucHi, H., Pentheroudakis, G., YamazaKi, K., Xu, R. H., et al. (2018). Pan-asian adapted ESMO consensus guidelines for the management of patients with metastatic colorectal cancer: A JSMO-ESMO initiative endorsed by CSCO, KACO, MOS, SSO and TOS. *Ann. Oncol.* 29 (1), 44–70. doi:10.1093/annonc/mdx738

Zhou, F., Wang, L., Jin, K., and Wu, Y. (2021). RecQ-like helicase 4 (RECQL4) exacerbates resistance to oxaliplatin in colon adenocarcinoma via activation of the PI3K/AKT signaling pathway. *Bioengineered* 12 (1), 5859–5869. doi:10.1080/21655979.2021.1964156

Zhou, H., Liu, Z., Wang, Y., Wen, X., Amador, E. H., Yuan, L., et al. (2022). Colorectal liver metastasis: Molecular mechanism and interventional therapy. *Signal Transduct. Target Ther.* 7 (1), 70. doi:10.1038/s41392-022-00922-2

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fgene.2023.1060504

OPEN ACCESS

EDITED BY
Pasquale Pisapia,
University of Naples Federico II, Italy

REVIEWED BY
Laura Cortesi,
University Hospital of Modena, Italy
Richarda Maria De Voer,
Radboud University Medical Centre, Netherlands

*CORRESPONDENCE
Rossella Ferrante,
rossellaferrante@yahoo.it

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

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

RECEIVED 03 October 2022
ACCEPTED 04 January 2023
PUBLISHED 01 February 2023

CITATION
Anaclerio F, Pilenzi L, Dell'Elice A, Ferrante R, Grossi S, Ferlito LM, Marinelli C, Gildetti S, Calabrese G, Stuppia L and Antonucci I (2023), Clinical usefulness of NGS multi-gene panel testing in hereditary cancer analysis.
Front. Genet. 14:1060504.
doi: 10.3389/fgene.2023.1060504

Clinical usefulness of NGS multi-gene panel testing in hereditary cancer analysis

Federico Anaclerio[1†], Lucrezia Pilenzi[1†], Anastasia Dell'Elice[1], Rossella Ferrante[1*], Simona Grossi[2], Luca Maria Ferlito[2], Camilla Marinelli[2], Simona Gildetti[2], Giuseppe Calabrese[3], Liborio Stuppia[1] and Ivana Antonucci[1]

[1]Center for Advanced Studies and Technology (CAST), G.d'Annunzio University of Chieti-Pescara, Chieti, Italy, [2]Eusoma Breast Centre, "G. Bernabeo" Hospital, Ortona, Italy, [3]UOSD Genetica Oncoematologica, Dipartimento di Oncologico-Ematologico, Pescara, Italy

Introduction: A considerable number of families with pedigrees suggestive of a Mendelian form of Breast Cancer (BC), Ovarian Cancer (OC), or Pancreatic Cancer (PC) do not show detectable *BRCA*1/2 mutations after genetic testing. The use of multi-gene hereditary cancer panels increases the possibility to identify individuals with cancer predisposing gene variants. Our study was aimed to evaluate the increase in the detection rate of pathogenic mutations in BC, OC, and PC patients when using a multi-gene panel.

Methods: 546 patients affected by BC (423), PC (64), or OC (59) entered the study from January 2020 to December 2021. For BC patients, inclusion criteria were i) positive cancer family background, ii) early onset, and iii) triple negative BC. PC patients were enrolled when affected by metastatic cancer, while OC patients were all submitted to genetic testing without selection. The patients were tested using a Next-Generation Sequencing (NGS) panel containing 25 genes in addition to *BRCA*1/2.

Results: Forty-four out of 546 patients (8%) carried germline pathogenic/likely pathogenic variants (PV/LPV) on *BRCA*1/2 genes, and 46 (8%) presented PV or LPV in other susceptibility genes.

Discussion: Our findings demonstrate the utility of expanded panel testing in patients with suspected hereditary cancer syndromes, since this approach increased the mutation detection rate of 15% in PC, 8% in BC and 5% in OC cases. In absence of multi-gene panel analysis, a considerable percentage of mutations would have been lost.

KEYWORDS

NGS, hereditary cancer, BRCA, cancer predisposition gene, multi-gene panel testing, breast cancer, ovarian cancer, pancreatic cancer

1 Introduction

In these years of personalized medicine, the study of individual's genotype is an important part of the determination of his specific susceptibility to several diseases, including cancer. The National Comprehensive Cancer Network Breast Cancer Risk Panel (NCCN) has for years been updating with publishing the indications for genetic testing of cancer patients and their family members (Sorscher 2019). The majority of Breast Cancer (BC), Ovarian Cancer (OC) and Pancreatic Cancer (PC) cases are sporadic (75%–80%), ~15%–20% are considered familial types and 5%–10% are hereditary (Russo et al., 2009; Antonucci et al., 2017b; Incorvaia et al., 2020). Over the past

20–30 years, molecular diagnosis of hereditary BC, OC or PC has focused primarily on two high-penetrance genes, *BRCA*1 and *BRCA*2 (Antonucci et al., 2017a). The identification of germline deleterious variants in *BRCA*1/*BRCA*2 has a significant impact on clinical management of both affected individuals and their family members (Babore et al., 2019; Lombardi et al., 2019; 2022). Nevertheless, an increasing number of families with pedigrees suggestive of a Mendelian form of BC, OC or PC have not detectable mutations in *BRCA*1/*BRCA*2. The problem of "missing heritability" can be explained with the presence of pathogenic gene variants in other susceptibility genes involved with low frequency or with reduced penetrance, usually not included in the diagnostic flowchart of patients with hereditary cancer, mainly due to the costs and the time required for the analysis in the Sanger sequencing era. Therefore, it has become mandatory to study many genes in a brief time and in an economic way. In this scenario, advances in genetic technology and implementation of NGS in clinical oncology have accelerated the discovery of new cancer-related genes revolutionizing cancer research, diagnosis and therapies (Rossi et al., 2022). The advent of NGS allows the simultaneous sequencing of multiple samples and genes (Fountzilas et al., 2018). Because of the advantage from cost-benefit reduction, this approach provides a powerful enforcement for patients with LPVs and PVs in other genes, beyond *BRCA*1/2. Several germline PVs in susceptibility genes as *CDH*1, *PALB*2, *PTEN*, STK11, *TP*53, *ATM*, *CHEK*2, *BARD*1, *BRLP*1, *RAD*51*C*, and *RAD*51*D* (Shah et al., 2016; Fanale et al., 2020) can be associated with hereditary tumors. Most of these genes are involved in cell cycle checkpoint and DNA damage repair mechanism, and function together in these physiological pathways (Nielsen et al., 2016; Piombino et al., 2020; Neiger et al., 2021); therefore, a fundamental comprehension of the disease drivers in the cascades would facilitate the accurate evaluation of the genetic risk of cancer development (Yoshimura et al., 2022). In our study we used a multi-gene panel including 27 genes in the diagnostic iter of 546 patients with BC, OC or PC (Table 1). The aims of this work were: 1) to investigate the prevalence of PVs or LPVs in susceptibility genes implicated in hereditary cancer predisposition, and 2) to assess the utility of carrying out a multi-gene panel testing in BC, OC or PC individuals who fulfill specific criteria on their familiar and personal history of tumor.

TABLE 1 Multi-gene panel including the 27 genes analyzed with NGS.

Multi-gene panel—next generation sequencing	
ATM	*BARD*1
*BRCA*1	*BRCA*2
*BRIP*1	*CDK*12
*CHEK*2	*NBN*
*PALB*2	*TP*53
EPCAM	*RAD*51*C*
*RAD*51*D*	*MSH*2
APC	*CDH*1
*CDKN*2*A*	*MKH*1
*MSH*6	*NF*1
*PMS*2	*PTEN*
*CDK*4	*MUTYH*
*POLD*1	*POLE*
*SMAD*4	

2 Materials and methods

2.1 Study population

Our study includes a cohort of individuals who referred to our Center between January 2020 and December 2021. We collected and analyzed DNA samples from 546 patients with BC (423), PC (64) or OC (59), averaging 54 years (range 25–70). For BC patients, inclusion criteria were 1) positive cancer family background, 2) early onset and 3) triple negative BC. PC patients were enrolled when affected by metastatic cancer, while OC patients were all submitted to genetic testing without selection. PC and OC patients were classified into 2 groups related to the age of disease onset: 1) early onset cancer (age at diagnosis ≤45 years) and 2) late onset cancer (age at diagnosis >45 years), while for BC patients the considered age of onset was 40 years. Among BC patients, 64 had early onset cancer and 359 had late onset cancer; among OC patients 9 had early onset cancer and 50 had late onset cancer; among PC patients 2 had early onset cancer and 62 had late onset cancer. Starting from 423 BC patients, 27 (6.4%) had triple-negative breast cancer (TNBC), including 25 patients with late onset BC and only 2 with early onset BC. Genetic counseling was performed in the presence of a geneticist and a psychologist to acquire the clinical personal and familiar history of patients. In addition, data about histological cancer type, any surgical operations and current therapies were acquired. All subjects signed an informed consent about the significance of the molecular genetic test.

2.2 Next-generation sequencing (NGS)

Genomic DNA of BC, OC and PC patients were collected using buccal swabs and extracted through MagPurix instrument and Forensic DNA Extraction Kit (Zinexts Life Science Corp.- CodZP01001) according to the manufacturer's protocol. NGS was executed by the Ion Torrent S5 system (Thermo Fisher Scientific, Waltham, MA, United States) after automatic library preparation using Ion Chef (Thermo Fisher Scientific, Waltham, MA, United States). Ion Chef consists of fragmentation and adapter ligation onto the PCR products, clonal amplification. The DNA libraries were quantified with Real-Time Step One PCR System (Thermo Fisher Scientific, Waltham, MA, United States) and the prepared samples were loaded onto an Ion 530™ chip by Ion Chef (Thermo Fisher Scientific, Waltham, MA, United States). Ion S5™ Plus (Thermo Fisher Scientific, Waltham, MA, United States) instrument was used for the sequencing. Specific plugins as "SampleId" and "Coverageanalysis" were used for NGS data analysis on the Torrent Suite 5.14.0 platform. The uniformity of base coverage was over 98% in all batches, and base coverage was over ×20 at all target regions. This NGS method cannot detect variations outside the +/−10 nucleotide coding sequence.

2.3 Sanger sequencing

Sanger Sequencing was performed using SeqStudio Genetic Analyzer System (Thermo Fisher Scientific) and BigDye

TABLE 2 All single PVs/LPVs recurrent in patients analyzed by multi-gene panel. All variants reported in the Table 2 are in heterozygous, except only one subject that had two PVs/LPVs on MUTYH gene (*).

Gene	Refseq	Omim	HGVS Nomenclature	Protein change	Variant interpretation	Number of patients
ATM	NM_000051.3	607585	c.2502dup	p.(Val835fs)	PV	1 (2.1)
ATM	NM_000051.3	607585	c.2113del	p.(Tyr705fs)	PV	1 (2.1)
ATM	NM_000051.3	607585	c.756_757del	p.(Cys252_Glu253delinsTer)	LPV/PV	1 (2.1)
ATM	NM_000051.3	607585	c.6095G > A	p.(Arg2032Lys)	LPV/PV	1 (2.1)
*BRIP*1	NM_03204.2	605882	c.2111T > A	p.(Leu704Ter)	PV	1 (2.1)
*CHEK*1	NM_03204.2	605882	c.2392C > T	p.(Arg798Ter)	PV	1 (2.1)
*CHEK*2	NM_007194.3	604373	c.1232G > A	p.(Trp411Ter)	LPV/PV	1 (2.1)
*CHEK*2	NM_007194.3	604373	c.1100del	p.(Thr367fs)	PV	2 (4.3)
*CHEK*2	NM_007194.3	604373	c.1427C > T	p.(Thr476Met)	PV	1 (2.1)
*CHEK*2	NM_007194.3	604373	c.349A > G	p.(Arg117Gly)	LPV/PV	2 (2.1)
*CHEK*2	NM_007194.3	604373	c.409C > T	p.(Arg137Ter)	PV	1 (2.1)
*CHEK*2	NM_007194.3	604373	c.470T > C	p.(Ile157Thr)	LPV	2 (4.3)
*CHEK*2	NM_007194.3	604373	c.499G > A	p.(Gly167Arg)	LPV/PV	2 (4.3)
*MSH*2	NM_000251.2	609309	c.2647dup	p.(Ile883fs)	PV	1 (2.1)
MUTYH	NM_001128425.2	608456	c.1187G > A	p.(Gly396Asp))	PV	7 (15.2)
MUTYH	NM_001128425.2	608456	c.1437_1439del	p.(Glu480del)	PV	1 (2.1)
MUTYH	NM_001128425.2	608456	c.536A > G	p.(Tyr179Cys)	PV	4 (8.7)
MUTYH	NM_001128425.1	608456	c.1012C > T	p.(Gln338Ter)	LPV	1 (2.1)
MUTYH	NM_001128425.2	608456	c.734G>A (*)	p.(Arg245His)	PV	3 (10.9)
MUTYH	NM_001128425.2	608456	c.884C>T (*)	p.(Pro295Leu)	PV	1 (2.1)
NBN	NM_002485.4	6,026,667	c.741_742dup	p.(Glu248fs)	PV	1 (2.1)
NBN	NM_002485.4	6026667	c.2140C > T	p.(Arg714Ter)	PV	1 (2.1)
*PALB*2	NM_024675.3	610355	c.661_662delGTinsTA	p.(Val221Ter)	PV	1 (2.1)
*PALB*2	NM_024675.3	610355	c.1050_1053del	p.(Thr351fs)	PV	1 (2.1)
POLE	NM_006231.3	174762	c.1458delC	p.(Met487fs)	LPV	1 (2.1)
*RAD*51*C*	NM_058216.2	602774	c.1026 + 5_1026 + 7del	-	PV/LPV	1 (2.1)
*RAD*51*C*	NM_058216.2	602774	c.905-2_905-1del	-	PV	2 (4.3)
*RAD*51*D*	NM_002878.3	602954	c.898C > T	p.(Arg300Ter)	LPV/PV	1 (2.1)
*TP*53	NM_000546.5	191170	c.646G > A	p.(Val216Met)	LPV/PV	1 (2.1)
*TP*53	NM_000546.5	191170	c.637C > T	p.(Arg213Ter)	PV	1 (2.1)
*TP*53	NM_000546.5	191170	c.993G > A	p.(Gln331Gln)	PV	1 (2.1)

Terminator 3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) to confirm all the PV/LPVs identified with NGS multi-gene panel.

2.4 Classification of the genetic variants

The genetic variants found in patients were classified into five classes: benign (C1), likely benign (C2), variant of uncertain significance (VUS, C3), likely pathogenic (C4), and pathogenic (C5), according to the guidelines of Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) (https://enigmaconsortium.org/). We focused on the LPVs and PVs that can be used for clinical purposes. Variants were referred in according to the nomenclature recommendations of the Human Genome Variation Society (https://www.hgvs.org). The clinical significance of the genetic variants found in this study was evaluated according to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), Varsome (https://varsome.com), Franklin Genoox (https://franklin.genoox.com) and, for some

10.3389/fgene.2023.1060504

TABLE 3 Different groups analyzed by the age of onset criteria.

Type of tumor	BC		OC		PC	
AGE OF ONSET	≤40	>40	≤45	>45	≤45	>45
BRCA PVs/LPVs	9	23	2	7	0	3
PM PVs/LPVs	7	26	1	2	1	9
TOTAL PVs/LPVs	16	49	3	9	1	12

susceptibility genes (*APC*, *MLH*1, *MSH*2, *MSH*6, *PMS*2, *EPCAM*, *MUTYH*, *CDH*1), according to LOVD-InSIGHT (https://www.insight-group.org/variants/databases/).

3 Results

In our study, 546 cases with BC, OC, or PC were enrolled from January 2020 to December 2021. PVs or LPVs on *BRCA*1/2 genes were detected in 44 patients (8%), specifically 32/423 (7%) with BC, 9/59 (15%) with OC and 3/64 (5%) with PC. On the other hand, 46 patients (8%), namely 33/423 (8%) with BC, 3/59 (5%) with OC and 10/64 (16%) with PC harbored germline PVs/LPVs in other cancer susceptibility genes, as follows: 17 (37%) in *MUTYH*, 11 (24%) in *CHEK*2, 4 (9%) in *ATM*, 3 (6%) in *RAD*51*C* and *TP*53, 2 (4%) *PALB*2, *BRIP*1, and *NBN*. In addition, a single PV in *POLE*, *MSH*2, and *RAD*51*D* was detected in two patients (Table 2).

Seven subjects enrolled showed two pathogenic variants in the genes analyzed.

According to age of onset, we found PVs/LPVs in 20 early onset patients (≤45 for PC and OC, ≤40 for BC) and in 70 late onset patients (>45 for PC and OC, >40 for BC). Eleven early onset patients with BC (14%) had PVs or LPVs mutations in *BRCA*1 or *BRCA*2 genes, whereas 17 patients (11%) reported mutations in one of the other genes included in the multi-gene panel. On the other hand, 27 late onset patients with BC (36%) had PVs or LPVs mutations in *BRCA*or *BRCA*2 genes, whereas 30 patients (40%) reported mutations in one of the several genes included in the multi-gene panel. On the OC and PC patients groups 2 early onset subjects (18%) had a PV or LPV in *BRCA*1/2, while 2 patients (18%) had PV or LPV in other gene. In the late onset group 10 patients (9%) had a PV or LPV in *BRCA*1/2 and 11(10%) with pathogenic variant in other gene. The distribution of PVs/LPVs in *BRCA*1/2 or in other genes in the different groups of patients is reported in Table 3.

MUTYH resulted as the gene with the higher percentage of mutation within the group analyzed by the multigene panel (16 out of 46 detected mutations), with the second most recurrent involved genes represented by *CHEK*2 with 11 cases (Table 2; Figure 1; Figures 2A, B). All *MUTYH* variants reported in this study are in heterozygous, except only one subject that had two PVs/LPVs on *MUTYH* gene, respectively c.734G>A and c.884C>T.

As to PVs/LPVs, the most frequent PV was c.1187G>A p.(Gly396Asp) of *MUTYH* gene, located in coding exon 13 and causing the substitution of a Glycine with Aspartate in codon position 396. This alteration, found in seven patients (15.2%) with BC, PC and OC, is frequently reported as founder mutation in multiple populations. M. Nielsen et al. have shown that this missense variant change the function of *MUTYH* protein (M. Nielsen et al., 2009).

The second recurrent PV found on the *MUTYH* gene was c.734G>A p. (Arg245His), in coding exon 9, results from the substitution of a Guanine to Adenine, and consequently the replacement of the arginine with a histidine at codon 245. Literature's data supports that this missense variant has a deleterious effect on protein structure/function (Viel et al., 2017). We found this mutation in five patients (10.9%).

Analyzing the second most gene mutated, *CHEK*2, the other most recurrent PVs were: c.1100delC p. (Thr367fs), c.470T>C p. (Ile157Thr) and c.499G>A p. (Gly167Arg). All subjects with *CHEK*2 variant, are carriers of only one PV/LPV.

Specifically, *CHEK*2 c.1100delC caused deletes of one Cytosine from exon 11 in position 1,100 causing a frameshift at codon 367, and

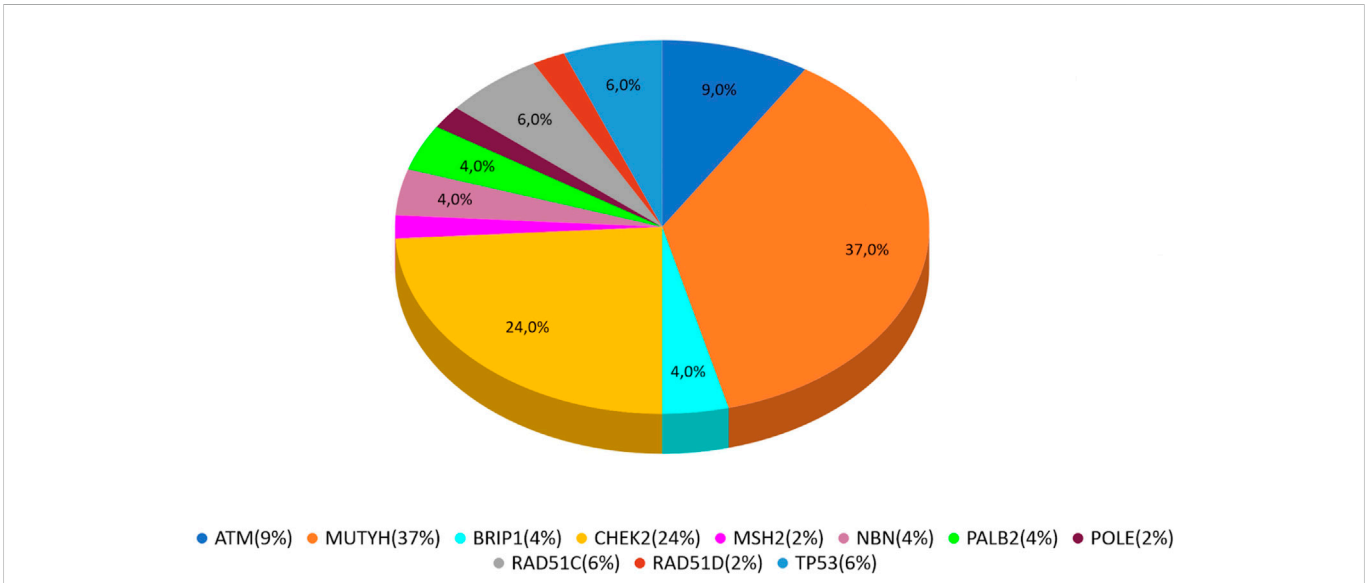


FIGURE 1
All cases analyzed with multi-gene panel.

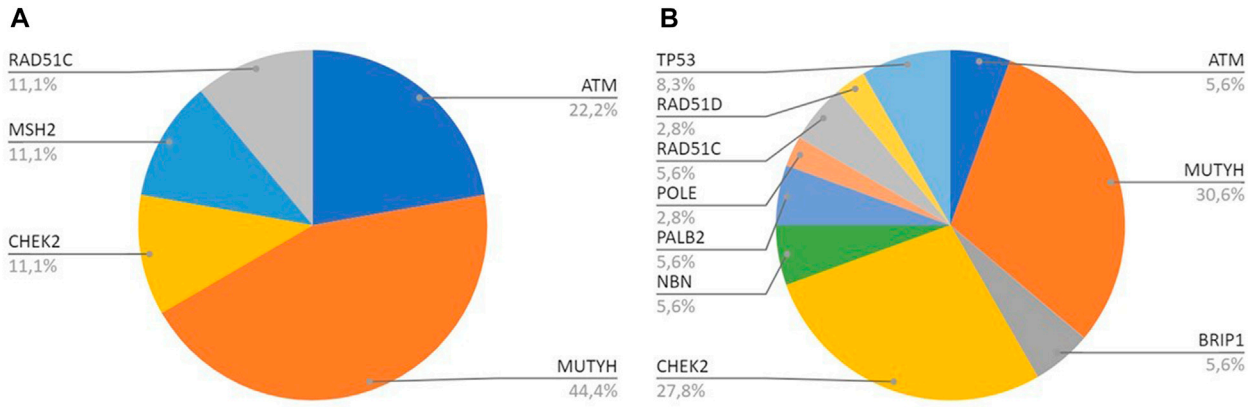


FIGURE 2
(A) All genes with PV/LPV in Breast Cancer cases; **(B)** All genes with PV/LPV in Pancreatic Cancer cases.

a premature translational stop signal p. (Thr367fs). This is expected to result in an absent or disrupted protein product. (Weischer et al., 2008). This variant is linked to increased risk of BC and OC.

4 Discussion

The identification of PV or LPV in genes responsible for hereditary cancers plays a key role in the prognosis, prevention and therapy of these conditions. In fact, cancer patients carriers of such gene variants must undergo specific protocols for the prevention of additional cancers but they can also benefit from specific drug therapies, such as those based on PARP inhibitors (PARPi), which represent a successful example of precision medicine (Slade, 2020). On the other hand, unaffected family members of a cancer patient carrier of a PV/LPV should be tested for the presence of the same variant and, when positive, specific prevention protocols, different from the common cancer screening programs used for the general populations, should be offered. In this view, a critical issue is represented by the number of genes to analyze in each condition, mainly in order to maintain a balanced cost/benefit ratio. While in a first moment it was suggested that each different type of cancer was related to one or a few specific genes (e.g., *BRCA*1/2 for BC and OC, *APC* for familial adenomatous polyposis, etc.), our study revealed that often there is not correspondence between tumor type and the associated mutated gene, raising the question about the need for more genes to be analyzed in hereditary cancers. Interestingly, our study showed that 94% of *MUTYH* carriers had a heterozygous variant. PVs/LPVs in *MUTYH* are associated with colorectal adenomatous polyposis autosomal recessive, while recent literature data revealed the association between monoallelic *MUTYH* variants and several type of cancer (Dell'Elice et al., 2021). BC, PC and OC, together with colon and prostate cancer, are the major tumors linked to clinical familiar history, as well as the major BRCA-associated cancers (Daly et al., 2021). Nevertheless, many of these patients result negative to the genetic testing for *BRCA*1/2 genes PVs and LPVs, even in presence of an evident familiar and/or personal cancer's background. This has been confirmed by data obtained in the present study, showing that no more that 8% of BC, OC or PC cancer show *BRCA*1/2 mutations even in the group of early onset cases.; that the use of multi-gene hereditary cancer panels increases the possibility to identify individuals with cancer predisposing gene variants (Shin et al., 2020; Hu et al., 2021). In an association analysis involving 113,000 women, the Breast Cancer Association Consortium, Dorling L, Carvalho S, et al. define the susceptibility genes that are most clinically useful for inclusion on panels for the prediction of breast cancer risk (Breast Cancer Association Consortium et al., 2021). By extending the test using a multi-gene panel, we found an additional 8% mutations in different susceptibility genes, such as *MUTYH*, *CHEK*2, *ATM*, *NBN*, *BRIP*1, and *TP*53 involved in several hereditary cancer syndromes (Desmond et al., 2015; Tsaousis et al., 2019; N; Tung et al., 2015). These results confirmed the studies already performed in 2021 by Bono et al., where a considerable percentage of PVs/LPVs have been lost without the use of multi-gene panel (Bono et al., 2021). Thus, our results evidenced that both in early and late onset cancer patients, using the classical approach of *BRCA*1/2 testing, we would have lost a large number of cases resulted *BRCA*1/2 negative, but actually carriers of a PV/LPV in other genes. In addition to the increased detection rate, the use of multigenic panel test allow the identification of specific prevention strategies based on the gene involved, in a precision medicine approach. For example, we diagnosed three patients with Li-Fraumeni syndrome (LFS) associated with PV/LPV in *TP*53 on chromosome 17p13.1. This syndrome represents a severe condition inherited in an autosomal dominant manner with very high penetrance. Prevention strategies of this condition are different from the one used for *BRCA*1/2 mutation carriers, since LFS component tumors also include soft tissue sarcomas, osteosarcoma, brain tumors, and adrenocortical carcinomas. Interestingly, in these patients no strong familiar history was found, but they all showed early onset cancer (≤35). In one case, a "*de novo*" origin of the mutation was demonstrated, allowing to suggest that the age of onset of the disease could be considered as a more reliable indicator of the presence of a genetic condition than the familiarity itself. Oncology therapy putting forth the concept of selective targeting of cancer cells thanks to precision medicine. According to our goal, one of the most interesting future perspectives is the therapy with poly-adenosine diphosphate-ribose polymerase (PARP). PARP inhibitors (PARPi) were a significant example of precision medicine (Slade, 2020). The identification of specific mutations in genes different from *BRCA*1/2 is relevant also for the therapeutical strategies. In fact, while the benefits of PARP inhibition have been well characterized for *BRCA*1/2 (Risdon et al., 2021), the efficiency of this

therapy in carriers of other mutations is so far a question of debate. For the therapy of metastatic breast cancer (MBC), is in progress a phase II study that are showing the efficacy of PARPi's Olaparib, in patients with germline/somatic (g/s) mutations in related genes (*PALB*2, *ATM* and *CHEK*2) other than *BRCA*1/2 (N. M. Tung et al., 2020). Responses were seen only with g*PALB*2 mutations, while there are not evidences for *ATM* or *CHEK*2 mutations respectively. For this reason, Olaparib could be used in patients with g*PALB*2 mutation beyond in g*BRCA*1/2 mutation carriers, significantly expanding the number of patients with MBC who would benefit from PARPi (Pommier et al., 2016; Lord and Ashworth., 2017; Cortesi et al., 2021). In conclusion, the multi-gene panel approach could be useful for targeting therapy in oncology patients that are carriers of mutations in susceptibility genes, beyond *BRCA*1/2.

Data availability statement

The datasets presented in this study can be found in online repositories. The link to the data can be found below: https://www.ncbi.nlm.nih.gov/sra/PRJNA927294. Accession to cite for these SRA data: PRJNA927294.

Author contributions

Writing—original draft preparation, FA and LP, writing—review and editing FA, LP, AD, and RF; performed the genetic analysis, FA, LP, AD, and RF; data curation RF; conceptualization and supervision, LS and IA; clinical investigation SGr, LMF, CM, SGi, GC, LS, and IA.

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

References

Antonucci, I., Provenzano, M., Sorino, L., Balsamo, M., Battista, P., Euhus, D., et al. (2017a). Comparison between CaGene 5.1 and 6.0 for BRCA1/2 mutation prediction: A retrospective study of 150 BRCA1/2 genetic tests in 517 families with breast/ovarian cancer. *J. Hum. Genet.* 62 (3), 379–387. doi:10.1038/jhg.2016.138

Antonucci, I., Provenzano, M., Sorino, L., Rodrigues, M., Palka, G., and Stuppia, L. (2017b). A new case of "de Novo" BRCA1 mutation in a patient with early-onset breast cancer. *Clin. Case Rep.* 5 (3), 238–240. doi:10.1002/ccr3.718

Babore, A., Bramanti, S. M., Lombardi, L., Stuppia, L., Trumello, C., Antonucci, I., et al. (2019). The role of depression and emotion regulation on parenting stress in a sample of mothers with cancer. *Support Care Cancer* 27 (4), 1271–1277. doi:10.1007/s00520-018-4611-5

Bono, M., Fanale, D., Incorvaia, L., Cancelliere, D., Fiorino, A., Calò, V., et al. (2021). Impact of deleterious variants in other genes beyond BRCA1/2 detected in breast/ovarian and pancreatic cancer patients by NGS-based multi-gene panel testing: Looking over the hedge. *ESMO Open* 6 (4), 100235. doi:10.1016/j.esmoop.2021.100235

Breast Cancer Association Consortium, Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Craig, L., Wahlström, C., et al. (2021). Breast cancer risk genes - association analysis in more than 113,000 women. *N. Engl. J. Med.* 384 (5), 428–439. doi:10.1056/NEJMoa1913948

Cortesi, L., Piombino, C., and Toss, A. (2021). Germline mutations in other homologous recombination repair-related genes than BRCA1/2: Predictive or prognostic factors? *J. Personalized Med.* 11 (4), 245. doi:10.3390/jpm11040245

Daly, M. B., Tuya Pal, M. P., Buys, S. S., Dickson, P., Domchek, S. M., Ahmed, E., et al. (2021). Genetic/familial high-risk assessment: Breast, ovarian, and pancreatic, version 2.2021, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Cancer Netw. JNCCN* 19 (1), 77–102. doi:10.6004/jnccn.2021.0001

Dell'Elice, A., Cini, G., Fornasarig, M., Franco, A., Barana, D., Bianchi, F., et al. (2021). Filling the gap: A thorough investigation for the genetic diagnosis of unsolved polyposis patients with monoallelic MUTYH pathogenic variants. *Mol. Genet. Genomic Med.* 9 (12), e1831. doi:10.1002/mgg3.1831

Desmond, A., Kurian, A. W., Gabree, M., Mills, M. A., Anderson, M. J., Kobayashi, Y., et al. (2015). Clinical actionability of multigene panel testing for hereditary breast and ovarian cancer risk assessment. *JAMA Oncol.* 1 (7), 943–951. doi:10.1001/jamaoncol.2015.2690

Fanale, D., Incorvaia, L., Filorizzo, C., Bono, M., Fiorino, A., Calò, V., et al. (2020). Detection of germline mutations in a cohort of 139 patients with bilateral breast cancer by multi-gene panel testing: Impact of pathogenic variants in other genes beyond BRCA1/2. *Cancers* 12 (9), E2415. doi:10.3390/cancers12092415

Fountzilas, C., and Kaklamani, V. G. (2018). Multi-gene panel testing in breast cancer management. *Cancer Treat. Res.* 173, 121–140. doi:10.1007/978-3-319-70197-4_8

Hu, C., Hart, S. N., Gnanaolivu, R., Huang, H., Lee, K. Y., Na, J., et al. (2021). A population-based study of genes previously implicated in breast cancer. *N. Engl. J. Med.* 384 (5), 440–451. doi:10.1056/NEJMoa2005936

Incorvaia, L., Fanale, D., Bono, M., Calò, V., Cancelliere, D., Castiglia, M., et al. (2020). Hereditary breast and ovarian cancer in families from southern Italy (Sicily)-Prevalence and geographic distribution of pathogenic variants in BRCA1/2 genes. *Cancers* 12 (5), E1158. doi:10.3390/cancers12051158

Lombardi, L., Trumello, C., Stuppia, L., Antonucci, I., Brandão, T., and Babore, A. (2022). BRCA1/2 pathogenetic variant carriers and reproductive decisions: Gender differences and factors associated with the choice of preimplantation genetic diagnosis (PGD) and prenatal diagnosis (PND). *J. Assisted Reproduction Genet.* 39 (7), 1433–1443. doi:10.1007/s10815-022-02523-y

Lombardi, L., Bramanti, S. M., Babore, A., Stuppia, L., Trumello, C., Antonucci, I., et al. (2019). Psychological aspects, risk and protective factors related to BRCA genetic testing: A review of the literature. *Support Care Cancer* 27 (10), 3647–3656. doi:10.1007/s00520-019-04918-7

Lord, C. J., and Ashworth., A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Sci. (New York, N.Y.)* 355 (6330), 1152–1158. doi:10.1126/science.aam7344

Neiger, H. E., Siegler, E. L., and Shi, Y. (2021). Breast cancer predisposition genes and synthetic lethality. *Int. J. Mol. Sci.* 22 (11), 5614. doi:10.3390/ijms22115614

Nielsen, F. C., Hansen, T. O., and Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: New genes in confined pathways. *Nat. Rev. Cancer* 16 (9), 599–612. doi:10.1038/nrc.2016.72

Nielsen, M., Jones, N., Vogt, S., Carli, M., Vasen, H. F. A., Sampson, J. R., et al. (2009). Analysis of MUTYH genotypes and colorectal phenotypes in patients with MUTYH-associated polyposis. *Gastroenterology* 136 (2), 471–476. doi:10.1053/j.gastro.2008.10.056

Piombino, C., Cortesi, L., Lambertini, M., Kevin, P., Grandi, G., and Toss, A. (2020). Secondary prevention in hereditary breast and/or ovarian cancer syndromes other than BRCA. *J. Oncol.* 2020, 6384190. doi:10.1155/2020/6384190

Pommier, Y., O'Connor, M. J., and de Bono, J. (2016). Laying a trap to kill cancer cells: PARP inhibitors and their mechanisms of action. *Sci. Transl. Med.* 8 (362), 362ps17. doi:10.1126/scitranslmed.aaf9246

Risdon, E. N., Chau, C. H., Price, D. K., Sartor, O., and Figg, W. D. (2021). PARP inhibitors and prostate cancer: To infinity and beyond BRCA. *Oncol.* 26 (1), e115–e129. doi:10.1634/theoncologist.2020-0697

Rossi, C., Cicalini, I., Cufaro, M. C., Ada, C., Upadhyaya, P., Sala, G., et al. (2022). Breast cancer in the era of integrating "omics" approaches. *Oncogenesis* 11 (1), 17. doi:10.1038/s41389-022-00393-8

Russo, A., Calò, V., Bruno, L., Schirò, V., Agnese, V., Cascio, S., et al. (2009). Is *BRCA*1-5083del19, Identified in Breast Cancer Patients of Sicilian Origin, a Calabrian Founder Mutation? *Breast Cancer Res. Treat.* 113 (1), 67–70. doi:10.1007/s10549-008-9906-7

Shah, P. D., Patil, S., Dickler, M. N., Offit, K., Hudis, C. A., and Robson, M. E. (2016). Twenty-one-gene recurrence score assay in BRCA-associated versus sporadic breast cancers: Differences based on germline mutation status. *Cancer* 122 (8), 1178–1184. doi:10.1002/cncr.29903

Shin, H-C., Lee, H-B., Yoo, T-K., Lee, E-S., Kim, R. N., Park, B., et al. (2020). Detection of germline mutations in breast cancer patients with clinical features of hereditary cancer syndrome using a multi-gene panel test. *Cancer Res. Treat.* 52 (3), 697–713. doi:10.4143/crt.2019.559

Slade, D. (2020). PARP and PARG inhibitors in cancer treatment. *Genes & Dev.* 34 (5–6), 360–394. doi:10.1101/gad.334516.119

Sorscher, S. (2019). Universal multigene panel testing in all breast cancer patients. *Am. J. Med.* 132 (11), e765–e766. doi:10.1016/j.amjmed.2019.03.012

Tsaousis, G.N., Papadopoulou, E., Apessos, A., Agiannitopoulos, K., Pepe, G., Kampouri, S., et al. (2019). Analysis of hereditary cancer syndromes by using a panel of genes: Novel and multiple pathogenic mutations. *BMC Cancer* 19 (1), 535. doi:10.1186/s12885-019-5756-4

Tung, N., Battelli, C., Allen, B., Kaldate, R., Bhatnagar, S., Bowles, K., et al. (2015). Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer* 121 (1), 25–33. doi:10.1002/cncr.29010

Tung, N. M., Robson, M. E., Ventz, S., Santa-Maria, C. A., Nanda, R., Marcom, P. K., et al. (2020). Tbcrc 048: Phase II study of Olaparib for metastatic breast cancer and mutations in homologous recombination-related genes. *J. Clin. Oncol.* 38 (36), 4274–4282. doi:10.1200/JCO.20.02151

Viel, A., Bruselles, A., Meccia, E., Fornasarig, M., Quaia, M., Canzonieri, V., et al. (2017). A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. *EBioMedicine* 20, 39–49. doi:10.1016/j.ebiom.2017.04.022

Weischer, M., Bojesen, S. E., Ellervik, C., Tybjaerg-Hansen, A., and Nordestgaard, B. G. (2008). CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J. Clin. Oncol.* 26 (4), 542–548. doi:10.1200/JCO.2007.12.5922

Yoshimura, A., Imoto, I., and Iwata, H. (2022). Functions of breast cancer predisposition genes: Implications for clinical management. *Int. J. Mol. Sci.* 23 (13), 7481. doi:10.3390/ijms23137481

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 06 April 2023
DOI 10.3389/fgene.2023.1029758

OPEN ACCESS

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

REVIEWED BY
Zhuan Liao,
First Affiliated Hospital of Naval Medical University, China
Shengjie Dai,
First Affiliated Hospital of Wenzhou Medical University, China

*CORRESPONDENCE
Seyed Mohammad Tavangar,
tavangar@ams.ac.ir

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

RECEIVED 27 August 2022
ACCEPTED 10 March 2023
PUBLISHED 06 April 2023

CITATION
Tamaddon M, Azimzadeh M, Gifani P and Tavangar SM (2023), Single-cell transcriptome analysis for cancer and biology of the pancreas: A review on recent progress.
Front. Genet. 14:1029758.
doi: 10.3389/fgene.2023.1029758

Single-cell transcriptome analysis for cancer and biology of the pancreas: A review on recent progress

Mona Tamaddon[1], Mostafa Azimzadeh[2,3,4], Peyman Gifani[5,6] and Seyed Mohammad Tavangar[1,7]*

[1]Chronic Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran, [2]Department of Medical Biotechnology, School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran, [3]Medical Nanotechnology and Tissue Engineering Research Center, Yazd Reproductive Sciences Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran, [4]Stem Cell Biology Research Center, Yazd Reproductive Sciences Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran, [5]AI VIVO Ltd., Bioinnovation Centre, Cambridge, United Kingdom, [6]Genetic Department, Institute of Systems Biology, University of Cambridge, Cambridge, United Kingdom, [7]Department of Pathology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

Single-cell sequencing has become one of the most used techniques across the wide field of biology. It has enabled researchers to investigate the whole transcriptome at the cellular level across tissues, which unlocks numerous potentials for basic and applied studies in future diagnosis and therapy. Here, we review the impact of single-cell RNA sequencing, as the prominent single-cell technique, in pancreatic biology and cancer. We discuss the most recent findings about pancreatic physiology and pathophysiology owing to this technological advancement in the past few years. Using single-cell RNA sequencing, researchers have been able to discover cellular heterogeneity across healthy cell types, as well as cancer tissues of the pancreas. We will discuss the new immunological targets and new molecular mechanisms of progression in the microenvironment of pancreatic cancer studied using single-cell RNA sequencing. The scope is not limited to cancer tissues, and we cover novel developmental, evolutionary, physiological, and heterogenic insights that have also been achieved recently for pancreatic tissues. We cover all biological insights derived from the single-cell RNA sequencing data, discuss the corresponding pros and cons, and finally, conclude how future research can move better by utilizing single-cell analysis for pancreatic biology.

KEYWORDS

single-cell sequencing, transcriptome analysis, pancreas biology, pancreas cancer, genetic heterogeneity

1 Introduction

What if we knew which genes are specifically expressed in individual cells of the pancreas? This is a question an enthusiastic researcher would ask not long ago, with a vision of a possible method. Today, however, this question seems funny, owing to the development and speed of next-generation sequencing. We are now able to characterize a piece of desired tissue and dig into the individual cells and see which, for example, RNAs are expressed where. Being able to do the same for the whole genome, we are currently able to molecularly

characterize single cells in tissues. To know the meaning of characterization, let us consider the matter of pancreatic tissue. By means of single-cell RNA sequencing, researchers are able to subtype the pancreatic tissue in more detail, find new immune-prominent cells and lineages, dig into tumor heterogeneity, trace different lineages toward development, and subtype various kinds of cancer and other pancreatic disorders, such as diabetes.

In this paper, we are going to review recent single-cell studies on the pancreas with a focus on RNA sequencing in a classified manner. To clarify, the reader will discover how single-cell RNA sequencing has helped us gather additional information on pancreatic biology. Therefore, we will narrate the surprising biological findings and compare and integrate the efforts in order to cast a light for future studies. The major part of this paper focuses on cancer similar to the literature; however, surprising findings about stem cells, new lineages in pancreatic normal tissue, efforts to characterize and map the overall pancreas, immunological studies and insights for the future are also discussed. We will incorporate the real data from selected landmark studies in order to provide a detailed comparative insight for the reader of the whole paper, which is not carried out in similar prestigious reviews in the field (Luo et al., 2020; Musa, 2020; Pompella et al., 2020; Han et al., 2021; Hematol et al., 2021; Mannarapu et al., 2021).

It is also important to observe that single-cell analysis is not limited to RNA sequencing or even sequencing overall. Figure 1 shows different techniques have been incorporated in all the main areas of single-cell analysis. Notably, nano-fabrication technology is the main progressive core for carrying out the single-cell research as all the modalities, such as cytometric devices and sequencing technologies, are built upon an intricate nano-technology. However, it is only through sequencing that an unbiased study of all the possible genes in charge of a biological mechanism becomes available. To recapitulate, other methods such as cytometry and microfluidic sorting are dependent on biological reagents that are all chosen based on previous biological knowledge, while sequencing can provide the data for the whole genome or transcriptome, allowing us to look for completely new mechanisms involved. However, all these methods came with their own pros and cons, as for instance, sequencing will not be able to provide wide proteomic data and this suggests using combinatory methods for a thorough scientific endeavor. Good examples of such combination are new technologies such as single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) (Ji et al., 2020), which provides noisy but analyzable data on the regulome for the whole genome, invading both genomic and transcriptomic landscapes, and single-cell bisulfite sequencing (scBS-seq) (Clark et al., 2017), which provides genome-wide cytosine methylation data, processed to generate information on the transcriptome (Smallwood et al., 2014).

2 Single-cell sequencing for pancreatic cancer studies

Single-cell sequencing can provide data on genetic mutations, transcription status, and level of intensity for susceptible genes, responses to therapy, immunological responses, and overall tumor microenvironment. In addition to diagnosis and therapy, this technology can also bring us knowledge on relying causes and molecular mechanisms of cancer, especially by combinatory

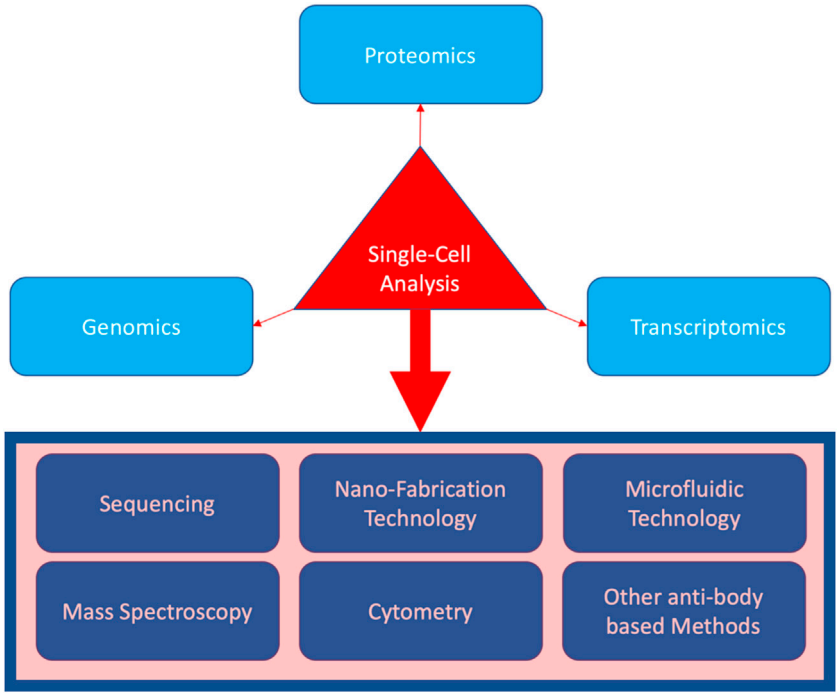


FIGURE 1
Three main biomolecular areas covered by single-cell analysis (in light blue), as well as the techniques incorporated (in the rectangular box).

methods like some of our suggestions for future work throughout the paper. In this manner, in order to classify the literature, we first discuss the prototype studies providing insights into the pancreatic tumor microenvironment both pathologically and immunologically using single-cell RNA sequencing. Since invasiveness is regarded as one of the most distinctive features of cancer incidence, we review the studies reporting discoveries of new mechanisms for cancer progression using the technology. Finally, as cancer incidence is becoming significantly more diverse in the language of personalized medicine, we review the studies using the sequencing technology to put tumor heterogeneity in a molecular perspective. However, let us first start with a discussion on the disease itself.

2.1 Pancreatic cancer

Pancreatic cancer has received significant attention from biomedical researchers worldwide. According to the American Cancer Society, most of the incidents of pancreatic cancer originated from the central tissues and are, therefore, called exocrine cancers, a great deal of which are pancreatic adenocarcinomas that originate from duct cells (Chen et al., 2021b). However, there are less prevalent subtypes, with the most famous being the one that originates from the acinar cells and is called the acinar cell carcinoma. Overall, there are endocrine incidents of pancreatic cancer called the pancreatic neuroendocrine tumors (NETs), or islet cell tumors. They are significantly less prevalent (less than 2% of pancreatic cancers); however, they have greater prognostic capability (Marini et al., 2021). Not surprisingly, most cancers originate from regions that had the stem cells present. To be more precise, cells with regenerative capabilities were uniquely found at the central ducts and exocrine parts of the pancreas. Similarity and proximity of genes for cancer prognosis and stemness are important to understand the prominent cause of this correlation. Both cancer prognosis and stemness are based on maintaining the mitotic capability in a specialized molecular way, which remains a great substrate to build our interpretation of genes contributing in both. This is exactly where single-cell transcriptome studies of the pancreas can become helpful in navigating us through complex molecular biology of both cancer and stem cell research. Hence, we begin by discussing pancreatic cancer research using the tool of single-cell RNA sequencing, while casting a glance on insights for non-cancer research.

2.2 Insights into the tumor microenvironment

Numerous groups around the world are focusing on one of the greatest factors affecting the tumors' faith in the body, i.e., the microenvironment they live in (Khatami and Tavangar, 2018; Khatami et al., 2018; Khatami et al., 2019; Raghavan et al., 2020; Shabani et al., 2020; Chen et al., 2021a). Ligorio et al. (2019) reported how the stromal microenvironment could navigate the intra-tumoral architecture. They managed to reconstruct the tumor tissue and identify eight subgroups according to the architecture of the tumor gland. As they conclude, cancer-associated fibroblasts play a key role in shaping cancer cells' microenvironment and, hence, the tumoral architecture. Reshaping the tumor architecture is not limited to one group, as for instance, Moncada et al. (2020) also attempted a tumor molecular configuration reconstruction using the microarray technology and single-cell sequencing. In order to cast light on the novelty of this type of research, we discuss the journey from the beginning, as it all started with the birth of spatial transcriptomics (ST) in science (Vickovic et al., 2014). With normal transcriptomics, we would have the molecular data for a pool of cells from a piece of tissue, out of which different lineages can be identified using their molecular attributes. However, with spatial transcriptomics, cell clusters are analyzed on microarrays that can serve as a vector to the position of the analyzed cluster. Hence, molecular reconstruction of the whole tissue architecture becomes available. Notably, ST does not offer a single-cell resolution, and Moncada et al. only attempted to push the limitation line forward as the combined ST data with single-cell RNA sequencing provided single-cellular details of the ST cell clusters. In other words, future research for spatial transcriptomics must rely on the way of increasing the resolution of analytical microarrays in order to enhance the reconstruction output resolution to a single-cell level, just as Moncada et al. (2020) attempted. Also, a combination of this technology with other convenient methods seems beneficial. A good example of such combination is the recent paper by Cui Zhou et al. (2022) that used single-cell RNA sequencing in collaboration with spatial transcriptomics, bulk proteogenomics, and cellular imaging. By mapping mutations and copy number events, they were able to notify normal cells and distinguish cells with multiple neoplasia. They also combined histology and spatial transcriptomics to identify the transitional subpopulations. To enlighten, they used histopathological data for deconvolution of ST data. This method is suitable for finding several gene expression relations as they discovered for TIGIT and Nectin in regulatory T cells and tumor cells, respectively (Cui Zhou et al., 2022).

On the other hand, sequencing in combination with modern genetic edition technologies has enabled us to molecularly monitor every single desired change in the microenvironment of any cell. To enlighten with an example, Chen et al. (2021c) deleted type 1 collagen in the myofibroblasts surrounding the pancreatic cancer cells and monitored the changes in the immune response and the progression of cancer. Meaningfully, they made the fibroblasts surrounding the tumor dysfunctional by eliminating their most abundant collagen type. Their results suggest that therapeutic methods severely interfering with the metabolism and molecular well-being of fibroblasts and healthy tissues surrounding the tumor site would be favorable for disease progression or possible reemergence of tumors if we are not successful in targeting all the cancer stem cells.

In order to further enlighten the role of RNA sequencing in the research study, we discuss another modern therapeutic method that must rely on single-cell sequencing for monitoring the resulting changes. CRISPR engineering has enabled the single-cell genome edition with a significant precision level, and using this cellular therapy, *in vivo* precise genetic therapy is becoming available (Zhang, 2021; Zhang et al., 2021). Notably, it is through single-cell sequencing that we can verify the genetic changes we make and the

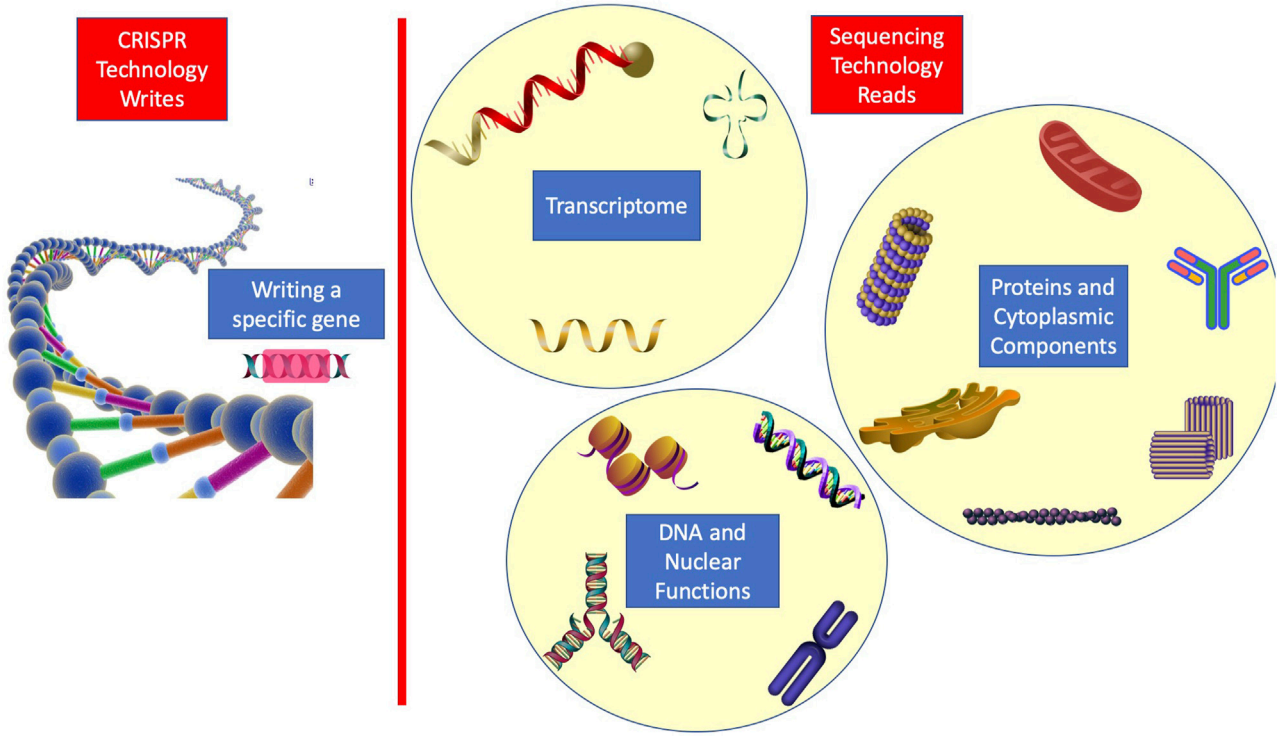


FIGURE 2
Illustration of how sequencing and CRISPR can be used in combination to further improve human knowledge on various biological functions.

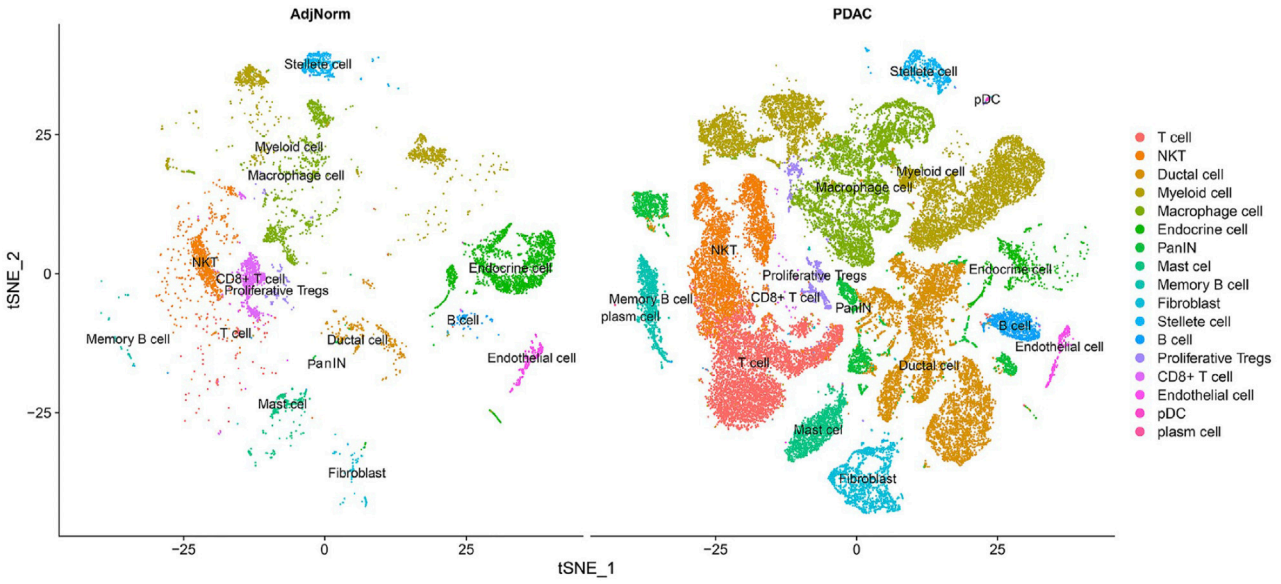


FIGURE 3
Characterizing all the immune cells in the pancreatic tumor microenvironment (Tang et al., 2021).

resulting molecular effects with the greatest accuracy. To recapitulate, CRISPR enables us to write in a genetic language, while sequencing provides the reading in genetic and transcriptome wide languages. Hence, it is beneficial for future works to focus on writing particular genetic codes and trying to figure out changes all across the transcriptome,

exploiting both technologies. Figure 2 shows the idea of how a wide range of biological functions can be viewed from a molecular perspective using the combination.

Moreover, single-cell data enables us to gather novel immunologic insights of cancer. For instance, Tang et al. (2021) performed a single-cell analysis on the tumor microenvironment to characterize the set of present immune cell types. As parts of their results are shown in Figure 3, all the present immune cells have been characterized and subset into populations matching the criteria of each particular category. The immunologic composition of the tumor microenvironment is of great importance, especially recently after numerous cross-talks and relationships between different immune cell types have been observed (Raghavan et al., 2020). Future therapy has no other way than to rely on these molecular findings regarding the tumor-immune microenvironment. Regarding the increasing immunological prominence in research, we discuss immunologic findings using single-cell RNA sequencing for the pancreas in the last section of this paper.

2.3 Discovery of new mechanisms of cancer progression

Although thorough knowledge of the underlying causes of pancreatic cancer and the molecular path from incidence to metastasis is yet to be investigated, some groups have gathered novel information on the matter using single-cell RNA sequencing. For instance, Zhu et al. reported a lethal molecular stage prior to invasiveness in pancreatic cancer (Id et al., 2021). To clarify, they have discovered the pre-invasive transition of adipose-derived stromal cells into *COL11A1*-expressing cancer-associated fibroblasts. With the transition being essential to invasiveness, this can be considered a discovery of a molecular target with therapeutic applications. In a similar finding, Dominguez et al. (2019) reported the transition of stromal cells into LRRC15 + myofibroblasts as the determinant of a patient's response to cancer immunotherapy. Notably, monitoring these kinds of lineage transitions has become part of our everyday research as a result of single-cell RNA sequencing incorporation. Not surprisingly, here both groups have used single-cell sequencing; the first was on the tissues of patients at different stages of pancreatic cancer, and the second was on an animal model. Their common approach was to look for common molecular signatures in different types of cells at various stages of the progression but present in some new lineages emerging during cancer. To restate, first, the lineage in charge of progression and invasiveness was found, and then, using the molecular signatures present in the transcriptome, ancestor cell lines were discovered. Next the figured-out transition was monitored as a corroboration to be ensured. This methodology can be beneficial in cancer research as each group digs out a bit more of the underlying molecular mechanism in charge of the disease. However, for the benefit of patients, more combinatory research studies to figure out a molecular map for cancer progression are needed. A map is presented from which the best personalized criteria of molecular targets can be administered to the patient. The complexity of this map refers mostly to the fact that it is about unlimited correlations and cross-talks, not only at the RNA level but also at DNA levels, and all the functional proteins and even lipids in the cell. To enlighten this multilayer nature, Yue et al., for instance, shed light on the prominence of genetic variations in controlling the transcriptome evolution in cancer (Khatami and Tavangar, 2018; Chan-seng-yue et al., 2020), while Dey et al. (2020) discovered the role of cytokines in cell reprogramming during cancer courses.

Single-cell RNA sequencing has made a significant contribution in further classification of the diseases. In other words, sometimes, new molecular findings change the classical way we previously looked at the classification of diseases, such as cancer, similar to Chung et al.'s (2019) prominent finding. They reported the signaling between endocrine and exocrine pancreas to be in charge of molecular progression in PDAC. Notably, this changes our classic view of considering PDAC as an exocrine pancreatic disorder standing alone.

One of the most important factors to be considered a significant help to pancreatic biology is the composition of modern genetic knockdown and single-cell sequencing. In other words, we are able to simultaneously alter the genome and monitor the effects at molecular and cellular levels. As a standard example for this kind of research, Shah et al. (2017) inspected in detail the molecular changes in APE1/Ref-1 knockdown. In this methodology, they took a gene of selection, knocked it down, and visited all the resulting molecular changes with sequencing to figure out the related pathways and genes with respective affinities. Sometimes, we know what the genes do and knock them down to observe the biological side effects in certain disorders such as cancer. Carstens et al. (2021), for instance, knocked *Snail*, *Twist*, and *ZEB1* genes to stabilize the tumor in the epithelial state. In all efforts, this ensured epithelial stabilization led to the promotion of cancer progression, immune suppression, and improvement in the liver's metastatic conditions. Although epithelial to mesenchymal transition occurs during cancer, they claimed that stabilizing the tumor in the epithelial state promotes collective cell migration and results in colonized, hence more powerful, metastasis. The possibility of classification of tumors in the epithelial–mesenchymal spectrum, as they did in Figure 4, is all again owed to single-cell sequencing.

2.4 Tumor heterogeneity in a molecular perspective

One of the most prominent roles of single-cell RNA sequencing in the current research is to enable us to discover the traces of cell lineages even with a few cells of that lineage present in the tissue. To restate, a piece of tissue such as the pancreas, which was classically categorized and identified based on different macro compartments, such as ducts and islets, with their specific cells, can now be characterized down to single cells; this way, an avalanche of newly discovered lineages with possible roles in immunotherapy or regenerative medicine is in line.

For pancreatic cancer, there have been several groups working on how single-cell RNA sequencing can reveal tumor heterogeneity (Quandt et al., 2011; Peng et al., 2019; Schlesinger et al., 2020). Since the output is more detailed molecular data on tumor cell types, this type of research is

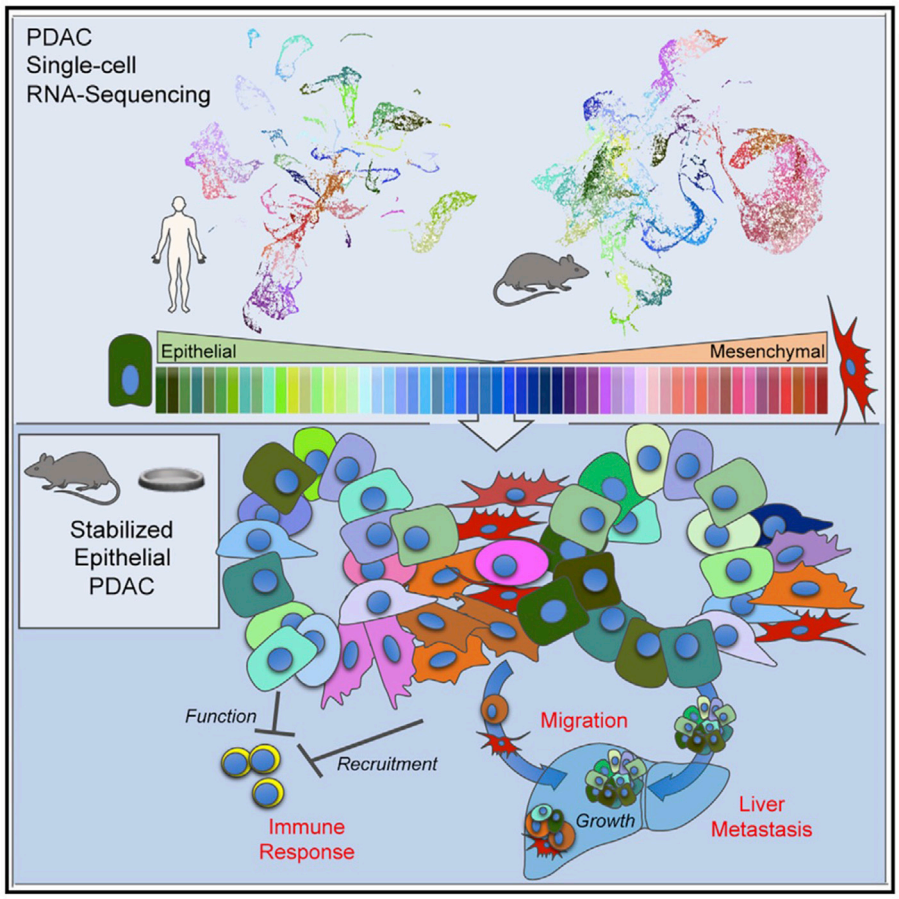


FIGURE 4
Carstens et al. (2021) provided evidence on how an epithelial phenotype of the primary tumors promotes metastasis and inhibits immune responses; they classified the primary tumors using single-cell RNA-seq.

favorable for personalized medicine, where particular therapy specialized for an individual is prescribed. More detailed data lead to more detailed classifications and that leads to more specialized classification of patients on the road to personalized medicine. To recapitulate, single-cell RNA sequencing can surely alter the way we categorize various diseases in the future, illuminating their molecular particularity. In this regard for pancreatic cancer, groups focus on descriptive studies to unveil the disease at a molecular level (Sho et al., 2017; Lin et al., 2020; Ren et al., 2021). They have all used pieces of cancerous or surrounding normal tissues to perform single-cell RNA sequencing and attempted to classify the cellular population based on discrepancies in transcriptome (Lin et al., 2020; Tamaddon et al., 2022). This categorization-based methodology of research can be undertaken in different orders of style. Hossein et al., for instance, took a stepwise approach in which they initiated the categorization into macro-types using the main molecular traces (i.e., all the fibroblasts into one category) (Hosein et al., 2019). They carried out the categorization for the early and late stages of progression and comparatively two distinct tumor subtypes. Then, they stepped up to navigate different subpopulations of fibroblasts and macrophages as two key elements affecting the tumor faith (Figure 5). To criticize this type of research, it is important to notice that single-cell RNA sequencing data from incidents of cancer in pieces of cancer tissue from mice cannot be considered the standard for the cancer progression phenomenon. However, it provides a strong basis for future similar research on human tissues and cross species in order to figure out the molecular path of incidence and progression for pancreatic cancer. Also, it is not limited to disease conditions as it can also be beneficial in increasing our knowledge of healthy pancreatic subpopulations. For instance, Baron et al. (2016) conducted a thorough research and an immense single-cell analysis to provide the transcriptomic map of human and mouse pancreases. They tried to reconstruct the overall pancreatic tissue based on 14 distinct cellular types. Efforts like these can navigate human beings' knowledge on precision anatomy. Also, this type of research is complementary to the style of Wollny et al. (2016) as they tried to subtype pancreatic islets. In other words, they took one more step toward the architecture of islets. The idea for all these types of research is shown in Figure 6, and the method provides a detailed image of pancreatic tissues.

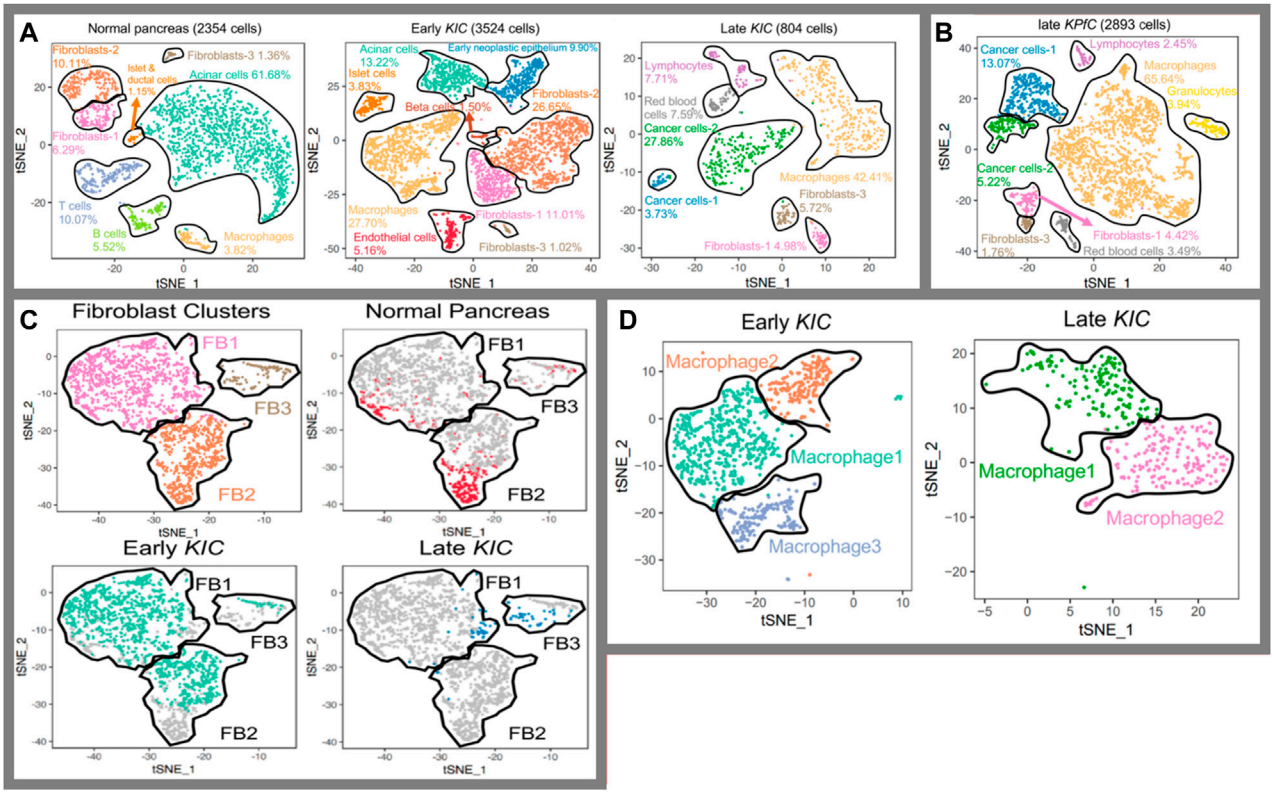


FIGURE 5
Identification of tissue composition at a molecular level (Hosein et al., 2019). (**A**) Cell populations in normal, early, and late cancerous tissue; **(B)** comparison in between two subtypes of KRAS-induced cancer; **(C)** fibroblasts into prospective in normal tissues as well as early and late cancer tissue,; and **(D)** populations of macrophages in magnification through the cancer course.

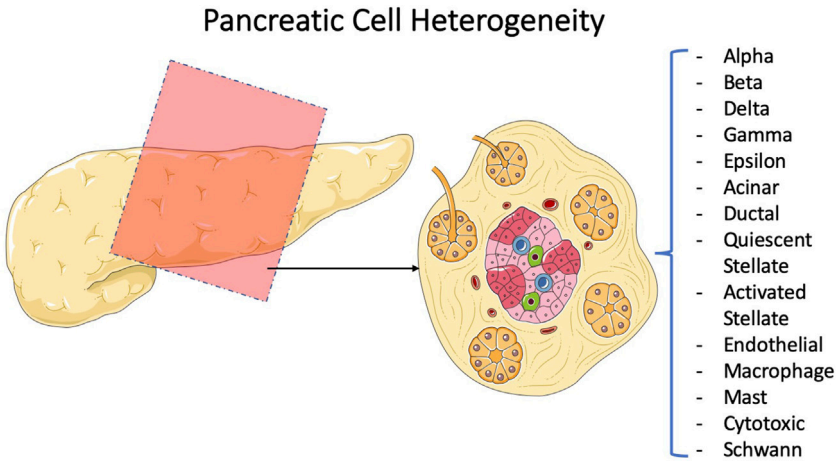


FIGURE 6
Efforts to characterize and subtype the pancreatic tissue. Comparing the percentages for each subpopulation of the pancreatic tissue under normal and disease conditions using sequencing technology has been the method of diagnosis for several groups. Baron et al. (2016)conducted a descriptive study on pancreatic cellular subtypes, and Wollny et al. (2016) shed light more on one subtype, namely, the diabetically prominent islets.

3 Pancreatic pathology, for abnormalities other than cancer

Applications of single-cell sequencing in pancreatic studies for sure are not limited to cancer. CP or chronic pancreatitis, for instance, was the topic of interest of Mao et al. (2022) as they characterized the ductal cell in the process of the disease. Their study is a good example of how a set of genes and markers responsible for a condition could be revealed in a single study. With clustering methods, they were able to sub-populate the cells and carried out precise genetic analysis to discover few hundred upregulated and downregulated genes. However, only overexpression for MMP7 and TTR was further verified. Other than CP, pancreatic disorders of various types such as acute (Sun et al., 2021) and chronic pancreatitis (Weiss et al., 2018) and diabetes have also been genetically investigated using the single-cell sequencing technology. Also, sequencing has shed light on our microbiome knowledge, enabling numerous groups to identify the relationship between various pancreatic disorders and gut microbiota with the help of sequencing (Ammer-Herrmenau et al., 2020; Qi-Xiang et al., 2022).

4 Single-cell RNA sequencing for pancreas biology

4.1 Multipotency in pancreatic tissue

Multipotent cells with regenerative abilities have recently been an interesting topic in biomedical research. Finding new lineages with stemness in the tissue will significantly increase hopes for regenerative capabilities and therapeutic applications (Hutton et al., 2021). To enlighten the prominence, suppose we discovered multipotent cells in the pancreas with the capability to differentiate into beta cells under certain conditions, which could help in the fight against diabetes. In this regard, numerous groups reported to have discovered new progenitor-like cells and characterized the previously known progenitors to uncover pancreatic development. Krentz et al. (2018), for instance, characterized known progenitor cells from embryos of mice to derive beta cells and provided a detailed description of the pancreas at the molecular level during development in mice. This can be used as a guide for efforts to form beta cells in both vitro and vivo from human embryonic and adult stem cells. The presence of stem cells in the adult pancreas and their ability to evolve to functional beta cells have been controversial in research in recent years. However, in a comprehensive research conducted by Mameishvili et al. (2019) published in PNAS in 2019, the controversy for the presence of isolable stemness in an adult pancreas somehow ended. Since, a signature gene (*Aldh1b1*) has been reported to be a certain flag for isolating cells with progenitor behavior and the capability of evolving to all three main subtypes in the adult pancreas. Not surprisingly, the same gene is reported to be responsible and core to pancreatic cancer (performed particularly on the KRAS-induced subtype) (Mameishvili et al., 2019). Some groups have moved further, exploiting technology to form pancreatic organoids *in vitro* in order to molecularly monitor their differentiation. For example, Wiedenmann et al. (2021) used microfluidic technology and single-cell transcriptome monitoring to unveil the molecular path of differentiation from pancreatic progenitor cells to duct-like organoids. The combinatory in Figure 7 shows how discussed groups unveiled different parts of the mystery of pancreatic stemness biology.

4.2 Pancreatic fibroblasts into perspective

There have been numerous groups publishing papers focusing on fibroblasts as key to biological mechanisms they look for in the pancreas (Elyada et al., 2019; Kieffer et al., 2020; Steele et al., 2021). Fibroblasts are common cell types prevalent in various parts of the body and different tissues. Day by day, new insights into the effectiveness of fibroblasts on various molecular mechanisms of cancer progression and immune responses are being discovered (Sebastian et al., 2020). Therefore, we designated a separate part discussing which pancreatic fibroblast findings are most thankful for single-cell RNA sequencing technology.

Concluding based on the literature (Elyada et al., 2019; Kieffer et al., 2020; Hutton et al., 2021), it seems that according to pancreatic ductal adenocarcinoma that fibroblasts play a key role in mediating the immune microenvironment. Even new subtypes of fibroblasts are discovered every now and then. For instance, Hutton et al. (2021) has discovered a new classification of fibroblasts using the CD105 marker, and they could engineer how this fibroblast lineage affects tumor development. To restate, the CD105-negative fibroblasts supported anti-tumor immunity, while CD105-positive ones were tumor permissive. Finding lineages like this can be much beneficial in moving toward better therapeutic regimens. Notably, they used mass cytometry to carry out the single-cell analysis. This method is suitable for working with known markers by figuring out a new combination of markers like what they did. However, as Wang et al. (2021) did for the same purpose of subtyping the fibroblasts, single-cell RNA sequencing has a greater precision generally and provides much more information, the pool from which we can fish out new biomarkers. Finally, as we also discussed in the previous sections, fibroblasts significantly contribute to the condition for the tumor microenvironment and play an important role in tumor heterogeneity.

4.3 Pancreatic tissue biology

Monitoring single cells at the transcriptome level can gather much information on various corners of biology. Evolution and development are prominent corners, for instance. In this regard, Enge et al. (2017) did a good job of uncovering more of the molecular pattern of aging. They tried to uncover the genes in charge of the transcriptional noise inherent in aging. Their idea and results seem surprising. However, it is important to note that all the 2,544 cell sequences were carried out from samples taken only from eight donors, spanning 60 years of age. Hence, individual discrepancies at a single-cellular level seem to have an inevitable effect. Stepping up these kinds of studies to apply on tissues from numerous human or animal models can better verify the genes in charge of developmental or aging processes.

One other corner of the biological planet covered by single-cell sequencing in the pancreas is immunology. Now and then, new

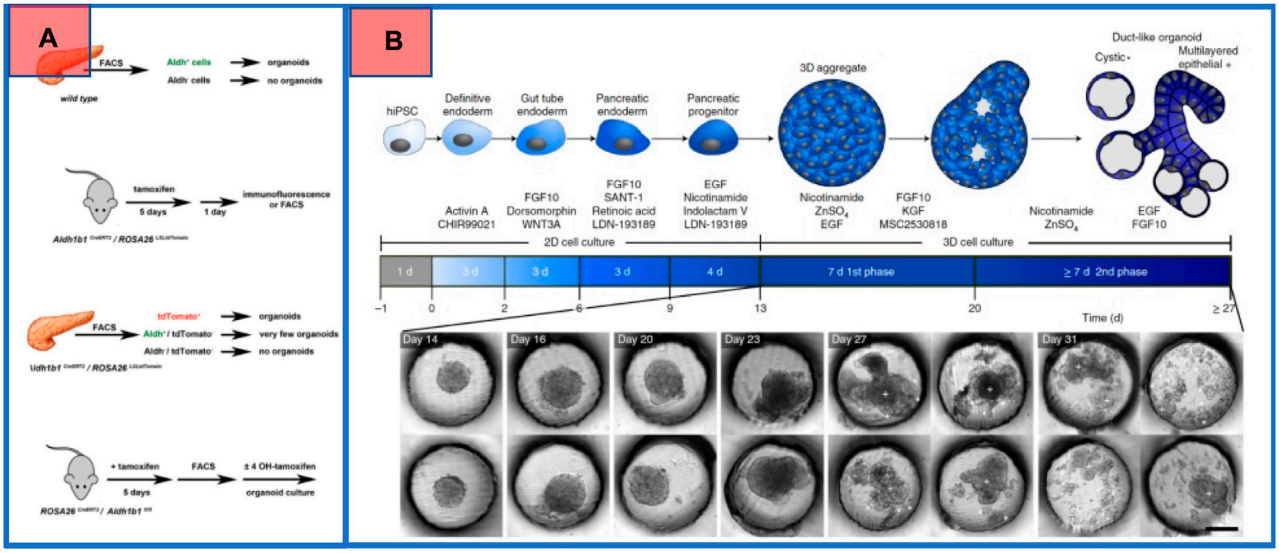


FIGURE 7
Effort to isolate and characterize human stem cells. The embryonic and adult stem cells have been characterized owing to single-cell RNA sequencing which now provides a substrate to look deeper for underlying mechanisms. **(A)** Wiedenmann et al. used microfluidic technology to grow duct-like organoids from pancreatic progenitors (Wiedenmann et al., 2021). **(B)** Mameishvili et al. provided evidence that *ALD1HB1* is the selector gene for pancreatic adult progenitors and is associated with KRAS-driven pancreatic cancer.

relationships between diverse types of immune cells and new lineages with particularly prominent functions are discovered (Pan et al., 2019; Li et al., 2021; Rodriguez et al., 2021). The fact that makes the immune system unique is the significant variety in the type of cells, with sometimes only few cells existing from one specialized lineage. This is exactly where a significant need for single-cell analysis is sensed, and single-cell sequencing both for DNA and RNA comes to the rescue. Accordingly, numerous groups have been focusing on single-cell analysis on pancreatic pathological conditions and discovering new immunological mechanisms. Kemp et al., for instance, fetched out a new way to molecularly detect pancreatic cancer, as they discovered special tumor-associated macrophages with the same traces present in the blood's monocytes (Kemp et al., 2021). They analyze the relatively small number of monocyte cells present in the blood sample down to the transcriptome owing to single-cell RNA sequencing. As another example, the same group published a detailed descriptive paper on the molecular landscape of immune cells in the tumor and peripheral blood, again using single RNA sequencing (Steele et al., 2020). Notably, they attained a pool of genetic data out of which several biomarker candidates could be selected. As another creative and more combinatory study example, Sivakumar et al. (2021) incorporated mass cytometry, single-cell RNA sequencing, immunohistochemistry, and spatial analysis focusing on T cells to reconstruct the immune system microenvironment. Using this approach, they discovered that unlike $CD4^+$ and $CD8^+$ cells that were coherently spread in the tissue, regulatory T cells were meaningfully present only in the stromal cancer tissue with minimum presence in the epithelial compartments, at both cancer and normal sides. This was the piece of evidence for their hypothesis of activated Treg cells being in charge of PDAC progression. Also, they looked into the indifference of $CD8^+$ T cells to the presence of a tumor and found senescence in charge, mostly analyzed by single-cell RNA sequencing.

5 Conclusion

Single-cell RNA sequencing is a powerful tool for studying pancreatic physiology and pathophysiology. However, there is a main difference in the way researchers utilized the technique. Some used complementary sorting methods, such as mass cytometry, to navigate the cellular population down to a more known subpopulation and then used sequencing. This methodology always provides easier analysis for the output sequencing data since we obtained good molecular clues for the sorted population. On the contrary, sequencing on unsorted tissue cells will produce sequenced data that are much harder to analyze but with much more novel information. In other words, completely new insights into biological mechanisms, such as new correlations, complex signaling networks between cells, or whole new lineages, must be extracted from a pool of raw sequenced data with all the subsets of the tissue present. Notably, pancreatic tissue is a relatively complex one, being involved in the diverse physiological processes of the complex endocrine, nervous, and digestive systems that ensures the complex unknown correlations between cells from different pancreas components, enlightening the prominence of the whole descriptive tissue single-cell study for the pancreatic, nervous, and endocrine biology. Also, a combination with other technologies such as microfluidics and nano-fabrication can always be beneficial as it has already provided us with significant inventory tools such as spatial transcriptomics. In other words, future single-cell studies on the pancreas have ways to improve, enabling us to perform single-cell sequencing for the whole tissue by scaling up the technology, also to be used in combination with other sorting methods for the whole tissue, not losing data of a single cell. The good news is that computers and the regrading algorithms have been

developed in advance and are ready to satisfy the need for scaling up the transcriptomics. Finally, the prominent early result of these combinations seems to be an enhancement in spatial transcriptomics, improving its limitations to the accessible goal of single-cell spatial resolution and for the whole tissues. Overall, having explained the applications of single-cell sequencing and analysis in pancreatic studies, the journey of potential future applications seems to have just started, and a load of new insights will emerge with advances in novel instruments and methodologies.

Author contributions

The first draft of the manuscript was written by MT, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflict of interest

PG was employed by AI VIVO Ltd.

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

Publisher's note

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

References

Ammer-Herrmenau, C., Pfisterer, N., Weingarten, M. F., and Neesse, A. (2020). The microbiome in pancreatic diseases: Recent advances and future perspectives. *United Eur. Gastroenterol. J.* 8 (8), 878–885. doi:10.1177/2050640620944720

Baron, M., Veres, A., Wolock, S. L., Faust, A. L., Gaujoux, R., Vetere, A., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell. Syst.* 3, 346–360. doi:10.1016/j.cels.2016.08.011

Carstens, J. L., Yang, S., Correa de Sampaio, P., Zheng, X., Barua, S., McAndrews, K. M., et al. (2021). Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *CellReports* 35 (2), 108990. ElsevierCompany. doi:10.1016/j.celrep.2021.108990

Chan-seng-yue, M., Kim, J. C., Wilson, G. W., Ng, K., Figueroa, E. F., O'Kane, G. M., et al. (2020). Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat. Genet.* 52, 231–240. doi:10.1038/s41588-019-0566-9

Chen, K., Wang, Q., Li, M., Guo, H., Liu, W., Wang, F., et al. (2021a). Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. *EBioMedicine* 66, 103315. doi:10.1016/j.ebiom.2021.103315

Chen, X., Zeh, H. J., Kang, R., Kroemer, G., and Tang, D. (2021b). Cell death in pancreatic cancer: From pathogenesis to therapy. *Nat. Rev. Gastroenterol. Hepatol.* 18 (11), 804–823. doi:10.1038/S41575-021-00486-6

Chen, Y., Kim, J., Yang, S., Wang, H., Wu, C. J., Sugimoto, H., et al. (2021c). Type I collagen deletion in a SMA + myofibroblasts augments immune suppression and accelerates progression of pancreatic cancer ll Article Type I collagen deletion in a SMA + myofibroblasts augments immune suppression and accelerates progression. *Cancer Cell.* 39 (4), 548–565. e6. doi:10.1016/j.ccell.2021.02.007

Chung, K. M., Singh, J., Lawres, L., Dorans, K. J., Garcia, C., Burkhardt, D. B., et al. (2019). Endocrine-exocrine signaling drives obesity-associated pancreatic ductal adenocarcinoma', *bioRxiv*. *Cold Spring Harb. Lab.* 12 (11), 663583. doi:10.1101/663583

Clark, S. J., Smallwood, S. A., Lee, H. J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* 12 (3), 534–547. doi:10.1038/nprot.2016.187

Cui Zhou, D., Jayasinghe, R. G., Chen, S., Herndon, J. M., Iglesia, M. D., Navale, P., et al. (2022). Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. *Nat. Genet.* 54 (9), 1390–1405. doi:10.1038/s41588-022-01157-1

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

Dominguez, C. X., Muller, S., Keerthivasan, S., Koeppen, H., Hung, J., Gierke, S., et al. (2019). Single-cell RNA sequencing reveals stromal evolution into LRRC15 + myofi broblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discov.* 10, 232–253. doi:10.1158/2159-8290.CD-19-0644

Elyada, E., Bolisetty, M., Laise, P., Flynn, W. F., Courtois, E. T., Burkhart, R. A., et al. (2019). Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* 9, 1102–1123. doi:10.1158/2159-8290.CD-19-0094

Enge, M., Arda, H. E., Mignardi, M., Beausang, J., Bottino, R., Kim, S. K., et al. (2017). Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell.* 171, 321–330.e14. doi:10.1016/j.cell.2017.09.004

Han, J., RonaldDePinho, A. A. M., and Maitra, A. (2021). Single-cell RNA sequencing in pancreatic cancer. *Nat. Rev. Gastroenterology Hepatology* 18, 451–452. doi:10.1038/s41575-021-00471-z

Hematol, J., Tang, R., Xu, J., Wang, W., Zhang, B., Liu, J., et al. (2021). Applications of single - cell sequencing in cancer research: Progress and perspectives. *J. Hematol. Oncol.* 14, 91–26. doi:10.1186/s13045-021-01105-2

Hosein, A. N., Huang, H., Wang, Z., Parmar, K., Du, W., Huang, J., et al. (2019). Cellular heterogeneity during mouse pancreatic ductal adenocarcinoma progression at single-cell resolution. *JCI Insight* 5, e129212. doi:10.1172/jci.insight.129212

Hutton, C., Heider, F., Blanco-Gomez, A., Banyard, A., Kononov, A., Zhang, X., et al. (2021). Single-cell analysis defines a pancreatic fibroblast lineage that supports anti-tumor immunity. *Cancer Cell.* 39 (9), 1227–1244. e20. doi:10.1016/j.ccell.2021.06.017

Id, K. Z., Cai, L., Cui, C., de Los Toyos, J. R., and Anastassiou, D. (2021). Single-cell analysis reveals the pan-cancer invasiveness-associated transition of adipose-derived stromal cells into COL11A1-expressing cancer-associated fibroblasts. *PLoS Comput. Biol.* 17, e1009228. doi:10.1371/journal.pcbi.1009228

Ji, Z., Zhou, W., Hou, W., and Hongkai, J. (2020). Single-cell ATAC-seq signal extraction and enhancement with SCATE. *Genome Biol.* 21 (1), 1–36. doi:10.1186/S13059-020-02075-3/FIGURES/12

Kemp, S. B., Steele, N. G., Carpenter, E. S., Donahue, K. L., Bushnell, G. G., Morris, A. H., et al. (2021). Pancreatic cancer is marked by complement-high blood monocytes and tumor-associated macrophages. *Life Sci. Alliance* 4 (6), e202000935. doi:10.26508/lsa.202000935

Khatami, F., Larijani, B., Nasiri, S., and Tavangar, S. M. (2019). Liquid biopsy as a minimally invasive source of thyroid cancer genetic and epigenetic alterations. *Int. J. Mol. Cell. Med.* 8, 19–29. doi:10.22088/IJMCM.BUMS.8.2.19

Khatami, F., Larijani, B., and Tavangar, S. M. (2018). The presence of tumor extrachomosomal circular DNA (ecDNA) as a component of liquid biopsy in blood. *Med. Hypotheses* 114, 5–7. doi:10.1016/j.mehy.2018.02.018

Khatami, F., and Tavangar, S. M. (2018). Liquid biopsy in thyroid cancer: New insight. *Int. J. Hematology-Oncology Stem Cell. Res.* 12 (3), 235–248.

Kieffer, Y., Hocine, H. R., Gentric, G., Pelon, F., Bernard, C., Bourachot, B., et al. (2020). Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov.* 10, 1330–1351. doi:10.1158/2159-8290.CD-19-1384

Krentz, N. A. J., Lee, M. Y. Y., Xu, E. E., Sproul, S. L. J., Maslova, A., Sasaki, S., et al. (2018). Single-cell transcriptome profiling of mouse and hESC-derived pancreatic progenitors', *stem cell reports*. *ElsevierCompany* 11 (6), 1551–1564. doi:10.1016/j.stemcr.2018.11.008

Li, D., Schaub, N., Guerin, T. M., Bapiro, T. E., Richards, F. M., Chen, V., et al. (2021). T cell-mediated antitumor immunity cooperatively induced by TGFβR1 antagonism and gemcitabine counteracts reformation of the stromal barrier in pancreatic cancer. *Mol. Cancer Ther.* 20, 1926–1940. doi:10.1158/1535-7163.MCT-20-0620

Ligorio, M., Sil, S., Malagon-Lopez, J., Nieman, L. T., Misale, S., Di Pilato, M., et al. (2019). Stromal microenvironment shapes the intratumoral architecture of pancreatic

cancer article stromal microenvironment shapes the intratumoral architecture of pancreatic cancer. *Cell.* 178 (1), 160–175.e27. doi:10.1016/j.cell.2019.05.012

Lin, W., Noel, P., Borazanci, E. H., Lee, J., Amini, A., Han, I. W., et al. (2020). Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Med.* 12, 80–14. doi:10.1186/s13073-020-00776-9

Luo, Q., Fu, Q., Zhang, X., Zhang, H., and Qin, T. (2020). Application of single-cell RNA sequencing in pancreatic cancer and the endocrine pancreas. *Adv. Exp. Med. Biol.* 1255, 143–152. doi:10.1007/978-981-15-4494-1_12

Mameishvili, E., Serafimidis, I., Iwaszkiewicz, S., Lesche, M., Reinhardt, S., Bolicke, N., et al. (2019). Aldh1b1 expression defines progenitor cells in the adult pancreas and is required for Kras-induced pancreatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* 116 (41), 20679–20688. doi:10.1073/pnas.1901075116

Mannarapu, M., Dariya, B., and Reddy, O. (2021). Application of single - cell sequencing technologies in pancreatic cancer. *Mol. Cell. Biochem.* 476 (6), 2429–2437. doi:10.1007/s11010-021-04095-4

Mao, X., Mao, S., Wang, L., Jiang, H., Deng, S., Wang, Y., et al. (2022). Single-cell transcriptomic analysis of the mouse pancreas: Characteristic features of pancreatic ductal cells in chronic pancreatitis. *Genes.* 13 (6), 1015. doi:10.3390/genes13061015

Marini, F., Giusti, F., Tonelli, F., and Brandi, M. L. (2021). Pancreatic neuroendocrine neoplasms in multiple endocrine neoplasia type 1. *Int. J. Mol. Sci.* 22 (8), 4041. doi:10.3390/IJMS22084041

Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J. C., Baron, M., et al. (2020). Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* 38, 333–342. doi:10.1038/s41587-019-0392-8

Musa, M. (2020). Single-cell analysis on stromal fibroblasts in the microenvironment of solid tumours. *Adv. Med. Sci.* 65 (1), 163–169. doi:10.1016/j.advms.2019.12.001

Pan, Y., Lu, F., Fei, Q., Yu, X., Xiong, P., Yu, X., et al. (2019). Single-cell RNA sequencing reveals compartmental remodeling of tumor-infiltrating immune cells induced by anti- CD47 targeting in pancreatic cancer. *J. Hematol. Oncol.* 6 (29), 124. doi:10.1186/s13045-019-0822-6

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

Pompella, L., Tirino, G., Pappalardo, A., Caterino, M., Ventriglia, A., Nacca, V., et al. (2020). Pancreatic cancer molecular classifications: From bulk genomics to single cell analysis. *Int. J. Mol. Sci.* 21, 2814. doi:10.3390/ijms21082814

Qi-Xiang, M., Yang, F., Ze-Hua, H., Nuo-Ming, Y., Rui-Long, W., Bin-Qiang, X., et al. (2022). Intestinal TLR4 deletion exacerbates acute pancreatitis through gut microbiota dysbiosis and Paneth cells deficiency. *Gut Microbes* 14 (1), 2112882. doi:10.1080/19490976.2022.2112882

Quandt, D., Fiedler, E., Boettcher, D., Marsch, W. C., and Seliger, B. (2011). B7-h4 expression in human melanoma: Its association with patients' survival and antitumor immune response', *Clinical cancer research: An official journal of the American Association for cancer research. Clin. Cancer Res.* 17 (10), 3100–3111. doi:10.1158/1078-0432.CCR-10-2268

Raghavan, S., Winter, P. S., Navia, A. W., Williams, H. L., DenAdel, A., Kalekar, R. L., et al. (2020). Transcriptional subtype-specific microenvironmental crosstalk and tumor cell plasticity in metastatic pancreatic cancer. BioRxiv.

Ren, X., Zhou, C., Lu, Y., Ma, F., Fan, Y., and Wang, C. (2021). Single-cell RNA-seq reveals invasive trajectory and determines cancer stem cell-related prognostic genes in pancreatic cancer. *Bioengineered* 12, 5056–5068. doi:10.1080/21655979.2021.1962484

Rodriguez, E., Boelaars, K., Brown, K., Eveline Li, R. J., Kruijssen, L., Bruijns, S. C. M., et al. (2021). Sialic acids in pancreatic cancer cells drive tumour-associated macrophage differentiation via the Siglec receptors Siglec-7 and Siglec-9. *Nat. Commun.* 12, 1270. doi:10.1038/s41467-021-21550-4

Schlesinger, Y., Yosefov-Levi, O., Kolodkin-Gal, D., Granit, R. Z., Peters, L., Kalifa, R., et al. (2020). Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. *Nat. Commun.* 11, 4516. doi:10.1038/s41467-020-18207-z

Sebastian, A., Hum, N. R., Martin, K. A., Gilmore, S. F., Peran, I., Byers, S. W., et al. (2020). Single-cell transcriptomic analysis of tumor-derived fibroblasts and normal tissue-resident fibroblasts reveals fibroblast heterogeneity in breast cancer. *Cancers (Basel)* 12, 1307. doi:10.3390/cancers12051307

Shabani, N., Sheikholeslami, S., Paryan, M., Zarif Yeganeh, M., Tavangar, S. M., Azizi, F., et al. (2020). An investigation on the expression of miRNAs including miR-144 and miR-34a in plasma samples of RET-positive and RET-negative medullar thyroid carcinoma patients. *J. Cell. Physiology* 235 (2), 1366–1373. John Wiley and Sons, Ltd. doi:10.1002/jcp.29055

Shah, F., Goossens, E., Atallah, N. M., Grimard, M., Kelley, M. R., and Fishel, M. L. (2017). APE1/Ref-1 knockdown in pancreatic ductal adenocarcinoma – characterizing gene expression changes and identifying novel pathways using single-cell RNA sequencing. *Mol. Oncol.* 11, 1711–1732. doi:10.1002/1878-0261.12138

Sho, S., Court, C. M., Winograd, P., Lee, S., Hou, S., Graeber, T. G., et al. (2017). Precision oncology using a limited number of cells: Optimization of whole genome amplification products for sequencing applications. *BMC Cancer* 17 (11301), 457. doi:10.1186/s12885-017-3447-6

Sivakumar, S., Abu-Shah, E., Ahern, D. J., Arbe-Barnes, E. H., Jainarayanan, A. K., Mangal, N., et al. (2021). Activated regulatory T-cells, dysfunctional and senescent T-cells hinder the immunity in pancreatic cancer. *Cancers (Basel)* 13, 1776. doi:10.3390/cancers13081776

Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., et al. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11 (8), 817–820. Nature Publishing Group. doi:10.1038/nmeth.3035

Steele, N. G., Biffi, G., Kemp, S. B., Zhang, Y., Drouillard, D., Syu, L., et al. (2021). Inhibition of hedgehog signaling alters fibroblast composition in pancreatic cancer. *Clin. Cancer Res.* 27 (19), 2023–2037. doi:10.1158/1078-0432.CCR-20-3715

Steele, N. G., Carpenter, E. S., Kemp, S. B., Sirihorachai, V. R., The, S., Delrosario, L., et al. (2020). Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nat. Cancer* 1, 1097–1112. doi:10.1038/s43018-020-00121-4

Sun, H. W., Dai, S. J., Kong, H. R., Fan, J. X., Yang, F. Y., Dai, J. Q., et al. (2021). Accurate prediction of acute pancreatitis severity based on genome-wide cell free DNA methylation profiles. *Clin. Epigenetics* 13 (1), 223. BioMed Central Ltd. doi:10.1186/s13148-021-01217-z

Tamaddon, M., Azimzadeh, M., and Tavangar, S. M. (2022). microRNAs and long non-coding RNAs as biomarkers for polycystic ovary syndrome. *J. Cell. Mol. Med.* 26 (3), 654–670. doi:10.1111/JCMM.17139

Tang, R., Liu, X., Liang, C., Hua, J., Xu, J., Wang, W., et al. (2021). Deciphering the prognostic implications of the components and signatures in the immune microenvironment of pancreatic ductal adenocarcinoma sources of datasets. *Front. Immunol.* 12, 1–15. doi:10.3389/fimmu.2021.648917

Vickovic, S., Salmen, F., Lundmark, A., Navarro, J. F., Magnusson, J., Asp, M., et al. (2014). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82. doi:10.1126/science.aaf2403

Wang, Y., Liang, Y., Xu, H., Zhang, X., Mao, T., Cui, J., et al. (2021). Single-cell analysis of pancreatic ductal adenocarcinoma identi fi es a novel fi broblast subtype associated with poor prognosis but better immunotherapy response. *Cell. Discov.* 7, 36. doi:10.1038/s41421-021-00271-4

Weiss, F. U., Skube, M. E., and Lerch, M. M. (2018). Chronic pancreatitis: An update on genetic risk factors', *Current opinion in gastroenterology. Curr. Opin. Gastroenterol.* 34 (5), 322–329. doi:10.1097/MOG.0000000000000461

Wiedenmann, S., Breunig, M., Merkle, J., von Toerne, C., Georgiev, T., Moussus, M., et al. (2021). Single-cell-resolved differentiation of human induced pluripotent stem cells into pancreatic duct-like organoids on a microwell chip. *Nat. Biomed. Eng.* 5, 897–913. doi:10.1038/s41551-021-00757-2

Wollny, D., Zhao, S., Everlien, I., Lun, X., Brunken, J., Brune, D., et al. (2016). Single-cell analysis uncovers clonal acinar cell heterogeneity in the adult pancreas. *Dev. Cell.* 39, 289–301. doi:10.1016/j.devcel.2016.10.002

Zhang, B. (2021). CRISPR/Cas gene therapy', *Journal of cellular physiology. J. Cell. Physiol.* 236 (4), 2459–2481. doi:10.1002/JCP.30064

Zhang, H., Qin, C., Zheng, X., Wen, S., Chen, W., Liu, X., et al. (2021). Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer. *Mol. Cancer* 20 (1), 126. doi:10.1186/S12943-021-01431-6

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 June 2023
DOI 10.3389/fgene.2023.1089023

OPEN ACCESS

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

REVIEWED BY
Tianshun Gao,
Sun Yat-Sen University, China
Weiwei Huang,
Peking Union Medical College, China

*CORRESPONDENCE
Hong Zhu,
zhuhong@kmmu.edu.cn
Tiehan Li,
litiehan@kmmu.edu.cn

†These authors share first authorship

RECEIVED 03 November 2022
ACCEPTED 06 June 2023
PUBLISHED 16 June 2023

CITATION
Zhu L, Tu D, Li R, Li L, Zhang W, Jin W, Li T and Zhu H (2023), The diagnostic significance of the ZNF gene family in pancreatic cancer: a bioinformatics and experimental study.
Front. Genet. 14:1089023.
doi: 10.3389/fgene.2023.1089023

The diagnostic significance of the ZNF gene family in pancreatic cancer: a bioinformatics and experimental study

Lei Zhu[1†], Dong Tu[2†], Ruixue Li[1†], Lin Li[1], Wenjie Zhang[1], Wenxiang Jin[1], Tiehan Li[1*] and Hong Zhu[1*]

[1]Department of Hepatobiliary and Pancreatic Surgery, Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China, [2]Department of Cardiothoracic Surgery, No. 920 Hospital of the PLA Joint Logistics Support Force, Kunming, China

Background: Pancreatic adenocarcinoma (PAAD) is among the most devastating of all cancers with a poor survival rate. Therefore, we established a zinc finger (ZNF) protein-based prognostic prediction model for PAAD patients.

Methods: The RNA–seq data for PAAD were downloaded from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) databases. Differentially expressed ZNF protein genes (DE-ZNFs) in PAAD and normal control tissues were screened using the "lemma" package in R. An optimal risk model and an independent prognostic value were established by univariate and multivariate Cox regression analyses. Survival analyses were performed to assess the prognostic ability of the model.

Results: We constructed a ZNF family genes-related risk score model that is based on the 10 DE-ZNFs (ZNF185, PRKCI, RTP4, SERTAD2, DEF8, ZMAT1, SP110, U2AF1L4, CXXC1, and RMND5B). The risk score was found to be a significant independent prognostic factor for PAAD patients. Seven significantly differentially expressed immune cells were identified between the high- and low-risk patients. Then, based on the prognostic genes, we constructed a ceRNA regulatory network that includes 5 prognostic genes, 7 miRNAs and 35 lncRNAs. Expression analysis showed ZNF185, PRKCI and RTP4 were significantly upregulated, while ZMAT1 and CXXC1 were significantly downregulated in the PAAD samples in all TCGA - PAAD, GSE28735 and GSE15471 datasets. Moreover, the upregulation of RTP4, SERTAD2, and SP110 were verified by the cell experiments.

Conclusion: We established and validated a novel, Zinc finger protein family - related prognostic risk model for patients with PAAD, that has the potential to inform patient management.

KEYWORDS

pancreatic adenocarcinoma, zinc finger protein family, prognostic risk model, TCGA, bioinformatics

1 Introduction

Pancreatic adenocarcinoma (PAAD) is a devastating malignancy with a very low 5-year survival rate (Kuninty et al., 2019). The onset of PAAD is insidious, and most patients are admitted to the hospital with clinical symptoms such as "jaundice and abdominal pain". Nearly 80% of patients have no chance of surgical resection (Mizrahi et al., 2020). Currently, the main diagnostic option for PAAD is imaging. Compared with other mainstream targeted therapies for malignant tumors, PAAD lacks effective target diagnosis and precise treatment options (Grossberg et al., 2020). The pathogenesis of PAAD is regulated by multiple factors, multiple genes and microenvironments. Therefore, the use of massively parallel sequencing technologies to mine PAAD datasets, and utilization of bioinformatics methods to accurately analyze interaction targets, can provide new ideas for precise treatment of PAAD.

The ZNF domain is present in about 5% of human proteins and is associated with pathogenesis of many solid tumors. Precisely, ZNF259 activates ERK/GSK3β by activating the ERK/GSK3β/Snail signaling pathway to promote breast cancer cell invasion and migration (Xiao et al., 2014; Liu et al., 2018). Due to multifunctional binding abilities of zinc finger proteins, ZNF plays an important role in cell differentiation, cell metabolism, autophagy, apoptosis, and stemness maintenance. There is a need to elucidate on the significance of zinc finger proteins in tumor pathogenesis. Through bioinformatics analyses, Sun et al. (2021) found that the zinc finger protein 2 gene (ZIC2) is positively correlated with immune infiltrating cells in hepatocellular carcinoma (HCC) patients, and elevated ZIC2 mRNA expressions in $CD4^+$ T cells are associated with the 5-year survival rate of HCC patients. These findings, imply that the ZIC2 gene can be used as a marker for liver cancer immune responses, and to predict HCC prognosis. In pancreatic ductal adenocarcinoma, ZNF185 and SERTAD2 are tumor immune targets, providing new ideas for treatment of tumor immune invasion (Chen et al., 2021; Zhang et al., 2022).

In summary, PAAD pathogenesis is regulated by multiple factors, multiple genes and microenvironments. The ZNF gene may be involved in cancer occurrence and progression as an oncogene. Therefore, the use of massively parallel sequencing technologies to mine PAAD datasets, and utilization of bioinformatics methods to accurately analyze interaction targets, can provide new ideas for precise treatment of PAAD. We mined TCGA, GEO, ICGC, as well as Uniprot databases, and used GSE as the training set to construct a PAAD-related risk score model via COX regression analysis. The constructed risk model is of great significance for further studies on the roles of ZNF in pancreatic cancer pathogenesis.

2 Methods

2.1 Data sources

The RNA-seq data and corresponding clinical data for TCGA-PAAD patients were downloaded from-the Genomic Data Commons (GDC) database (GDC Data Portal, RRID:SCR_014514) (https://portal.gdc.cancer.gov), which included 178 PAAD cases and 4 control samples. The RNA-seq data of 167 and eight healthy pancreatic tissues samples were obtained from the GTEx and ANTE databases, respectively. The RNA-seq data and survival information for PAAD patients from two countries (PACA-AU: 455 cases; PACA-CA: 455 cases) were downloaded from the International Cancer Genome Consortium (ICGC) database (ICGC Data Portal, RRID:SCR_021722) (https://dcc.icgc.org/). Three PAAD-related datasets (GSE62452, GSE15471, and GSE28735) were downloaded from the Gene Expression Omnibus (GEO) database- (GEO, RRID:SCR_005012) (https://www.ncbi.nlm.nih.gov/geo/). The ICGC and GSE62452 datasets were used as external validation datasets for the prognostic model, and the GSE15471, GSE28735 datasets, as well as the ANTE-normal cohorts, were used to validate the expression levels of the prognostic model genes. The 1986 zinc finger protein family genes were extracted from the universal Protein Resource UniProt database (universal Protein Resource, RRID:SCR_002380) UniProt database (https://www.uniprot.org/) (The UniProt Consortium, 2017).

2.2 Analysis of differentially expressed ZNF protein family genes in PAAD

According to the previous literature, the differential analysis was performed between 178 PAAD samples from the TCGA database and 171 normal samples from the GTEx and TCGA databases using the "limma"package (limma, RRID:SCR_010943) (http://bioinf.wehi.edu.au/limma/) in R (Ritchie et al., 2015; Wen et al., 2020). The principal component analysis (PCA) plot in the TCGA-PAAD and GTEx-normal cohorts between the case and normal samples were displayed in Supplementary Figure S1, indicating an excellent distinction. In which, genes with adjusted p-value (adj. p) < 0.05 and-|log2 (fold change)|>1 were considered significantly expressed (Wang et al., 2022). Intersections of differentially expressed genes in PAAD and 1936 PAAD-related ZNF family genes in the UniProt database were used as differentially expressed ZNF family genes (DE-ZNFs). Results were visualized using a heatmap and a volcano plot.

2.3 Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on the DE-ZNFs using clusterProfiler (ClusterProfiler, RRID:SCR_016884) (http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) (Yu et al., 2012). The GO system consists of three components: biological process (BP), molecular functions (MF) and cellular components (CC). KEGG (KEGG, RRID:SCR_012773) (https://www.kegg.jp) (http://www.genome.jp/kegg/-) is a biological systems database that integrates genomic, chemical and systemic functional information (Kanehisa et al., 2008). The enrichment results were visualized by withusing the "ggplot2" package (ggplot2, RRID:SCR_014601) (https://cran.r-project.org/web/packages/ggplot2/index.html).

2.4 Construction of the prognostic signature

Using TCGA as the training set, univariate Cox proportional hazards regression analysis was performed on DE-ZNFs to screen for the genes that were significantly associated with overall survival (OS) outcomes in the TCGA-PAAD training set ($p < 0.01$). Then, patients were divided into high and low expression groups according to expressions of DE-ZNFs from univariate Cox analysis for KM survival analysis to obtain the DE-ZNF with significant different survival rate between the 2 expression groups ($p < 0.01$). These DE-ZNFs were subjected to multivariate Cox regression analyses to identify suitable ZNF - related genes to construct the model. Model genes were screened to calculate the risk scores. A prognostic risk score was calculated for each patient using the formula: risk score = $\beta 1\ X1 + \beta 2\ X2 + \ldots\ldots + \beta n\ Xn$, whereby β represents the coefficient, X represents prognostic gene expressions and n represents the number of genes. The median risk score was used as the cut-off value to divide the TCGA-PAAD patients into the high and low-risk group. Then, the K-M survival curve was constructed after which, the log-rank test was used to assess survival differences between the risk groups. Sensitivity and specificity of the prognostic model were assessed by ROC analysis and AUC values indicated discrimination. Effectiveness of the prognostic model was validated using the ICGC and GSE63452 datasets.

2.5 Survival and risk score analyses

Stratified analysis was performed to establish the correlations between high and low risk groups and survival outcomes in patients with different clinico-pathological characteristics (age >65, age ≤ 65, female, male, M0, T3-T4, stage I-stage II, stage III-stage IV; G1/G2, G3/G4; race-white). To establish the correlations between clinico-pathological characteristics and risk score, the 82 samples with clinical information (stage, age, gender, grade, race and TMN) in the in TCGA-PAAD training set were extracted and compared between subgroups in each clinico-pathological characteristics. The Wilcoxon test was used for comparisons between groups while the Kruskal-Wallis test was used for comparisons among groups.

2.6 Assessment of independent prognostic value

Univariate and multivariate Cox regression analyses were conducted to establish whether the ZNF-related risk score can be used as an independent predictor of OS for PAAD patients. Stage, age, gender, grade, race, T stage, M stage, N stage and risk score were used as the covariates. Clinical factors with $p < 0.05$ after the two cox analyses were considered to be independent prognostic factors that were used to establish the prognostic model. A prognostic nomogram for assessing the 1-, 3- or 5- year survival probability for PAAD patients was established using the "rms" package (RMS, RRID:SCR_007415) (http://www.rms.org.uk/) (Kandimalla et al., 2020). The C-index and calibration curves of the nomogram were used to calculate the discrimination and calibration between the nomogram predicted value and the true survival.

2.7 Analysis of the ZNF gene family signature

To investigate the biological processes that are relevant to the ZNF gene family, first, we determined the correlations between all ZNF family genes and risk scores (Pearson $|R| > 0.4$). The "corrr" package was used for correlation network construction of the obtained genes, while the "clusterProfiler" package was used for GO and KEGG function enrichment analyses of the correlation genes. To establish the differences in immune cell infiltrations, the proportions of 22 immune cell types in the high- and low-risk groups were calculated using the CIBERSORT algorithm (Chen et al., 2018). The "ggplot2" package was used to draw violin diagrams to present the comparison results. To establish the risk score-associated inflammatory activities, the "correlogram" package was used to investigate the correlations between 7 metagenic clusters (HCK, IgG, Interferon, LCK, MHC-I, MHC-II, and STAT1) and risk scores. Relationships between ZNF-related prognostic genes and immunotherapeutic responses were determined by calculating the differences in tumor mutational load (TMB), neoantigen number, clonal neoantigen number and subclonal neoantigen number between the high and low-risk groups.

2.8 Regulatory mechanisms of the prognostic genes

Correlation coefficients between prognostic factor expression levels and their methylation levels were determined using the Pearson's correlations method. Differential analysis was performed on 178 PAAD samples and 4 normal samples using the "limma" package to obtain differentially expressed miRNAs and lncRNAs. Combined with expression trends of prognostic genes, the competing endogenous RNA (ceRNA) regulatory network was constructed based on the lncRNA-miRNA-mRNA regulatory mechanism.

2.9 Validation of prognostic gene expression

In order to verify the expression of the prognostic genes, the TCGA-PAAD cohorts were analyzed and compared with the normal individuals both in the GTEx (with the normalization tool of RLE from DESeq2 R package to obtain normalized count data) and ANTE datasets by the wilcox. test as well as GSE28735 and GSE15471. Immunohistochemical results of prognostic genes in PAAD tissues were searched using Human Protein Atlas (HPA,RRID:SCR_006710) (HPA: https://www.proteinatlas.org/).

2.10 Cell RT-qPCR validation

Four strains of pancreatic cancer cells were cultured to establish a group of normal pancreatic epithelial cells as the control group for cytological verification of model genes. Cells were cultured in 10% FBS complete medium to enter the log phase. Then, RNA extraction was performed using a kit (Tiangen, cat#DP430). cDNA synthesis from the extracted RNA was performed using the iScript™ cDNA Synthesis Kit (BIO-RAD, cat#1708891). Cell RT-qPCR validation

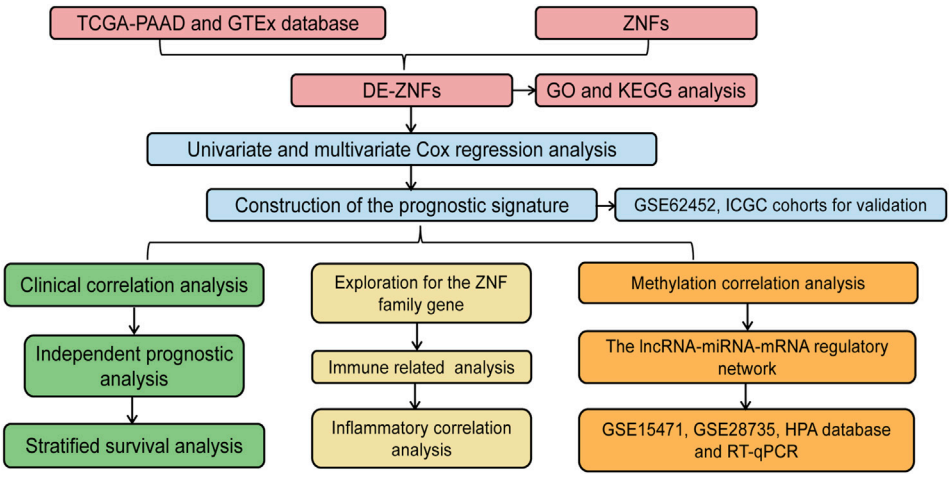


FIGURE 1
Workflow diagram of this paper.

amplification was performed in a 10 ul system using the iTaq™ universal SYBR® Green SupermixRNA (BIO-RAD, cat#1725121). The experiments were conducted in triplicates, and $p < 0.05$ was set as the threshold for statistical significance (Supplementary Figure S2). The PANC-1, RRID: KCB 200809YJ, BxPC-3, RRID: KCB 200428YJ, SW 1990, RRID: KCB 2012113YJ cells, were acquired from the Kunming Institute of Zoology, Chinese Academy of Sciences. The ASPC-1, RRID: TCHu 8, cells were procured from the cell bank of the Chinese Academy of Sciences while. HPDE6-C7, RRID: BFN60807571 cells were obtained from, Qingqi (Shanghai) Biotechnology Development Co. Ltd.

3 Results

3.1 Identification of DE-ZNFs and functional analysis in PAAD

The workflow for this study is shown in Figure 1. A total of 407 DE-ZNFs (150 were upregulated and 257 were downregulated) were identified between the PAAD and normal control tissue samples (Figures 2A, B). The 407 DE-ZNFs were found to be enriched in 226 BPs, 10 CCs, 61 MFs, and 18 KEGG signaling pathways. These included biological processes such as protein autoubiquitination, intracellular receptor signaling pathway, protein polyubiquitination (Figure 2C), and signaling pathways such as herpes simplex virus 1 infection, Th17 cell differentiation, and NF-kappa signaling pathway (Figure 2D).

3.2 Construction of the prognostic model for PAAD

Univariate Cox regression analysis revealed 36 DE-ZNFs that were significantly associated with OS (Supplementary Figure S3). The significantly expressed genes were subjected to multivariate Cox regression analysis to construct the prognostic model. The forest map was used for visualization (Figure 3A). The prognostic model exhibited the best performance when 10 DE-ZNFs were included. The risk score for each sample was calculated based on expression levels of the 10 prognostic genes. Risk score = ZNF185* 0.340812 + DEF8 * −0.91561 + ZMAT1 *-0.77978 + PRKCI * 0.6138 + SP110* 0.70712+U2AF1L4 * −0.57379 + RTP4 * 0.351366 + CXXC1 *1.352264 + RMND5B * −0.96994 + SERTAD2 * 0.373257. Based on the median risk score, PAAD samples were divided into high and low-risk groups. The Kaplan–Meier curve revealed that samples in the high-risk group exhibited worse OS outcomes than those in the low-risk group (Figure 3B). The risk curve and scatter plot were generated to show the risk score and survival status for each PAAD sample. The risk coefficient and mortality in the high-risk group were higher than those in the low-risk group (Figure 3C). A heat map of the 10 prognostic gene expression profiles in PAAD samples revealed that DEF8, RMND5B, CXXC1, ZMAT1, and U2AF1L4 were highly expressed in the low-risk group, while RTP4, ZNF185, PRKCI, SERTAD2 and SP110 were highly expressed in the high-risk group (Figure 3D). The ROC curves were plotted at time nodes of 1, 3, and 5 years. The AUCs of the ROC curves were all greater than 0.7, indicating a good efficacy of the prognostic model (Figure 3E). Validation of the prognostic model was performed using the GSE62452 and ICGC datasets, and the results were comparable to those of the training set (Supplementary Figure S4).

3.3 Risk score performance

After stratifying the clinical characteristics, there were significant differences in risk scores between the high and low risk groups in Age>65, Age<=65, female, male; M0, T3-T4, stage I-stage II, stage III-stage IV, G1/G2, G3/G4 and race-white (Figure 4A). Correlation analysis of clinico-pathological factors such as stage, age, gender, grade, race, and TMN with prognostic models for the 82 TCGA - PAAD samples revealed significant

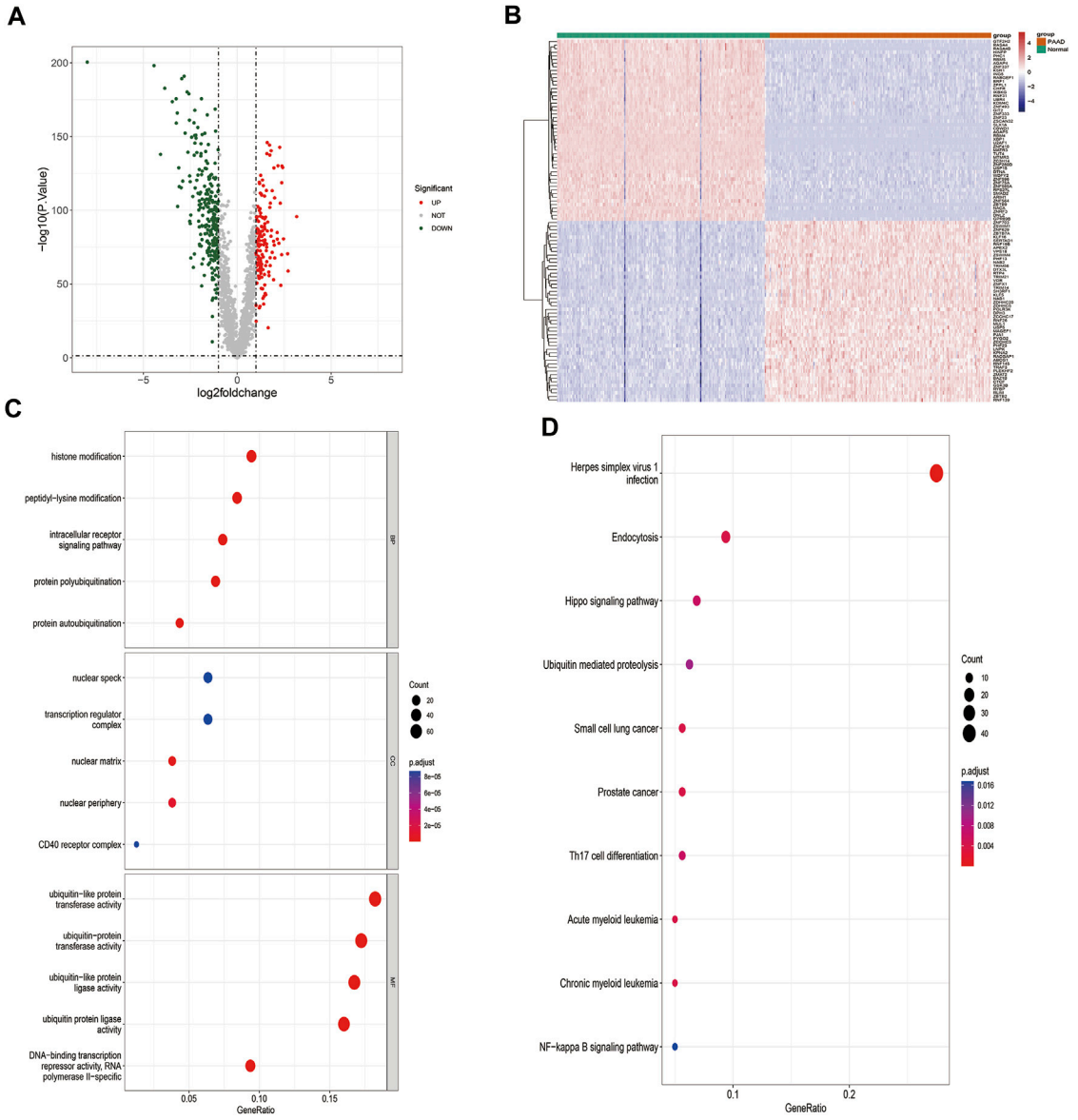


FIGURE 2
Identification and functional analysis of DE-ZNFs. **(A)**. Volcano plot of differentially expressed genes in PAAD-vs-Normal comparison group **(B)**. Heat map of differentially expressed genes, 150 upregulated, 257 downregulated, |log2 (fold change)|>1 and $p < 0.05$ **(C)**. Top 5 GO BP, CC, and MF enrichment results of DE-ZNFs **(D)**. Top 10 enriched KEGG pathways of DE-ZNFs.

differences in risk scores only in grade Figure 4B; Supplementary Figure S5.

3.4 Independent prognostic factors in PAAD

Clinico-pathological factors, such as stage, age, gender, grade, race, TMN, and risk score were subjected to univariate and multivariate Cox regression analyses to establish the independent prognostic factors for PAAD. The risk score was found to be a significant prognostic factor in both Cox analyses ($p \leq 0.05$), suggesting that the risk score was is an independent prognostic factor for PAAD patients (Figures 5A, B).

3.5 Construction of the nomogram

The independent prognostic factors were used to establish a nomogram for prediction of 1-, 3- and 5-year OS outcomes in TCGA - PAAD cohorts. Ten prognostic genes, including DEF8, RMND5B, CXXC1, ZMAT1, U2AF1L4, RTP4, ZNF185, PRKCI, SERTAD2 and SP110 were included in the model (Figure 6A). The points of the factors indicate their corresponding contribution to the survival probability. The actual OS and nomogram-predicted OS outcomes at 1 and 3 years matched well, as shown by the calibration curves (Only 1 patient survived for 5 years, thus, the calibration curve for 5 years was not shown.) (Figure 6B). The AUCs at 1-, 3- and 5- years time nodes were 0.796, 0.725 and 0.826, respectively,

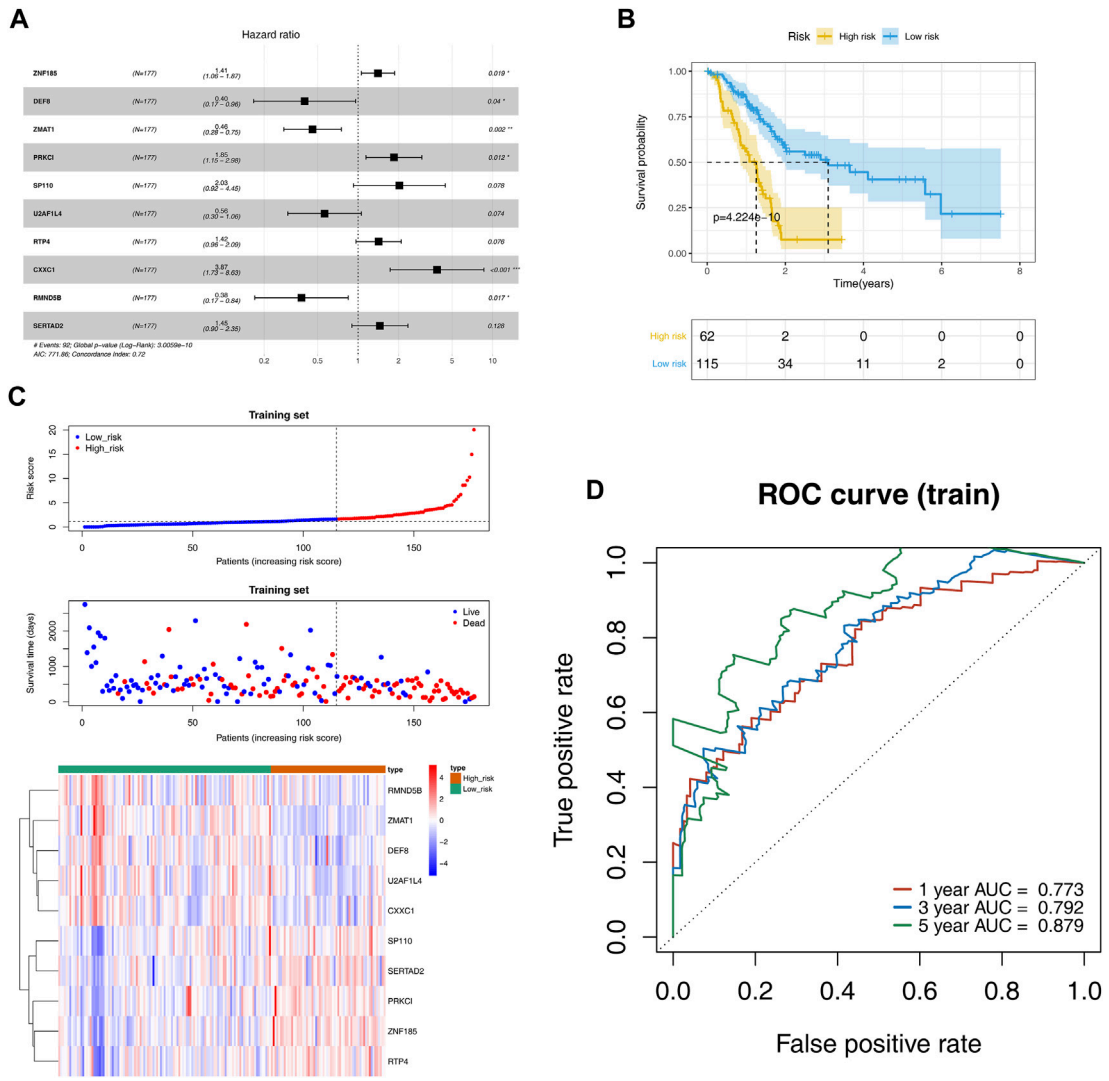


FIGURE 3
Evaluation and validation of the prognostic risk models. **(A)**. Cox regression analysis forest plot shows that 10 DE-ZNFs were used as parameters to construct the best prognosticis model **(B)**. OS survival curves showing thes survival probabilities of high and low risk groups **(C)**. The scatter plot of the risk score and survival time as well as heatmap of gene expression for each PAAD sample in high and low risk groups, which were sorted from left to right according to the risk score **(D)**. ROC curves of the prognostic model at the 1-, 3-, and 5-year time nodes.

revealing that the predictive ability of the nomogram was accurate (Figure 6C).

3.6 Biological processes of ZNF family gene signatures

The 115 ZNF family signatures that were closely associated with risk scores were identified by correlation analyses (Figure 7A). There were positive correlations between most of the ZNF family signatures and the risk scores (Figure 7B). The 115 ZNF family signatures were enriched in 56 BP, 28 CC, 16 MF, and 8 KEGG signaling pathways, which were significantly associated with epidermis development, protein processing, keratinocyte proliferation biological functions and protein digestion and absorption, insulin secretion, as well as ECM-receptor interaction signaling pathways (Figures 7C, D).

3.7 Immune cell infiltration landscape

The proportions of 7 immune cells (naïve B cells, memory B cells, plasma cells, resting NK cells, monocytes, activated dendritic cells and neutrophils) differed between the risk groups ($p < 0.05$; Figures 8A, B). These results reveal a dysregulated tumor immune microenvironment. Correlation analysis of seven meta genic clusters (HCK, IgG, interferons, LCK, MHC-I, MHC-II, and STAT1) with risk scores showed that the risk score was negatively correlated with IgG and LCK and weakly positively correlated with HCK, interferons, MHC-I, MHC-II, and STAT1 (Figure 8C).

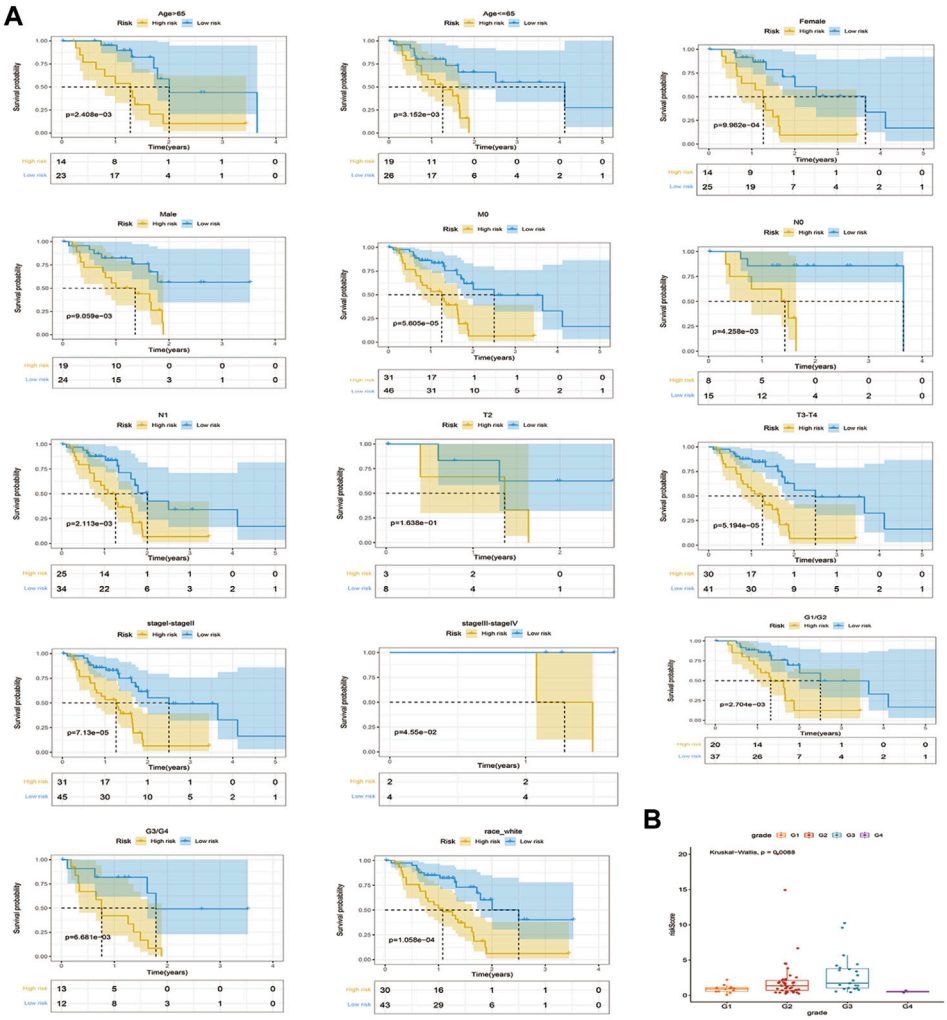


FIGURE 4
Stratified survival analysis of risk scores and correlation analysis of clinicopathological characteristics. **(A)**. K-M curves of PAAD patients in high and low risk groups of Age > 65, Age <= 65, female, male, M0, T2, T3-T4, stagel-stage II, stage III-stageIV, Race white, G1/G2, and G3/G4 **(B)**. Correlations between stages of grade and risk models.

3.8 Relationships between immunotherapy and ZNF family gene signatures

Relationships between ZNF family signatures and immunotherapeutic responses were analyzed. The number of TMB was significantly high in the low risk group, compared to the high risk groups ($p < 0.05$; Figure 9A). Exclusion and PD-L1 were significantly differentially expressed between the high and low risk groups ($p < 0.05$), while TIDE was not significantly differentially expressed between the groups (Figure 9B).

3.9 Regulatory mechanisms of prognostic genes

Analysis of correlations between prognostic genes and their methylation levels arevealed significant negative correlations between RTP4 and SP110 and their methylation levels (Figures 10A, B; Supplementary Figure S6). There were 66 differentially expressed miRNAs between the 178 PAAD samples and 4 normal samples, of which 14 were upregulated while, 52 were downregulated (Figure 10C). Moreover, there were; and 199 differentially expressed lncRNAs, of which 49 were upregulated while 150 were downregulated (Figure 10D). Among the 10 prognostic genes, ZNF185, PRKCI, RTP4, and SERTAD2 were upregulated while DEF8, ZMAT1, SP110, U2AF1L4, CXXC1, and RMND5B were downregulated. We sequentially extracted three expression matrices from mRNA/miRNA/lncRNA. The expression data for 10 prognostic genes (up-4, down-6), differentially expressed miRNAs (up-14, down-52) and differentially expressed lncRNA (up-49, down-150) were extracted. Finally, based on potential regulatory relationships in mRNA/miRNA/lncRNA, we constructed a ceRNA network consisting of 5 prognostic, 7 miRNAs and 35 lncRNAs (Figure 10E).

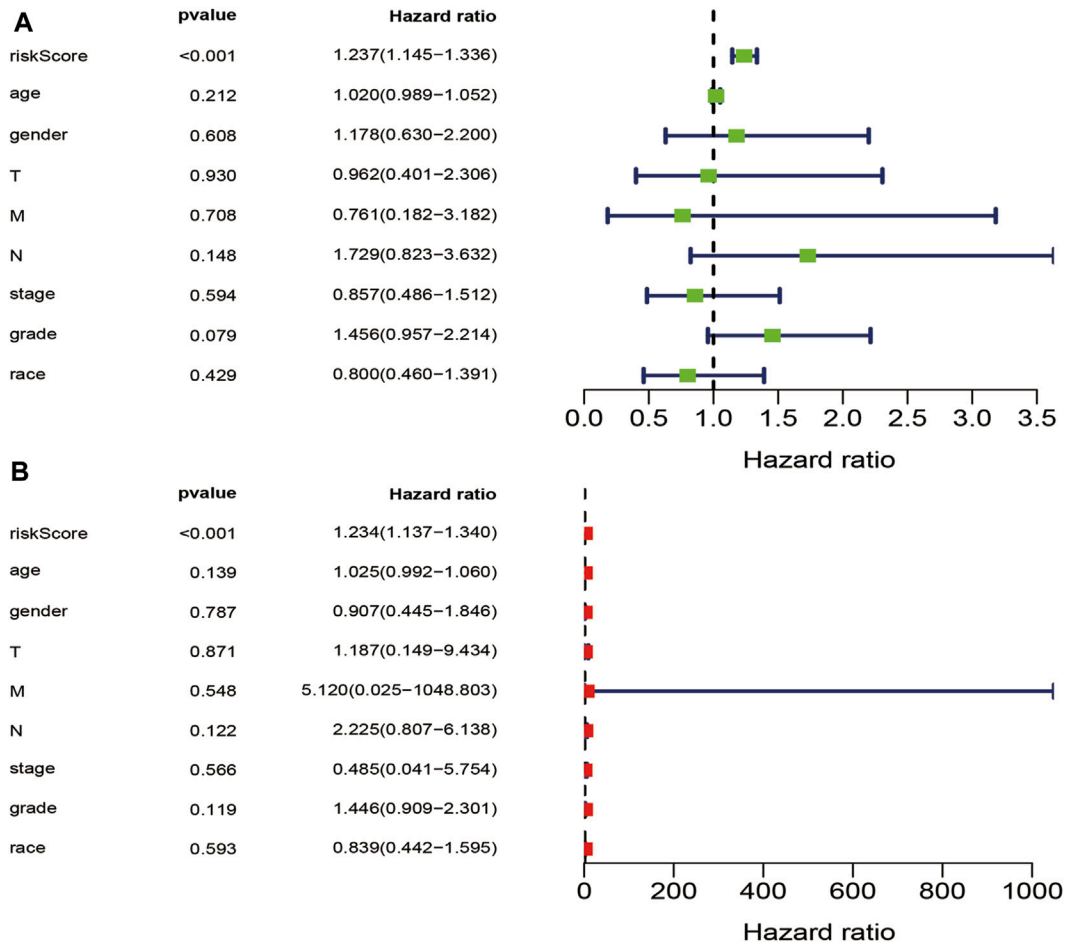


FIGURE 5
Univariate and multivariate independent prognostic analysis. **(A)**. Univariate Cox independent prognostic analysis of stage, age, gender, grade, race, T stage, M stage, N stage, and riskScore **(B)**. Multivariate Cox independent prognostic analysis of stage, age, gender, grade, race, T stage, M stage, N stage, and riskScore.

3.10 Expressions of the prognostic genes

The expression levels of ten prognostic genes (ZNF185, PRKCI, RTP4, SERTAD2, DEF8, ZMAT1, SP110, U2AF1L4, CXXC1, and RMND5B) were examined in the datasets TCGA, GSE28735, and GSE15471. The results between the TCGA-PAAD and GTEx-normal data after normalization exhibited that among these genes expression was significantly different between the two groups (Figure 11A). Similarly, the expressions of eight risk genes between TCGA-PAAD and ANTE-normal cohorts had apparent differences except for DEF8 and RMND5B (Figure 11B). However, it was noticed that the expression trends of SERTAD2 and U2AF1L4 were opposite in two cohorts above. For the two GEO datasets, it was showed that the expression patterns of various prognostic genes were similar to those in the TCGA cohort except for SERTAD2 and U2AF1L4 as well. Furthermore, it was noteworthy that the SP110 expression was upregulated in PAAD samples compared to controls, while it expressed higher in GTEx-normal and ANTE-normal samples (Figures 11C, D).

Analysis of protein levels of prognostic genes in PAAD and normal tissues in the HPA database further revealed that SERTAD2 (in Cytoplasmic, membranous) proteins were expressed at significantly higher levels in tumor tissues than in normal controls, confirming the expression results in (Figures 11A, D rather than that in (Figures 11B,C; While the expression of U2AF1L4 proteins (in Nuclear) was not significantly different (Figure 12), and no immunohistochemical result was available for SP110 protein in the HPA database, indicating that the expression patterns of U2AF1L4 and SP110 remained to be further studied. Besides the expression levels of PRKCI (in Cytoplasmic, membranous), ZMAT1 (in Cytoplasmic, membranous), CXXC1 (in Nuclear), DEF8 (in Cytoplasmic, membranous), RTP4 (in Cytoplasmic, membranous), RMND5B (in Nuclear), and ZNF185 were confirmed and in accordance with the results of public datasets.

3.11 PADD cell validation

The HPDE6-C7 cell line was used as the control group, and pancreatic carcinoma *in situ* and metastatic cancer cell lines were used as the experimental group to investigated the expression of

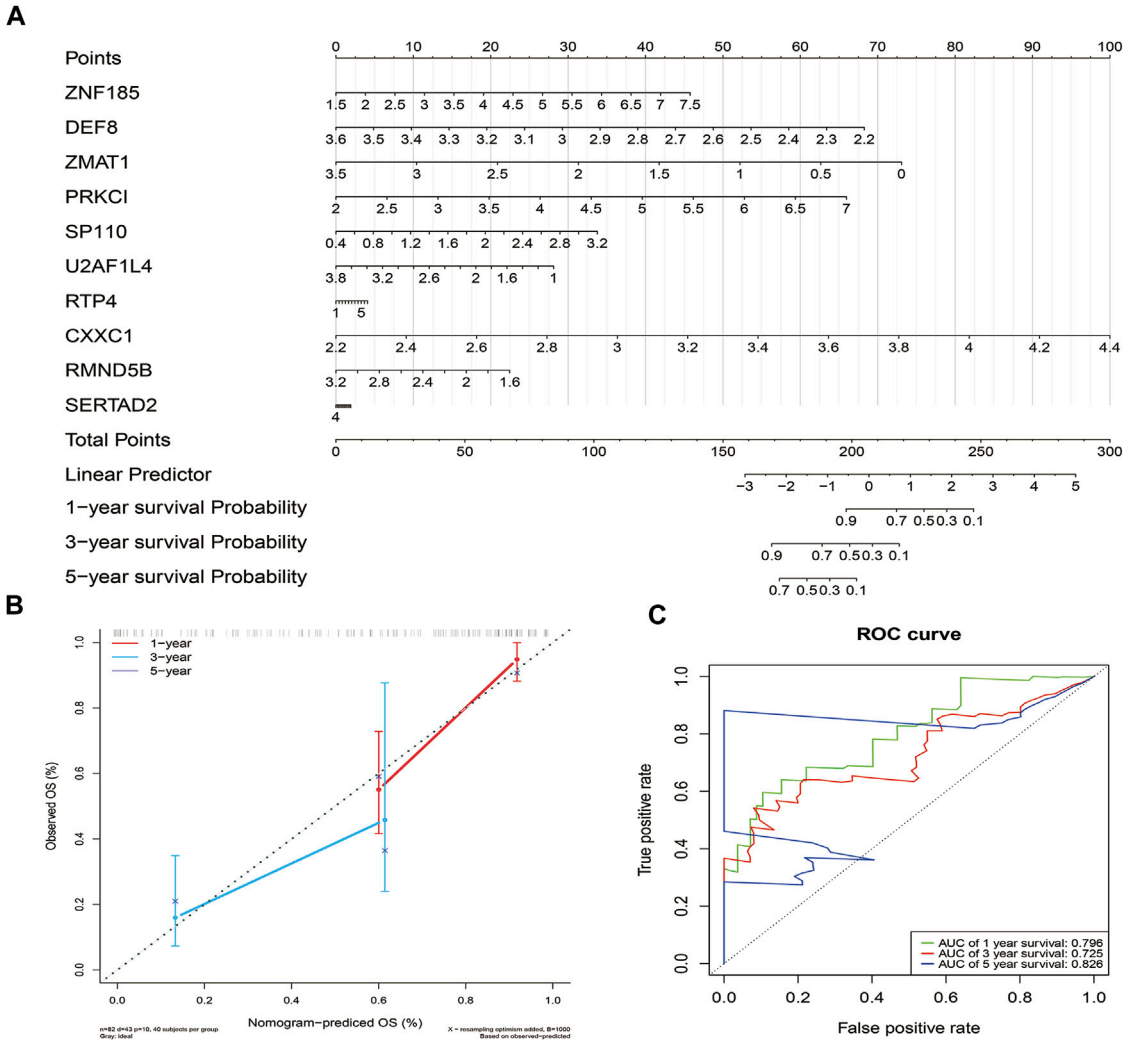


FIGURE 6
Construction and validation of the nomogram. **(A)**. The nomogram based on the 10 prognostic genes of the risk score **(B)**. Calibration curve of the nomogram.The diagonal dotted line slope is 1. **(C)** ROC curves of the nomogram.

10 risk model genes, where the expression levels in the control group were basically the same. It could be seen that RTP4, SERTAD2, and SP110 were all highly expressed in the four pancreatic cancer cell lines, which were in concordance with publicly available GEO data, suggesting that they are closely related to the occurrence and metastasis of the pancreas at the cellular level. The expression of DEF8, RRKCI, U2AF1L4 were distinctly higher in PANC-1, BXPC-3 and Aspc-1 metastatic cell lines. Moreover, the expression levels of the ZNF185, CXXC1, RMND5B, and ZMAT1 genes changed significantly with different cell lines (Figure 13).

4 Discussions

Pancreatic cancer is a highly invasive malignant tumor, that invades many organs, including the stomach, common bile duct, duodenum, superior mesenteric vein, and celiac artery in the terminal stages. Extensive lymphatic metastasis is often accompanied by nerve sheath metastasis, which results in extremely high fatality rates. Pancreatic cancer development involves complex biological processes. It is closely associated with cell phenotypes that are related to autophagy, histone methylation, hypoxia tolerance, and apoptosis among others and is crucial to improve the treatment plans and prognostic outcomes of PAAD (Gupta et al., 2021; Liu et al., 2021; Wang Y. et al., 2021). Studies on correlations of biological targets for early pancreatic cancer diagnosis are still in early stages. In this study, survival and RNA-seq data for PAAD patients from TCGA and GSE databases were downloaded, and the differentially expressed genes and ZNF data obtained and analyzed to construct a prognostic risk model. The prognostic risk model including 10 independent risk factors, such as ZMAT1, PRKCI, ZNF185, SERTAD2, CXCC1, U2AF1L4, RTP4, SP110, and DEF8, was established and validated in GEO and ANTE datasets. The q-PCR analysis showed that the Panc-1 cell line, derived from human ductal cell carcinoma and the pancreatic adenocarcinoma Bxpc-3 basically showed a higher expression level, while the expression in the metastatic carcinoma

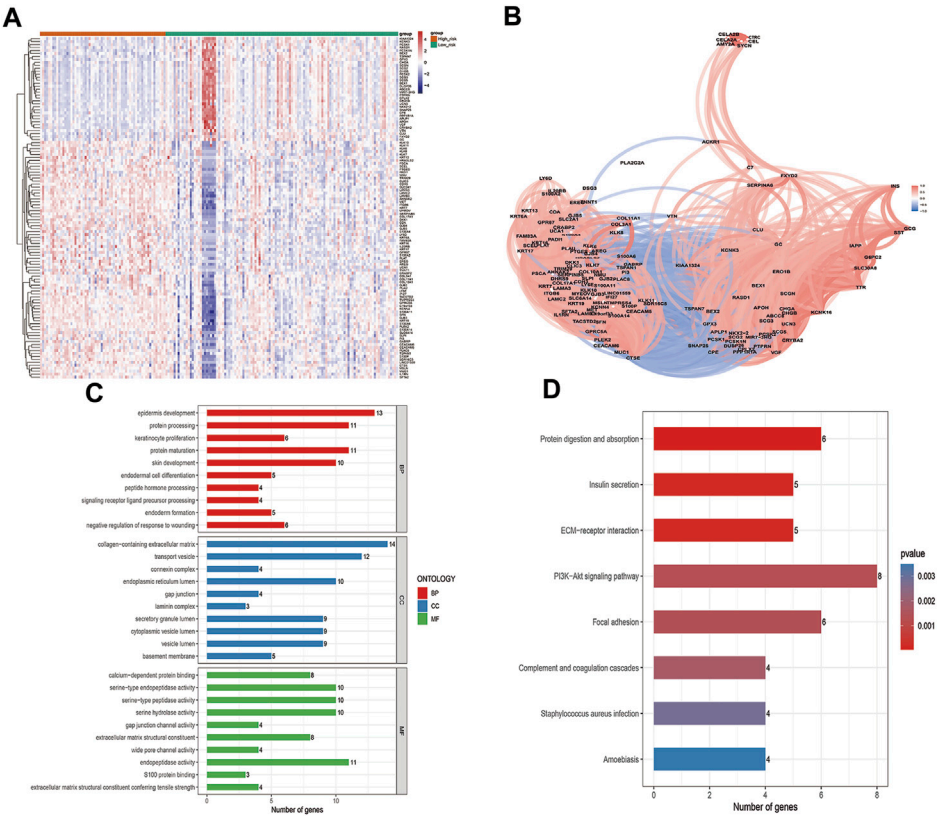


FIGURE 7
Biological processes involved in ZNF family gene signaling. **(A)**. Heatmap of 115 genes that were closely related to the risk score **(B)**. Correlation network of 115 genes that were closely related to the risk score **(C)**. GO enrichment analysis results of 115 genes that were closely related to the risk score, top 10 BP, CC and MF enriched terms **(D)**. Eight enriched KEGG pathways in which the115 genes that were closely related to risk score were enriched.

cells of pancreatic cancer was different. It was found that RTP4, SERTAD2, and SP110 were significantly expressed in all pancreatic cancer cells, including those of pancreatic ductal carcinoma, pancreatic adenocarcinoma and metastatic carcinoma, providing a reliable basis for subsequent elucidation of pancreatic cancer pathogenesis.

It has been reported that ZMAT1 induces p21 expressions via the SIRT3-p53 signaling pathway to inhibit pancreatic cancer cell proliferation and induce S/G2 cell cycle arrest (Ma et al., 2022). PRKCI-mediated ablation of pancreatic acinar cells resulted in p62 aggregation and loss of autophagic vesicles. Pancreatic PRKCI knockdown significantly increased pancreatic immune cell infiltrations acinar cell DNA damage, apoptosis, and promoted KrasG12D mediated pancreatic intraepithelial neoplasia, promoting tumor growth (Inman et al., 2022). Pancreatic cancer chemotherapy tolerance is associated with poor survival and prognostic outcomes of patients. Pancreatic cancer HEAT repeat-containing protein 1 (HEATR1) deficiency can affect pancreatic cancer chemotherapy sensitization via the upregulation of ZNF185 (Fang Y. et al., 2020). In the prognostic risk model of this study, ZMAT1 is a low-risk gene for pancreatic cancer, while PRKCI and ZNF185 are high-risk genes for pancreatic cancer.

A part from the above three genes, the other independent risk factors have not been reported to be pancreatic cancer-related. However, basic experiments and high-throughput data analysis of other solid tumors confirmed that the other risk model genes are involved in tumor pathogenesis. This evidence provides the author with more reliable information and enthusiasm for exploration. Elucidation of the importance of zinc finger proteins in pancreatic cancer mayinform on novel approaches for identification of treatment targets and overcoming chemotherapeutic resistance among others. As transcriptional regulators of the largest mammalian system, zinc finger proteins are involved in regulation of tumor mechanisms via multiple pathways. The central member of the family of TLS polymerases (REV1) upregulates SERTAD2 expressions in a Rad18-dependent manner, thereby enhancing lung cancer development (Chen et al., 2022). Wei Wang et al. performed whole-genome sequencing of invasive and *in situ* patients with cutaneous squamous cell carcinoma and found that DEF8 was highly enriched in invasive cutaneous squamous cell carcinoma (Wang et al., 2018). Through bioinformatics and meta-analysis, Shenying Fang et al. identified 3 melanoma risk-related genes, including DEF8, among 330 unique melanoma genes (Fang S. et al., 2020). In bioinformatics studies of other related tumors, it was found that U2AF1L4 a prognostic factor for renal cancer, especially renal clear cell carcinoma (Wang B. et al., 2021). Kuo-Wei Chang et al. performed whole-exome sequencing of a p53-deficient murine

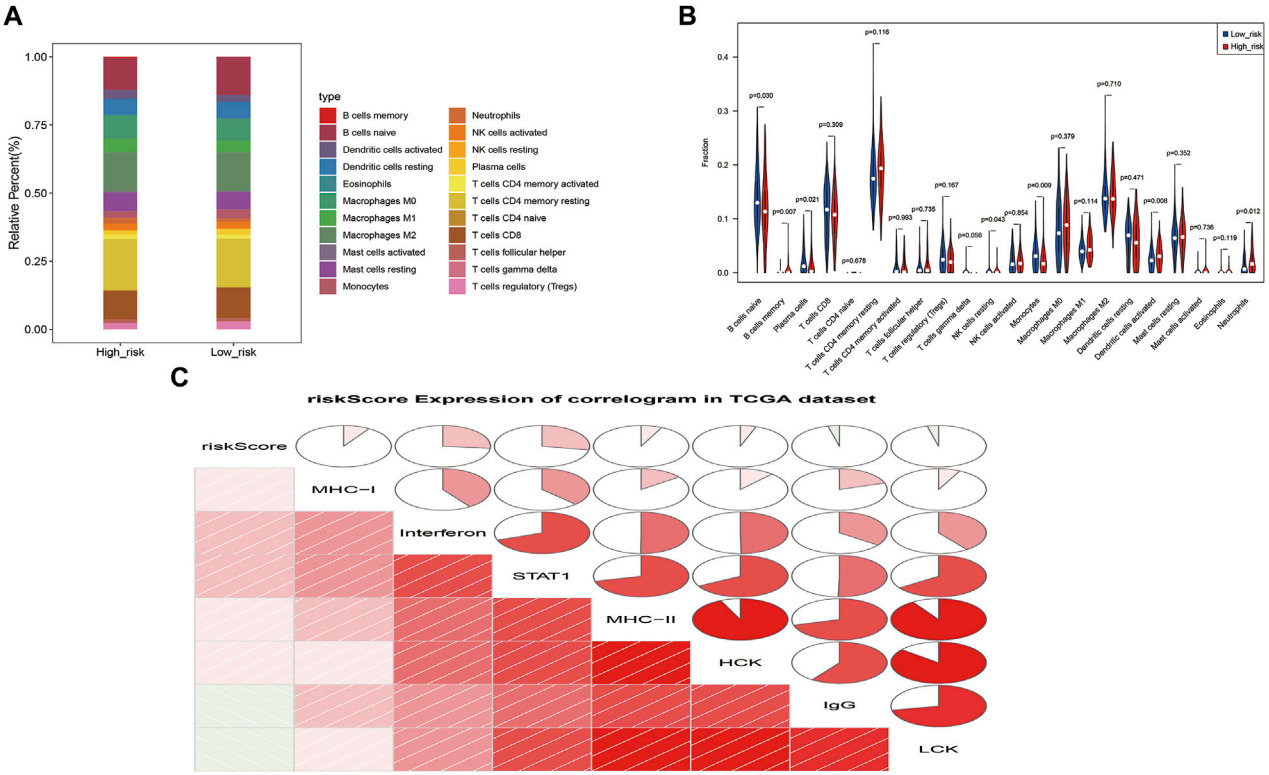


FIGURE 8
Correlation analysis of ZNF family gene signaling with cellular immunity and inflammation. **(A)**. Heat map of the proportions of 22 immune cells in high and low risk groups **(B)**. Violin plot of the infiltration abundance of 22 immune cells in the high and low risk groups **(C)**. Correlation plot of the risk score and seven metagene clusters.

oral cancer cell line and found that SP110 exhibited comparable mutations to those in chemical carcinogenesis-related tongue cancer cell lines in the human TCGA database (Chang et al., 2020).

We intersected the PAAD-related differentially expressed genes in the TCGA database with DE-ZNF to obtain 407 ZNF-related functional genes. The GO and KEGG enrichment analysis revealed that the functional genes were enriched in 226 BP, 10 CC, 61 MF and 18 KEGG signaling pathways. The main enriched related biological signals and pathways were: protein autoubiquitination, intracellular receptor signaling pathway, Th17 cell differentiation, HSV1 infection, and NF-κB signaling pathway. These results show that the functional genes are involved in tumor proliferation, apoptosis, cycle, metastasis, and tumor immunity. Correlation analysis of the ZNF gene family and PAAD showed that the DE-ZNF functional gene risk score was significantly correlation from various tumor immune infiltrating cells and inflammatory cells. In the high-risk group, NK cells, dendritic cells, neutrophils, STAT family, MHC-I, and MHC-II were positively correlated with PAAD pathogenesis, and the immunotherapy target PDL-1 was also significantly different. The α-Enolase (ENO1) specific Th17 cells have specific anticancer effects in PAAD patients, and compared with healthy mucosa, the abundance of Th17 in peripheral blood of tumor patients is low, while the proportion of FOXP3+Tregs is high. The FOXP3+RORγt + Tregs secrete both Th17 and Th2-related pro-inflammatory cytokines, corresponding to elevated Th17- and Th2-mediated immune responses in PDAC patients (Amedei et al., 2013; Chellappa et al., 2016). In solid malignant tumors, such as pancreatic cancer, NF-κB is the main regulatory signaling pathway that promotes malignancy and chemoresistance. Expressions of GPR87 are significantly upregulated in pancreatic cancer and clinical tissues, and activation of the NF-κB signaling pathway promotes pancreatic cancer metastasis (Wang et al., 2017). In enrichment analysis, herpes simplex virus type 1 (HSV-1) infection was specifically proposed. *In vivo* and *in vitro* studies confirmed that pancreatic cancer cells are highly sensitive to HSV1 virus replication, which can be evaluated as an effective treatment scheme. A study on oral squamous cell carcinoma showed that the co-expression gene of zinc finger protein 71 (ZNF71) was mainly enriched in the HSV1 infection pathway (Gayral et al., 2015; Jiang et al., 2022). 1-Methyl-D-tryptophan (1-MT) has been shown to significantly reduces the activities of cancer stem cells. A high abundance of CD133 + and PDL-1 expressions in the tumor immune microenvironment, suppresses NF-κβ and Wnt/β-catenin signaling pathways in tumors, and decreases the abundance of intra-tumor Treg cells (Alahdal et al., 2018). Partially mature dendritic cells in peripheral blood of PAAD patients, significantly enhanced the expressions of CD83, CD40, B7H3, PDL-1, CCR6 and CCR7, decreased the expressions of ICOSL and DCIR, and improved the survival and prognostic outcomes of patients (Tjomsland et al., 2010). The 10 risk models that we analyzed have so far not been reported to be related to tumor

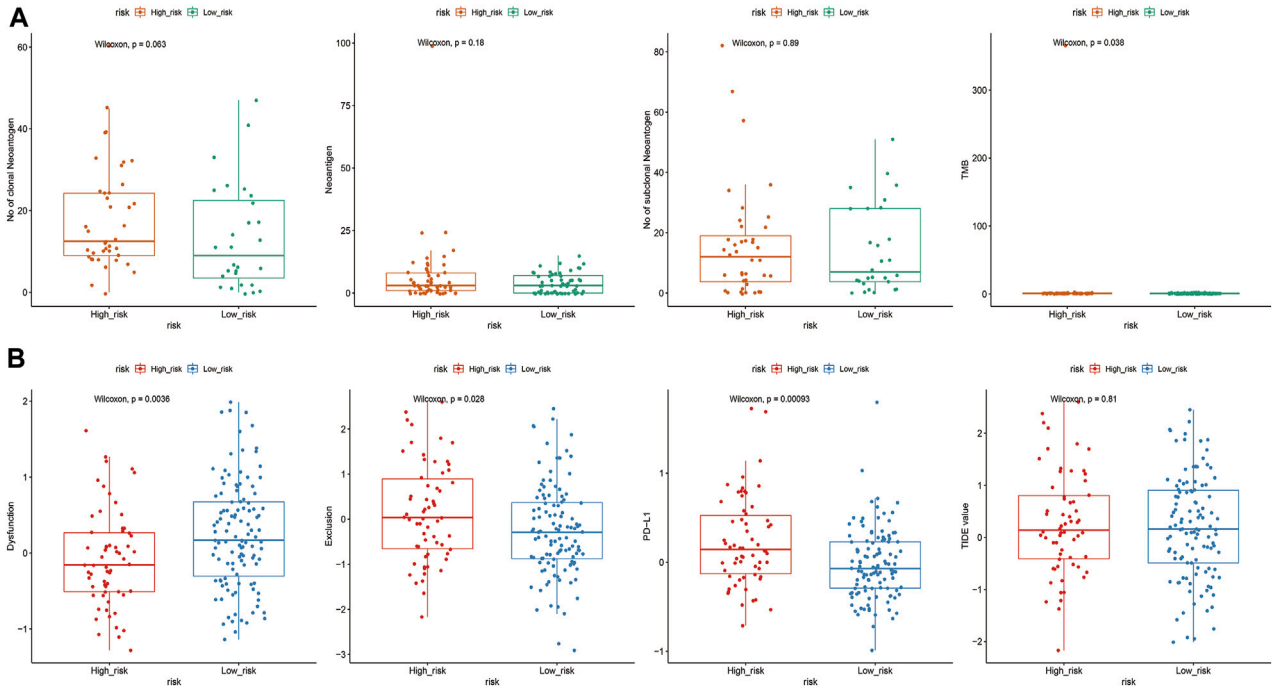


FIGURE 9
Correlation analysis of ZNF family gene risk signals and immunotherapy. **(A)**. Differences in abundance of TMB, neoantigens, cloned neoantigens and subcloned neoantigens between high and low risk groups **(B)**. Expressions of TIDE, Dysfunction, Exclusion and PD-L1 in high and low risk groups.

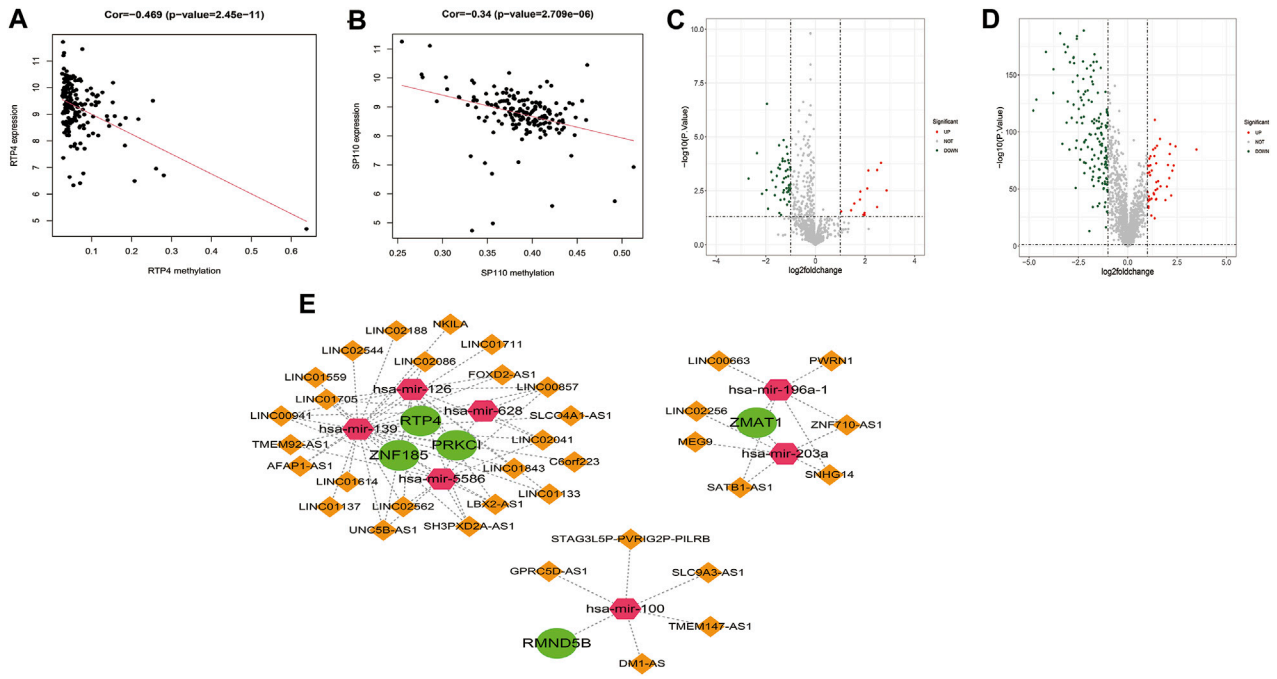


FIGURE 10
Regulatory mechanisms of risk model genes. **(A)**. Scatter plot of the correlation between risk model genes RTP4 and their methylation levels **(B)**. Scatter plot of the correlation between risk model genes SP110 and their methylation levels. **(C)**. Volcano plot of differentially expressed miRNAs in PAAD-vs-Normal comparison group **(D)**. Volcano plot of differentially expressed lncRNAs in PAAD-vs- Normal comparison group **(E)**. The ceRNA regulatory network with 5 risk model genes, 7 miRNAs and 35 lncRNAs. The green circles represent the risk model genes, the pink hexagons represent miRNAs, and the orange diamonds represent lncRNAs.

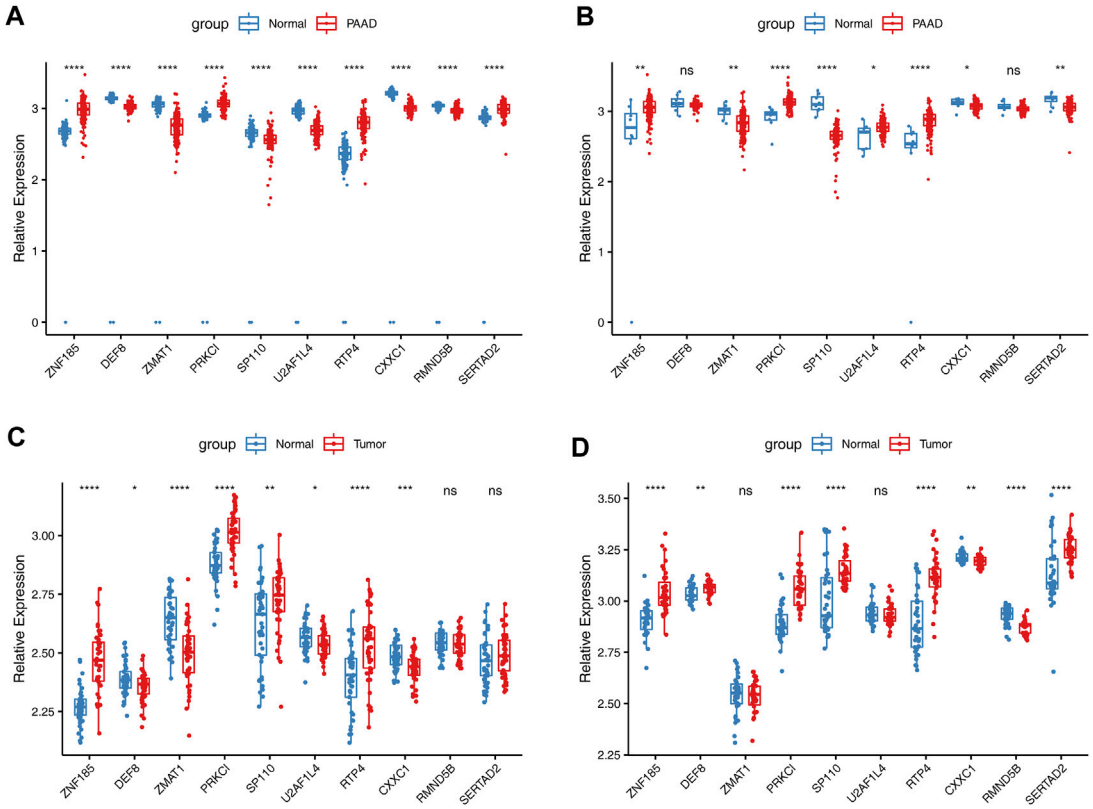


FIGURE 11
Validation of expressions of risk model genes in TCGA and GEO datasets. **(A)**. The expressions of 10 risk genes between PAAD TCGA-PAAD and normal samples GTEx-normal cohorts in TCGA **(B)**. The expressions of 10 risk genes between TCGA-PAAD and ANTE-normal cohorts. The expressions of 10 risk genes in GSE28735 **(C)**. and GSE15471 **(D)** datasets.

immunity and NF-κB signaling pathway, which will provide new insights into studies on pancreatic cancer cells and tissue variations as well as functions. In summary, studies on pancreatic cancer tumor immunity and other related biological activities that are important in development of optimal immunotherapeutic approaches are in the initial stage.

Tumor epigenetics is regulated by protein methylation. Studies on tumor epigenetics are key in gene mutation research and targeted therapy. Uncontrolled methylation leads to changes in chromatin structure, and increased protein synthesis mediates infinite pancreatic cancer progression (Wang S. S. et al., 2021). Mishra NK et al. performed differential expression analyses on PAAD tissue data and normal samples from the TCGA database, and found that most differential CpG island (CpG) sites were hypermethylated in PAAD, and promoter methylation as well as 5′UTR were associated with gene expressions, and most of them were negatively correlated, while gene body and 3′UTR-related methylation were positively correlated (Mishra and Guda, 2017). The Ras-mediated cancers utilize the METTL13-eEF1AK55me2 dimethylation axis to increase the translational output, and enhance protein synthesis to promote pancreatic cancer progression. Human Arginine Methyltransferase 1 (PRMT1) overexpression enhances HSP70 binding and BCL2 mRNA stability via elements of the 3′UTR. Increased HSP arginine methylation of HSP70 regulates cell malignancy and is involved in pancreatic cancer drug resistance (Liu et al., 2019; Wang et al., 2020). These findings are in tandem with ours. Knockout of CXCC1, also known as CxxC Finger Protein-1 (Cfp1), affects cytosine methylation and regulation of histone H3K4 on chromatin structure and function. A DNA methyltransferase (DNMT) inhibitor disrupted the DNMT1/CFP1 complex and enhanced mouse glioma chemosensitivity (Cheray et al., 2014). We also found that RPT4 is closely associated with methylation. In previous studies, RTP4 was shown to regulates prostate cancer via methylation and is regarded as a precise target, whose expression levels can be used to independently predict the prognosis of HER2(+) breast cancer (Laurin et al., 2013; Xu et al., 2019).

The carboxy terminus of LisH (CTLH) complex representing RMND5B can promote tumor maintenance and rapid proliferation under extreme conditions and is associated with EMT and, wnt/β-catenin pathway. Overexpressed RMND5B has been shown to inhibit NKX3.1 factor in prostate cancer to suppress its ubiquitination and nuclear levels so as to promote tumor proliferation (Huffman et al., 2019). In eukaryotic cells, including tumor cells, protein degradation is mainly achieved via the ubiquitin-protease degradation system, and ubiquitin ligase E3 is the key dominant factor in this degradation system. During tumor progression, ubiquitin ligase E3inhibits gene induction, suppresses the expression regulation function of the star tumor suppressor gene pP53, and then mediates tumor occurrence and development through the cell cycle or apoptosis. RMND5B has E3 ligase

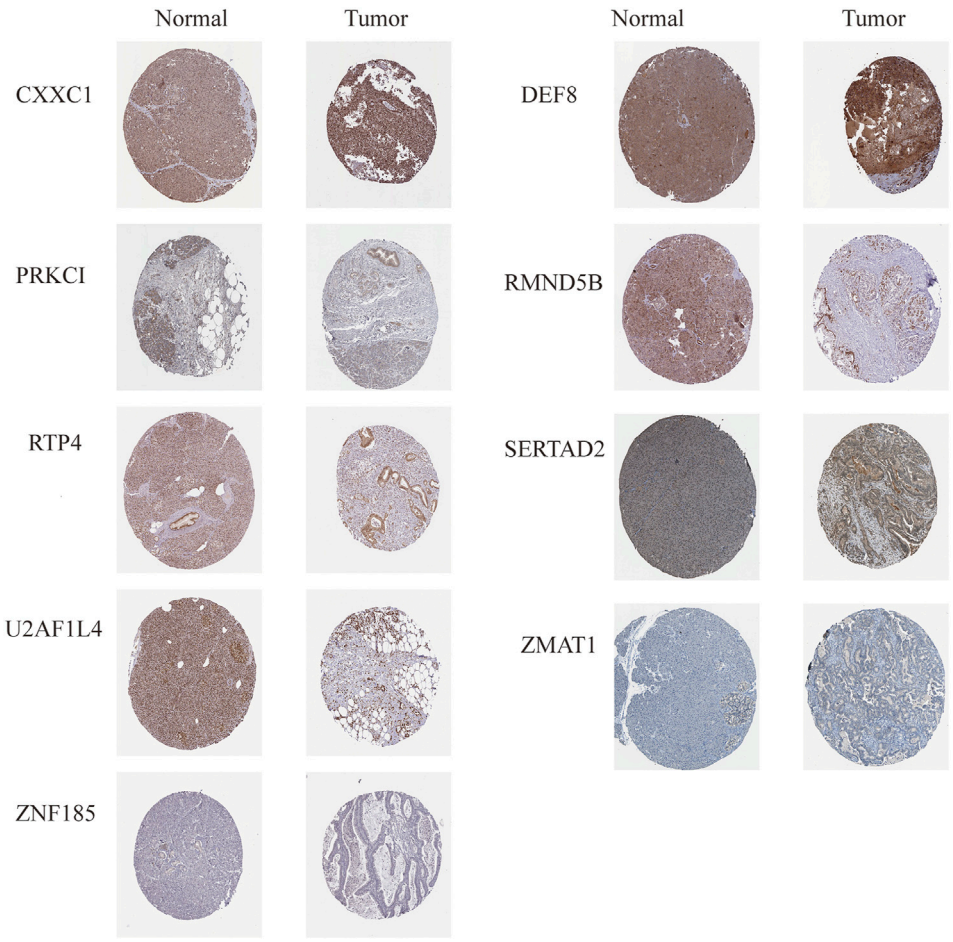


FIGURE 12
Expression levels of 9 risk model genes were validated in HPA database.

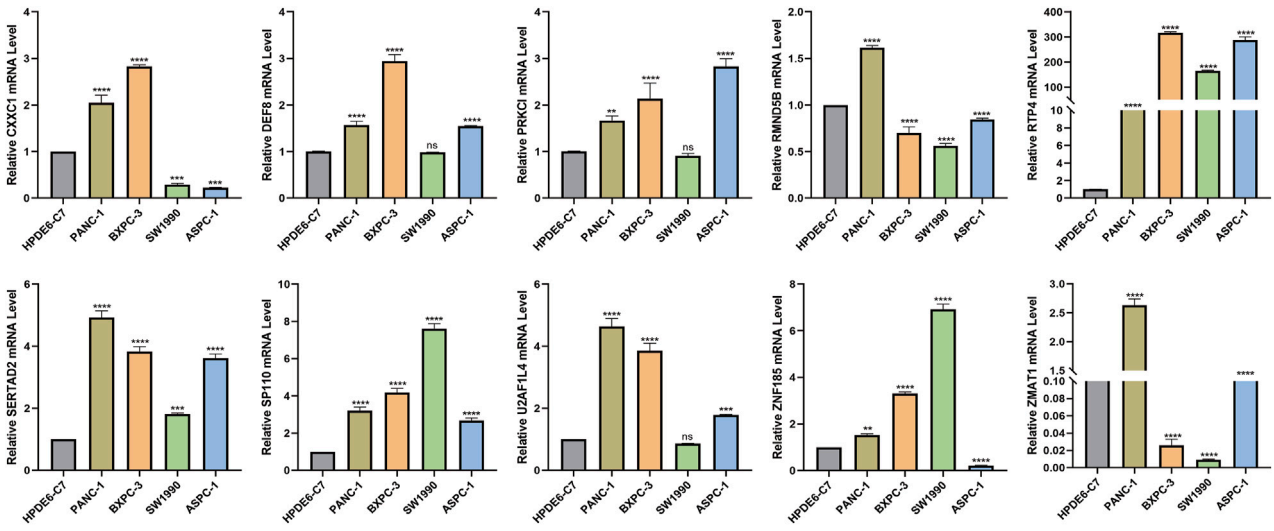


FIGURE 13
Validation of expression levels of 10 risk model genes in pancreatic cancer cells by Cell RT-qPCR validation * represents $p < 0.05$, **, represents $p < 0.01$, *** represents $p < 0.001$, and **** represents $p < 0.0001$.

activities and its overexpression in tumors is critical for cancer cell therapy resistance (Tjomsland et al., 2010). In this study, RMND5B was found to be low risk in the risk model, and Cell RT-qPCR validation revealed that its levels were suppressed in multiple pancreatic cancer cells, implying that it may exist as a tumor suppressor gene in pancreatic cancer, which contradicts the findings from other studies. However, the specific regulatory factors have yet to be determined. Abnormalities of the ubiquitin-proteasome system are key in PAAD pathogenesis, and the ubiquitin-proteasome UCHL5 can promote tumor progression and dry expression depending on involvement of the ELK3 protein (Yang et al., 2022). These findings are in tandem with our enrichment analysis results.

Due to its early metastasis, difficult operation and low survival characteristics, pancreatic cancer is a "lethal" cancer. In this study, we identified the PAAD-related DE-ZNF functional genes, and conducted an in-depth analysis of the possible mechanisms of PAAD. We identified ten risk model genes that can be used as independent prognostic factors for PAAD. Epigenetic modifications include methylation, ubiquitination, tumor immune microenvironment, and ceRNA gene regulatory networks. Our findings provide a novel basis for in-depth assessment of immunotherapy and clinical diagnosis of pancreatic cancer.

Finally, it is worth noting that there are differences in gene expression among SERTAD2, U2AF1L4, SP110, *etc.*, in the four online datasets, cell validation, and immunohistochemistry results. A search of relevant literature in the past decade found that SP110 is a special transcription factor of tumor involved in the carcinogenic regulation of breast cancer and ovarian cancer (Korakiti et al., 2020; Rooda et al., 2020). U2AF1L4 was reported to be involved in renal clear cell carcinoma, but there is a lack of corresponding mechanism research (Wang B. et al., 2021). SERTAD2 has been reported as an oncogene in pancreatic cancer (Zhang et al., 2022). This result is consistent with our cell verification, and believes that cell experiments are more reliable. The mechanism exploration of pancreatic cancer is a huge challenge for clinical research. The relevant data from the database provides research direction, but it still needs long-term exploration and large sample size research support to obtain accurate basic data.

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

Conceptualization:HZ; Methodology: TL; Formal analysis and investigation: LZ,D DT, WZ, LL, WJ; Writing–original draft preparation: RL, DT; Writing–review and editing: RL, LZ, DT. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the "National Natural Science Foundation of China", project number: 81460132. Yunnan Provincial Clinical Center of Hepato-Biliary-Pancreatic Diseases (No specific number).

Acknowledgments

I would like to thank HZ and Director Li Tiehan for proposing the design concept of this project and for their guidance during the project. I would like to thank the National Natural Science Foundation of China and the Second Affiliated Hospital of Kunming Medical University for their financial and platform support for this project.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1089023/full#supplementary-material

References

Alahdal, M., Xing, Y., Tang, T., and Liang, J. (2018). 1-Methyl-D-tryptophan reduces tumor CD133+ cells, wnt/β-catenin and NF-κβp65 while enhances lymphocytes NF-κβ2, STAT3, and STAT4 pathways in murine pancreatic adenocarcinoma. *Sci. Rep.* 8, 9869. doi:10.1038/s41598-018-28238-8

Amedei, A., Niccolai, E., Benagiano, M., Della Bella, C., Cianchi, F., Bechi, P., et al. (2013). *Ex vivo* analysis of pancreatic cancer-infiltrating T lymphocytes reveals that ENO-specific Tregs accumulate in tumor tissue and inhibit Th1/Th17 effector cell functions. *Cancer Immunol. Immunother.* 62, 1249–1260. doi:10.1007/s00262-013-1429-3

Chang, K. W., Lin, C. E., Tu, H. F., Chung, H. Y., Chen, Y. F., and Lin, S. C. (2020). Establishment of a p53 null murine oral carcinoma cell line and the identification of genetic alterations associated with this carcinoma. *Int. J. Mol. Sci.* 21, 9354. doi:10.3390/ijms21249354

Chellappa, S., Hugenschmidt, H., Hagness, M., Line, P. D., Labori, K. J., Wiedswang, G., et al. (2016). Regulatory T cells that co-express RORγt and FOXP3 are pro-inflammatory and immunosuppressive and expand in human pancreatic cancer. *Oncoimmunology* 5, e1102828. doi:10.1080/2162402X.2015.1102828

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, Y., Chen, D., Wang, Q., Xu, Y., Huang, X., Haglund, F., et al. (2021). Immunological classification of pancreatic carcinomas to identify immune index and provide a strategy for patient stratification. *Front. Immunol.* 12, 719105. doi:10.3389/fimmu.2021.719105

Chen, Y., Jie, X., Xing, B., Wu, Z., Yang, X., Rao, X., et al. (2022). REV1 promotes lung tumorigenesis by activating the Rad18/SERTAD2 axis. *Cell Death Dis.* 13, 110. doi:10.1038/s41419-022-04567-5

Cheray, M., Nadaradjane, A., Bonnet, P., Routier, S., Vallette, F. M., and Cartron, P. F. (2014). Specific inhibition of DNMT1/CFP1 reduces cancer phenotypes and enhances chemotherapy effectiveness. *Epigenomics* 6, 267–275. doi:10.2217/epi.14.18

Fang, Y., Han, X., Li, J., Kuang, T., and Lou, W. (2020). HEATR1 deficiency promotes chemoresistance via upregulating ZNF185 and downregulating SMAD4 in pancreatic cancer. *J. Oncol.* 2020, 3181596. doi:10.1155/2020/3181596

Fang, S., Lu, J., Zhou, X., Wang, Y., Ross, M. I., Gershenwald, J. E., et al. (2020). Functional annotation of melanoma risk loci identifies novel susceptibility genes. *Carcinogenesis* 41, 452–457. doi:10.1093/carcin/bgz173

Gayral, M., Lulka, H., Hanoun, N., Biollay, C., Selves, J., Vignolle-Vidoni, A., et al. (2015). Targeted oncolytic herpes simplex virus type 1 eradicates experimental pancreatic tumors. *Hum. Gene Ther.* 26, 104–113. doi:10.1089/hum.2014.072

Grossberg, A. J., Chu, L. C., Deig, C. R., Fishman, E. K., Hwang, W. L., Maitra, A., et al. (2020). Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. *CA Cancer J. Clin.* 70, 375–403. doi:10.3322/caac.21626

Gupta, V. K., Sharma, N. S., Durden, B., Garrido, V. T., Kesh, K., Edwards, D., et al. (2021). Hypoxia-Driven oncometabolite L-2HG maintains stemness-differentiation balance and facilitates immune evasion in pancreatic cancer. *Cancer Res.* 81, 4001–4013. doi:10.1158/0008-5472.CAN-20-2562

Huffman, N., Palmieri, D., and Coppola, V. (2019). The CTLH complex in cancer cell plasticity. *J. Oncol.* 2019, 4216750. doi:10.1155/2019/4216750

Inman, K. S., Liu, Y., Scotti Buzhardt, M. L., Leitges, M., Krishna, M., Crawford, H. C., et al. (2022). Prkci regulates autophagy and pancreatic tumorigenesis in mice. *Cancers (Basel)* 14, 796. doi:10.3390/cancers14030796

Jiang, F. C., Li, G. S., Luo, J. Y., Huang, Z. G., Dang, Y. W., Chen, G., et al. (2022). Downregulation of zinc finger protein 71 expression in oral squamous cell carcinoma tissues and its underlying molecular mechanism. *Pathol. Res. Pract.* 238, 154109. doi:10.1016/j.prp.2022.154109

Kandimalla, R., Tomihara, H., Banwait, J. K., Yamamura, K., Singh, G., Baba, H., et al. (2020). A 15-gene immune, stromal, and proliferation gene signature that significantly associates with poor survival in patients with pancreatic ductal adenocarcinoma. *Clin. Cancer Res.* 26, 3641–3648. doi:10.1158/1078-0432.CCR-19-4044

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi:10.1093/nar/gkm882

Korakiti, A. M., Moutafi, M., Zografos, E., Dimopoulos, M. A., and Zagouri, F. (2020). The genomic profile of pregnancy-associated breast cancer: A systematic review. *Front. Oncol.* 10, 1773. doi:10.3389/fonc.2020.01773

Kuninty, P. R., Bansal, R., De Geus, S. W. L., Mardhian, D. F., Schnittert, J., van Baarlen, J., et al. (2019). ITGA5 inhibition in pancreatic stellate cells attenuates desmoplasia and potentiates efficacy of chemotherapy in pancreatic cancer. *Sci. Adv.* 5, eaax2770. doi:10.1126/sciadv.aax2770

Laurin, M., Huber, J., Pelletier, A., Houalla, T., Park, M., Fukui, Y., et al. (2013). Rac-specific guanine nucleotide exchange factor DOCK1 is a critical regulator of HER2-mediated breast cancer metastasis. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7434–7439. doi:10.1073/pnas.1213050110

Liu, B., Xing, X., Li, X., Guo, Q., Xu, T., and Xu, K. (2018). ZNF259 promotes breast cancer cells invasion and migration via ERK/GSK3β/snail signaling. *Cancer Manag. Res.* 10, 3159–3168. doi:10.2147/CMAR.S174745

Liu, S., Hausmann, S., Carlson, S. M., Fuentes, M. E., Francis, J. W., Pillai, R., et al. (2019). METTL13 methylation of eEF1A increases translational output to promote tumorigenesis. *Cell* 176, 491–504. doi:10.1016/j.cell.2018.11.038

Liu, X. Y., Guo, C. H., Xi, Z. Y., Xu, X. Q., Zhao, Q. Y., Li, L. S., et al. (2021). Histone methylation in pancreatic cancer and its clinical implications. *World J. Gastroenterol.* 27, 6004–6024. doi:10.3748/wjg.v27.i36.6004

Ma, Z., Li, Z., Wang, S., Zhou, Z., Liu, C., Zhuang, H., et al. (2022). ZMAT1 acts as a tumor suppressor in pancreatic ductal adenocarcinoma by inducing SIRT3/p53 signaling pathway. *J. Exp. Clin. Cancer Res.* 41, 130. doi:10.1186/s13046-022-02310-8

Mishra, N. K., and Guda, C. (2017). Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget* 8, 28990–29012. doi:10.18632/oncotarget.15993

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

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

Rooda, I., Hensen, K., Kaselt, B., Kasvandik, S., Pook, M., Kurg, A., et al. (2020). Target prediction and validation of microRNAs expressed from FSHR and aromatase genes in human ovarian granulosa cells. *Sci. Rep.* 10, 2300. doi:10.1038/s41598-020-59186-x

Sun, L., Lin, Y., Wang, G., Zhang, L., Hu, L., and Lu, Z. (2021). Correlation of zinc finger protein 2, a prognostic biomarker, with immune infiltrates in liver cancer. *Biosci. Rep.* 41. doi:10.1042/BSR20203115

The UniProt Consortium (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169. doi:10.1093/nar/gkw1099

Tjomsland, V., Spangeus, A., Sandstrom, P., Borch, K., Messmer, D., and Larsson, M. (2010). Semi mature blood dendritic cells exist in patients with ductal pancreatic adenocarcinoma owing to inflammatory factors released from the tumor. *PLoS One* 5, e13441. doi:10.1371/journal.pone.0013441

Wang, L., Zhou, W., Zhong, Y., Huo, Y., Fan, P., Zhan, S., et al. (2017). Overexpression of G protein-coupled receptor GPR87 promotes pancreatic cancer aggressiveness and activates NF-κB signaling pathway. *Mol. Cancer* 16, 61. doi:10.1186/s12943-017-0627-6

Wang, W., Jorgenson, E., Whittemore, A. S., and Asgari, M. M. (2018). Susceptibility loci-associated cutaneous squamous cell carcinoma invasiveness. *J. Invest. Dermatol* 138, 557–561. doi:10.1016/j.jid.2017.09.034

Wang, L., Jia, Z., Xie, D., Zhao, T., Tan, Z., Zhang, S., et al. (2020). Methylation of HSP70 orchestrates its binding to and stabilization of BCL2 mRNA and renders pancreatic cancer cells resistant to therapeutics. *Cancer Res.* 80, 4500–4513. doi:10.1158/0008-5472.CAN-19-1738

Wang, Y., Qin, C., Yang, G., Zhao, B., and Wang, W. (2021). The role of autophagy in pancreatic cancer progression. *Biochim. Biophys. Acta Rev. Cancer* 1876, 188592. doi:10.1016/j.bbcan.2021.188592

Wang, B., Zhao, H., Ni, S., and Ding, B. (2021). Integrated analysis of the roles of RNA binding proteins and their prognostic value in clear cell renal cell carcinoma. *J. Healthc. Eng.* 2021, 5568411. doi:10.1155/2021/5568411

Wang, S. S., Xu, J., Ji, K. Y., and Hwang, C. I. (2021). Epigenetic alterations in pancreatic cancer metastasis. *Biomolecules* 11, 1082. doi:10.3390/biom11081082

Wang, Y., Zhou, W., Chen, Y., He, D., Qin, Z., Wang, Z., et al. (2022). Identification of susceptibility modules and hub genes of osteoarthritis by WGCNA analysis. *Front. Genet.* 13, 1036156. doi:10.3389/fgene.2022.1036156

Wen, X., Shao, Z., Chen, S., Wang, W., Wang, Y., Jiang, J., et al. (2020). Construction of an RNA-binding protein-related prognostic model for pancreatic adenocarcinoma based on TCGA and GTEx databases. *Front. Genet.* 11, 610350. doi:10.3389/fgene.2020.610350

Xiao, Y., Xiang, T., Luo, X., Li, C., Li, Q., Peng, W., et al. (2014). Zinc-finger protein 545 inhibits cell proliferation as a tumor suppressor through inducing apoptosis and is disrupted by promoter methylation in breast cancer. *PLoS One* 9, e110990. doi:10.1371/journal.pone.0110990

Xu, N., Wu, Y. P., Ke, Z. B., Liang, Y. C., Cai, H., Su, W. T., et al. (2019). Identification of key DNA methylation-driven genes in prostate adenocarcinoma: An integrative analysis of TCGA methylation data. *J. Transl. Med.* 17, 311. doi:10.1186/s12967-019-2065-2

Yang, Y., Cao, L., Guo, Z., Gu, H., Zhang, K., and Qiu, Z. (2022). Deubiquitinase UCHL5 stabilizes ELK3 to potentiate cancer stemness and tumor progression in pancreatic adenocarcinoma (PAAD). *Exp. Cell Res.* 421, 113402. doi:10.1016/j.yexcr.2022.113402

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

Zhang, D., Cui, F., Peng, L., Wang, M., Yang, X., Xia, C., et al. (2022). Establishing and validating an ADCP-related prognostic signature in pancreatic ductal adenocarcinoma. *Aging (Albany NY)* 14, 6299–6315. doi:10.18632/aging.204221

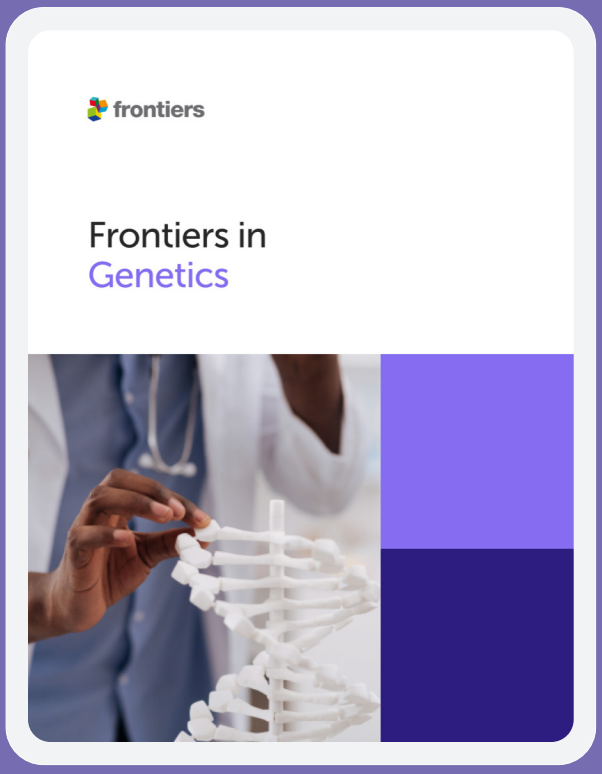
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact