frontiers | Research Topics

Epitranscriptomic RNA modification in non-apoptotic forms of regulated cell death

Edited by
Xing Niu, Xiang Xue and Min Sun

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-4205-7
DOI 10.3389/978-2-8325-4205-7

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

Epitranscriptomic RNA modification in non-apoptotic forms of regulated cell death

Topic editors
Xing Niu — China Medical University, China
Xiang Xue — University of New Mexico, United States
Min Sun — Hubei University of Medicine, China

Citation
Niu, X., Xue, X., Sun, M., eds. (2024). *Epitranscriptomic RNA modification in non-apoptotic forms of regulated cell death*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-4205-7

Table of contents

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fgene.2022.1000652

Risk model of hepatocellular carcinoma based on cuproptosis-related genes

Zhiqiang Liu[1†], Yong Qi[1†], Haibo Wang[1], Qikun Zhang[1], Zhengsheng Wu[2]* and Wenyong Wu[1]*

[1]Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China,
[2]Department of Pathology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Meirong Chi,
Fujian Agriculture and Forestry University, China
Sarun Juengpanich,
Sir Run Run Shaw Hospital, China
Wei Li,
Shenzhen Longhua District Central Hospital, China

*CORRESPONDENCE
Zhengsheng Wu,
wuzhengsheng@ahmu.edu.cn
Wenyong Wu,
m13805694400@163.com

[†]These authors have contributed equally to this work

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

RECEIVED 22 July 2022
ACCEPTED 31 August 2022
PUBLISHED 15 September 2022

CITATION
Liu Z, Qi Y, Wang H, Zhang Q, Wu Z and Wu W (2022), Risk model of hepatocellular carcinoma based on cuproptosis-related genes.
Front. Genet. 13:1000652.
doi: 10.3389/fgene.2022.1000652

Background: Owing to the heterogeneity displayed by hepatocellular carcinoma (HCC) and the complexity of tumor microenvironment (TME), it is noted that the long-term effectiveness of the cancer therapy poses a severe clinical challenge. Hence, it is essential to categorize and alter the treatment intervention decisions for these tumors.

Materials and methods: "ConsensusClusterPlus" tool was used for developing a secure molecular classification system that was based on the cuproptosis-linked gene expression. Furthermore, all clinical properties, pathway characteristics, genomic changes, and immune characteristics of different cell types involved in the immune pathways were also assessed. Univariate Cox regression and the least absolute shrinkage and selection operator (Lasso) analyses were used for designing the prognostic risk model associated with cuproptosis.

Results: Three cuproptosis-linked subtypes (clust1, clust2, and clust3) were detected. Out of these, Clust3 showed the worst prognosis, followed by clust2, while Clust1 showed the best prognosis. Three subtypes had significantly different enrichment in pathways related to Tricarboxylic Acid (TCA) cycle, cell cycle, and cell senescence ($p < 0.01$). The clust3 subtype with poor prognosis had a low "ImmuneScore" and low immune cell infiltration, and the three subtypes had significant differences in the antigen processing and presentation pathway of the macrophages. Clust1 had a low TIDE score and was sensitive to immunotherapy. Then, according to the prognosis-related genes of cuproptosis, a prognosis risk model related to cuproptosis was constructed, containing seven genes (KIF2C, PTTG1, CENPM, CDC20, CYP2C9, SFN, and CFHR3). "High" group had a higher TIDE score compared to the TIDE score value shown by the "Low" group, which benefited less from immunotherapy, whereas the "High" group patients were more sensitive to the conventional drugs. Finally, the prognosis risk model related to cuproptosis was combined with clinical pathological characteristics to further improve the prognostic model and survival prediction.

Conclusion: Three new molecular subgroups based on cuproptosis-linked genes were revealed, and a cuproptosis-related prognostic risk model comprising seven genes was established in this study, which could assist in predicting the prognosis and identifying the patients benefit from immunotherapy.

KEYWORDS

liver cancer, cuproptosis, molecular subtype, risk score, prognosis, tumor immunity

Introduction

Liver cancer includes primary liver cancer and secondary forms of liver cancer. Hepatocellular carcinoma (HCC) is a very prevalent type of primary liver cancer, followed by intrahepatic cholangiocarcinoma and other rare cancers (sarcoma, hemangioendothelioma, etc.) (Li et al., 2022). Liver cancer shows a poor prognosis. The frequency of liver cancer has significantly increased in the past few years, while its 5-year survival OS rate is <20% (Pham et al., 2022). Hepatitis B and C viruses, non-alcoholic fatty liver disease (NAFLD), alcohol consumption, and other factors (i.e., aflatoxin and microcystin) are among the primary causes of liver cancer. Out of these, HBV and HCV are seen to be the major risk factors for liver cancer (Lin et al., 2020). Although numerous high- or low-expression genes linked to the onset of liver cancer and carcinogenesis have been identified, the probable molecular mechanism of liver cancer is not entirely understood. Precision medicine can introduce a fresh perspective for individualized cancer diagnosis and focused therapy by considering the heterogeneity of every patient. Therefore, clinicians should propose more specific diagnosis and treatment methods for the subtype of the disease for optimizing the efficacy of treatment, thereby decreasing the resulting side effects (Liang et al., 2018).

From bacteria and fungi to plants and animals, copper is the basic element of life. In the human body, it combines with enzymes to help blood clots, hormone maturation, and cell energy processing and is also involved in many biological behaviors. However, too much copper will kill cells and cause pathological damage to multiple organs. Studies have shown that copper ion is both a key cofactor of many enzymes, and excessive copper ion will lead to cell death. The exact mechanism of cuproptosis involves the induction of cell death after combining the tricarboxylic acid cycle (TCA)-linked enzymes, leading to a protein toxic stress response, which differs from the cell death mechanisms discussed in the past (Tsvetkov et al., 2022). Many studies have shown that copper metabolism is involved in many pathophysiologies of chronic hepatitis. A long-term exposure to a higher concentration of copper ions or the long-term usage of unqualified copper water pipes and tableware could lead to chronic copper poisoning, thereby causing chronic liver disease (Guo et al., 2021; Nakaichi et al., 2021). In addition, cuproptosis has aroused widespread concern in a variety of liver diseases. Excessive copper exposure can lead to oxidative stress, due to excessive reactive oxygen species (ROS) production and reduced antioxidant function, and then promote hepatocyte apoptosis through mitochondrial apoptosis. Earlier reports also stated that the TNF-R1 signaling pathway played a vital role in the Cu-induced apoptosis pathway (Liu et al., 2020). Copper metabolism is closely related to human-related genetic disease hepatolenticular degeneration (Xu et al., 2021), and such patients have been associated with copper storage disorders for a long time. Hence, it becomes important to determine novel molecular markers and identify the cuproptosis-linked downstream signaling pathways, for understanding the regulatory role played by cuproptosis in the pathophysiology of liver cancer.

In this report, the cuproptosis-linked genes were used to identify stable molecular subtypes through consistent clustering. Thereafter, the clinical characteristics, pathway characteristics, and immune characteristics were compared between the different subtypes. Finally, genes related to the cuproptosis phenotype were detected using the expression difference analysis and least absolute shrinkage and selection operator (Lasso). Furthermore, the risk model and clinical prognostic model were constructed, which can assist in the personalized treatment of liver cancer patients.

Materials and methods

Data collection and processing

The Cancer Genome Atlas (TCGA) GDC API was used for downloading TCGA-LIHC dataset containing RNA-seq data, copy number variation (CNV) and mutation data used in this study. Primary tumor samples were remained. Samples with no survival information were removed. After the screening, 50 normal and 360 primary tumor samples were included in this study. The Gene Expression Omnibus (GEO) database provided the gene expression data for the GSE14520 dataset. Following identification, 242 liver carcinoma samples were used in the study. Here, the TCGA-LIHC was used as a training set, while the GSE14520 dataset was used as an independent verification set. The cuproptosis-linked genes in this study came from the study of Tsvetkov et al. (2022), a total of 13 cuproptosis-related genes, i.e. ATP7A, LIAS, LIPT1, DLD, DBT, DLST, FDX1, PDHA1, DLAT, GCSH, PDHB, SLC31A1, and ATP7B. The bioinformatics analysis of this study was supported by the Sangerbox tool (http://vip.sangerbox.com/) (Shen et al., 2022). The work flow of this study was shown in Supplementary Figure S1.

Data preprocessing

The RNA-seq data downloaded from the TCGA database were preprocessed as mentioned below: 1) All samples without any clinical follow-up data were discarded; 2) All samples without information regarding their survival duration were eliminated; 3) All samples without their OS were eliminated; 4) Ensembl was converted to the Gene symbol; and 5) Median values of the expressions with multiple gene symbols were considered. On the other hand, the GEO data were pre-processed as follows: For the GEO data set, the annotation information of the corresponding chip platform was downloaded. According to the annotation information, the probe was mapped to the gene, and the probe that matched multiple genes was discarded. If a gene matched multiple probes, the median value was regarded as its gene expression value.

Molecular subtypes of the cuproptosis-linked genes

ConsensusClusterPlus was used for consistent clustering to build a consistency matrix, and the samples were clustered and typed (Wilkerson and Hayes, 2010). The molecular subtypes of all the samples were derived using the expression data of the cuproptosis-linked genes. “Pam” algorithm and “Euclidean” were used as the distance measurement, and 500 bootstraps were conducted, wherein every bootstrap process included 80% of all patients in a training set. The cluster number was defined as between 2 and 10, and the best classification was selected by determining the consistency matrix and consistency cumulative distribution function for determining the molecular subtype of the sample.

Constructing a risk model

1) Through the molecular subtypes identified previously, the cuproptosis-linked genes with differences between the subtypes were identified. Here, the differences between the clust1 vs. non-clust1 subtypes, clust2 vs. non-clust2, and clust3 vs. non-clust3 subtypes, were identified through the Limma package (Ritchie et al., 2015). The differentially expressed genes (DEGs) were also identified based on their FDR<0.05 and | log2FC |>1 values.
2) Univariate Cox analysis was conducted through the Cox function in the survival package, and DEGs with significant prognosis (| logFC |>1 & FDR<0.05) were selected.
3) Lasso regression (Tibshirani, 1997) was used to decrease the number of genes. Stepwise regression was then utilized, using the Akaike Information Criterion (AIC), which considered the model’s statistical fit and the no. of parameters that could be used for fitting. The most complex model was used to start the stepAIC technique in the MASS package (Zhang, 2016), and one variable was eliminated at a time to lower AIC. The model performed better with a smaller value, indicating that it had achieved an acceptable degree of fit with fewer parameters.

The RS of each patient was estimated using the formula as follows: RiskScore (RS) = $\Sigma\beta_i \times EXP_i$. EXP_i refers to the gene expression level of gene characteristics related to the prognosis of cuproptosis-related phenotypes, while β_i refers to a Cox regression coefficient for the respective gene. To categorize patients into high-risk and low-risk RS groups, survminer R package (http://www.sthda.com/english/rpkgs/survminer/) was used to calculate the optimal cut-off. KM curve was used for drawing the survival curve for prognostic analysis, while the log-rank test was employed for determining the significant difference between the groups.

Gene set enrichment analysis technique

For investigating the pathways associated with various biological processes in numerous molecular subtypes, the “GSEA” technique was utilized for pathway analysis (Subramanian et al., 2005). Here, GSEA was analyzed using the c2. cp.kegg.v7.0. symbols.gmt as a background set through GSEA software, and identified with NP < 0.05. In addition, the TCA cycle-associated genes and pathways were downloaded from the MSigDB database in GSEA (http://www.GSEa-msigdb.org/GSEa/msigdb/search.jsp), and the ssGSEA was used for calculating the score of the TCA related pathways. Then, the pathways and genes related to cell growth and death were downloaded from KEGG’s official website (https://www.kegg.jp/kegg/pathway.html), and the score of the cell growth and death-related pathways was calculated by ssGSEA. In addition, the NK Cytotoxicity Score, Toll-Like Receptor Score, and the Antigen Processing and Presentation Score for every sample were determined using the ssGSEA process, with the help of the relevant genes involved in these pathways, derived from the GSEA-based MSigDB database.

Calculation of invasion abundance of tumor microenvironment cells

The relative abundance of the 22 immune cells involved in lung cancer was measured using the CIBERSORT algorithm (https://cibersort.stanford.edu/) (Newman et al., 2015). Simultaneously, the percentage of immune cells was determined using the Estimation of Stromal and Immune Cells in Malignant Tumors Using Expression Data (ESTIMATE) software, and the Wilcoxon rank sum test was employed for comparing the degree of immune cell

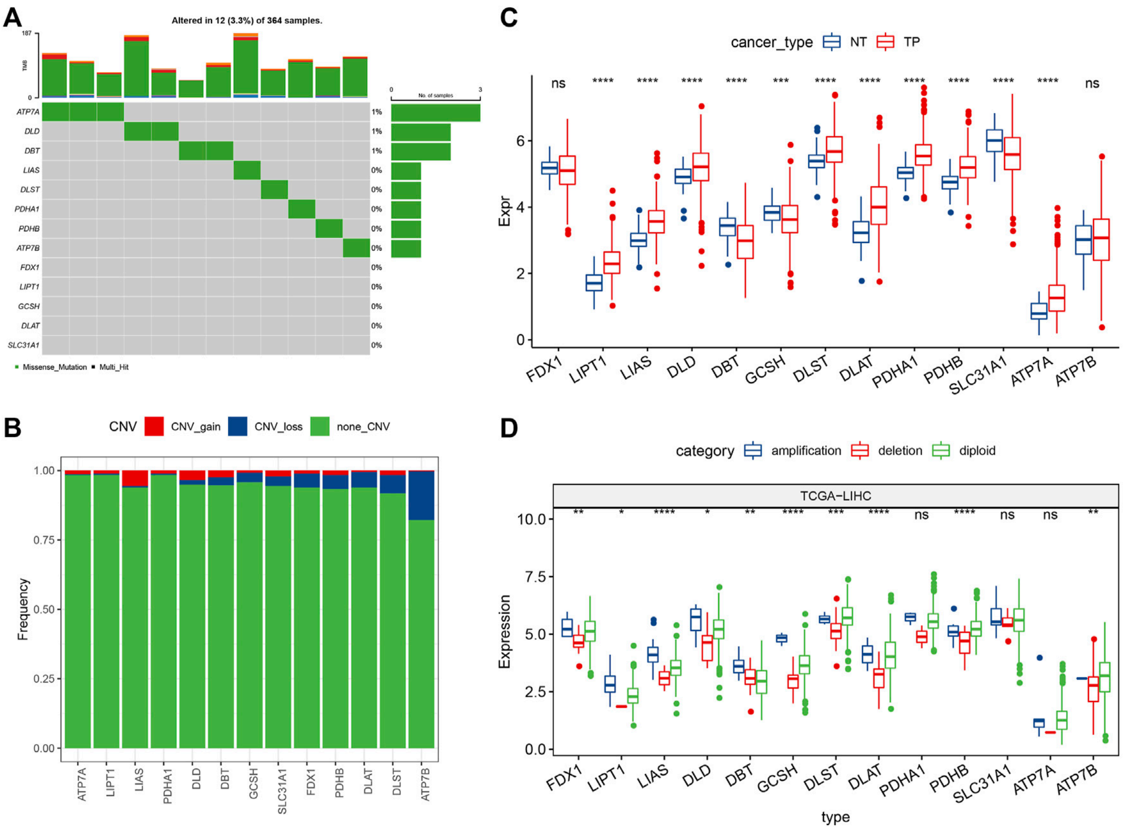


FIGURE 1
Mutation map and expression characteristics of cuproptosis-related genes in liver cancer. **(A)** Mutation map of cuproptosis-linked genes in the primary tumor samples; **(B)** CNVs of cuproptosis-linked genes in the primary tumor samples. Vertical axis indicates the percentage of CNV types of cuproptosis-related genes; **(C)** The differences of gene expression levels between different CNV types in primary tumor samples; **(D)** Differential analysis of transcriptional expression levels of cuproptosis-linked genes in primary tumor and adjacent normal tissue samples. Log2 (expression) was selected in C and D. ns, not significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

infiltration between the high-risk and the low-risk groups (Runa et al., 2017).

Prediction of responsiveness to immunotherapy

The effect of the Immune Checkpoint Inhibitor Score (IMS) on predicting the Immune Checkpoint inhibitors' (ICI) clinical reactivity was confirmed using the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm. Immune Checkpoint Blockade (ICB) reactivity is predicted using the gene expression profile by the TIDE algorithm (Jiang et al., 2018). The TIDE algorithm assessed two distinct mechanisms of the tumor immune escape scores, such as tumor-infiltrating Cytotoxic T Lymphocytes (CTLs) dysfunction score (dysfunction) and the immunosuppressive factor rejection score (exclusion), as well as 3 cell types that restricted T cell infiltration into the tumors, such as M2 subtype of the cancer-associated fibroblasts (CAF), myeloid-derived suppressor cells (MDSCs), and the tumor-associated macrophages (TAMs). The potential clinical consequences of immunotherapy in the new molecular subgroups were assessed in this study using the TIDE software (http://tide.dfci.harvard.edu). The likelihood of immunological escape increases with increasing TIDE prediction score, indicating that patients are less likely to benefit from immunotherapy.

Results

Gene mutations and transcriptional changes of cuproptosis-related genes

In this study, 13 cuproptosis-related genes were obtained. For determining the genetic changes caused by "cuproptosis" in liver cancer, the gene mutation rate of the somatic mutations in 13 cuproptosis genes was evaluated. Among 364 TCGA-LIHC primary tumor samples, 12 (3.3%) had mutations in cuproptosis-

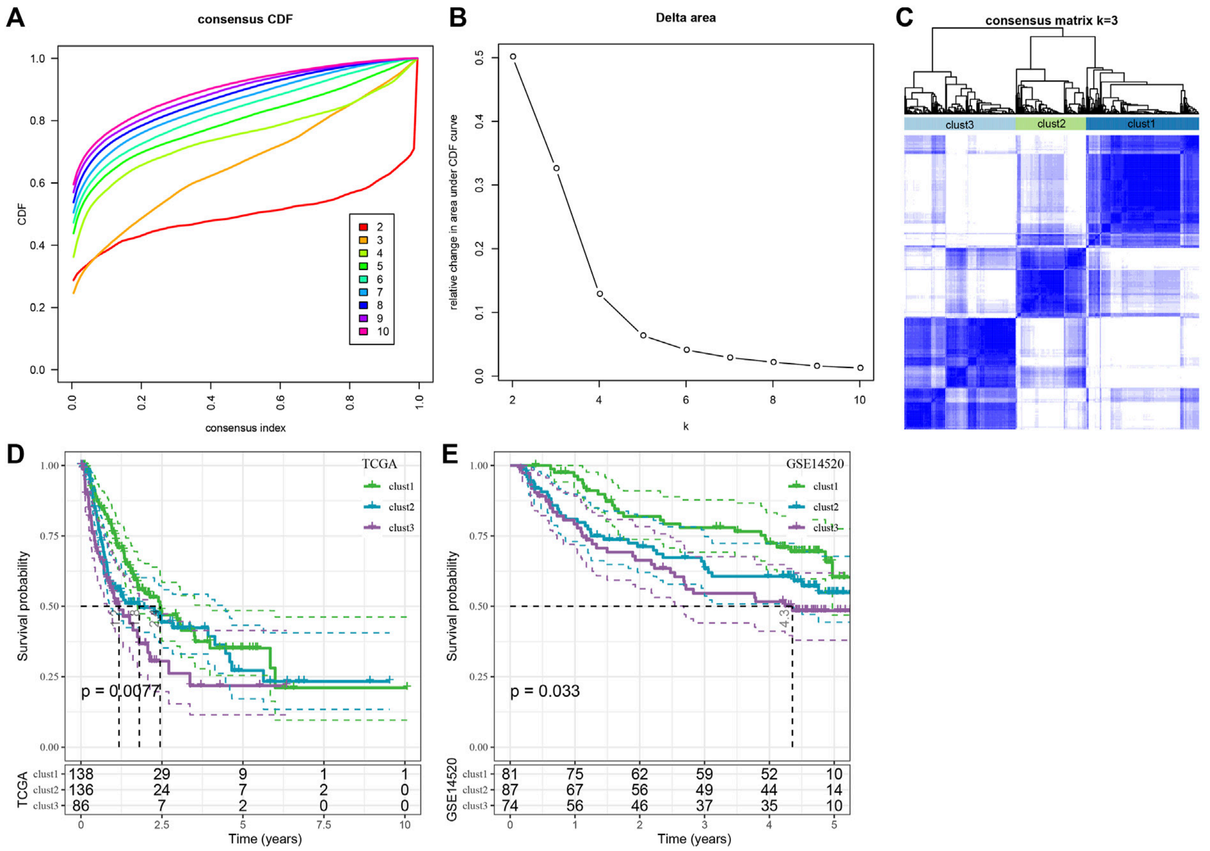


FIGURE 2
Consensus clustering analysis based on the prognosis of cuproptosis-linked genes in liver cancer. **(A)** CDF curve of TCGA-LIHC dataset samples; **(B)** CDF-delta area curve for the TCGA-LIHC dataset, Delta area curve for consensus clustering, which indicates the relative difference in the area under the CDF curve for every category number, k, in comparison to the k—1. The *X*-axis axis denotes the category number, k, whereas the *Y*-axis indicates the relative change in the area under the CDF curve; **(C)** Sample clustering-related heat map when the consumption $k = 3$; **(D)** KM curves denoting the correlation between the prognosis of 3 subtypes, identified using the TCGA-LIHC dataset; **(E)** KM curve of prognosis of three subtypes in GSE14520 cohort.

linked genes (Figure 1A). Among them, only ATP7A, DLD, and DBT had gene mutations. Then, we analyzed the somatic copy number changes of these cuproptosis-related genes in primary liver cancer and found that cuproptosis-related genes had a low CNV amplification/deletion frequency (Figure 1B). To determine if the genes related to cuproptosis are differentially expressed in primary tumors and normal tissues, the mRNA changes of cuproptosis-linked genes between the primary tumor samples and the adjoining normal tissue samples were compared, showing that a majority of the cuproptosis-linked genes were differentially expressed (Figure 1C). Further, to explore the difference in CNV value in mRNA expression in primary tumor tissues, patients with primary liver cancer were categorized into 3 groups according to CNV value, including increased CNV, CNV loss, and no significant change in CNV. Then, the mRNA expressions of the cuproptosis-linked genes between all groups were compared (Figure 1D). The results indicated that most of these cuproptosis-linked genes showed higher expression in patients with increased CNV and patients with lost CNV, and there was no significant change compared with CNV.

Molecular typing depending on the cuproptosis-linked genes

For understanding the expression pattern of the cuproptosis-linked genes, the liver cancer samples in the TCGA-LIHC dataset containing clinical information were used to classify patients through the consistent clustering of the expression profiles of these 13 cuproptosis-related genes. Then, an optimal no. of clusters was determined based on the cumulative distribution function (CDF), and the CDF Delta area curve showed that if the selected cluster was 3, it showed a very stable clustering outcome (Figures 2A,B), Finally, the k-value of 3 was selected to determining 3 molecular subtypes (Figure 2C). Analysis of the prognostic features of the 3 molecular subtypes showed that they displayed significant

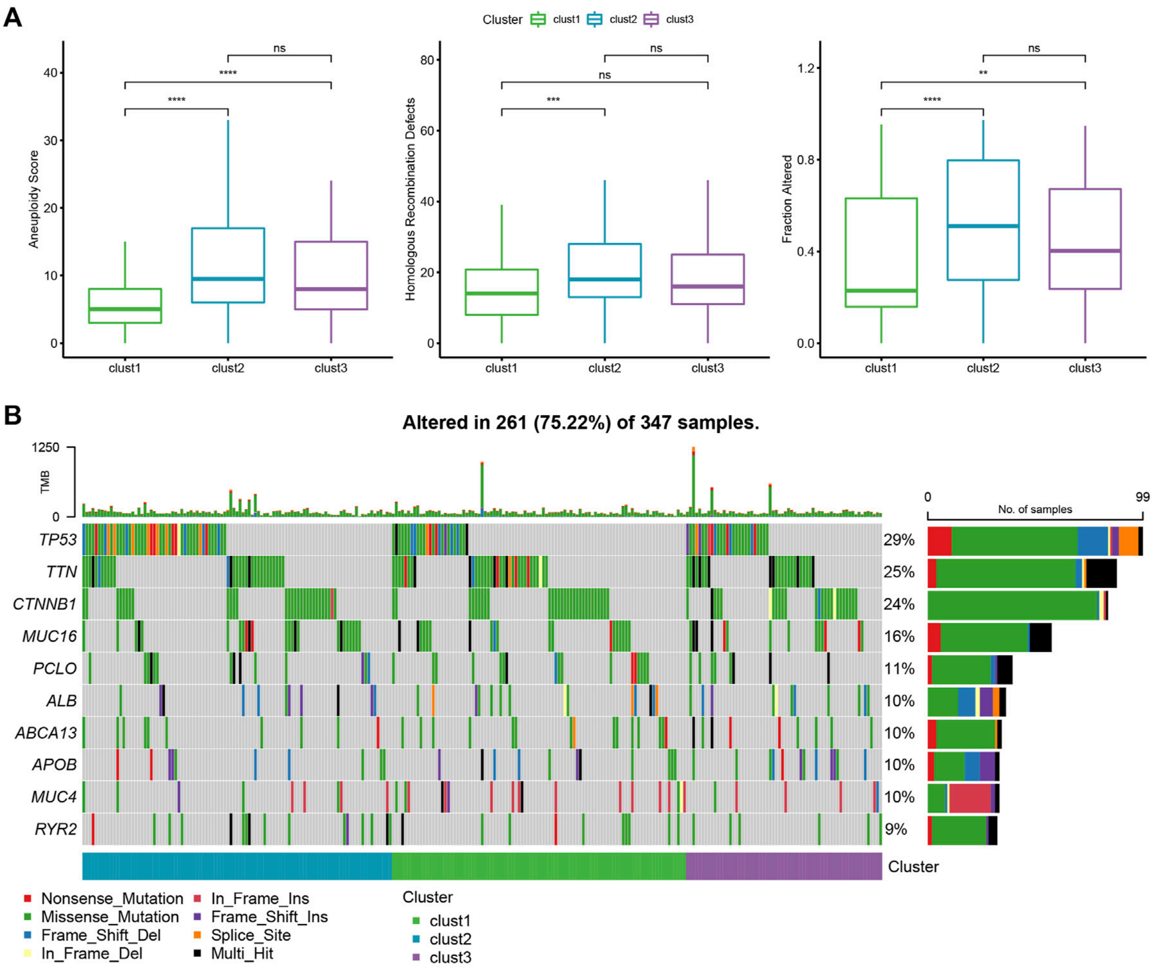


FIGURE 3
Genome changes of molecular subtypes in the TCGA-LIHC dataset. **(A)** The differences among the molecular subtypes of the TCGA-LIHC dataset were compared in terms of Aneuploidy Score, Homologous Recombination Defects, Fraction Altered. **(B)** The top 10 mutated genes in three subtypes. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

prognostic differences (Figure 2D). It was noted that clust3 showed the worst prognosis, followed by clust2, and clust1 had the best prognosis. This same technique was used for verifying the GSE14520 dataset and the results showed that significant differences existed in the prognosis of the 3 molecular subtypes (Figure 2E), which was in agreement with the TCGA-LIHC dataset.

Clinical characteristics and mutation characteristics between molecular subtypes

The clinicopathological characteristics of the numerous molecular subtypes in the TCGA-LIHC dataset were assessed. Then, the distribution of various clinical characteristics in 3 molecular subtypes was compared, and all distribution differences in the clinical characteristics of different subtypes were determined. The results revealed significant differences between the clust1 and clust2/clust3 in T-stage, Stage, and Grade (Supplementary Figure S2A). T1/T2 accounted for a relatively high proportion in clust1/clust2, and T3/T4 accounted for an increase in clust3. In the stage distribution, T1/T2 accounted for a relatively high proportion in clust1/clust2, and T3/T4 accounted for an increase in clust3. In the grade distribution, the proportion of clust1 in G1/G2 was relatively high, and the proportion of clust2/clust3 in G3/G4 was increased.

Mutation characteristics between the molecular subtypes

The different genomic changes noted in the 3 molecular subtypes existing in the TCGA-LIHC dataset were analyzed. Here, the molecular characteristics of TCGA-LIHC were

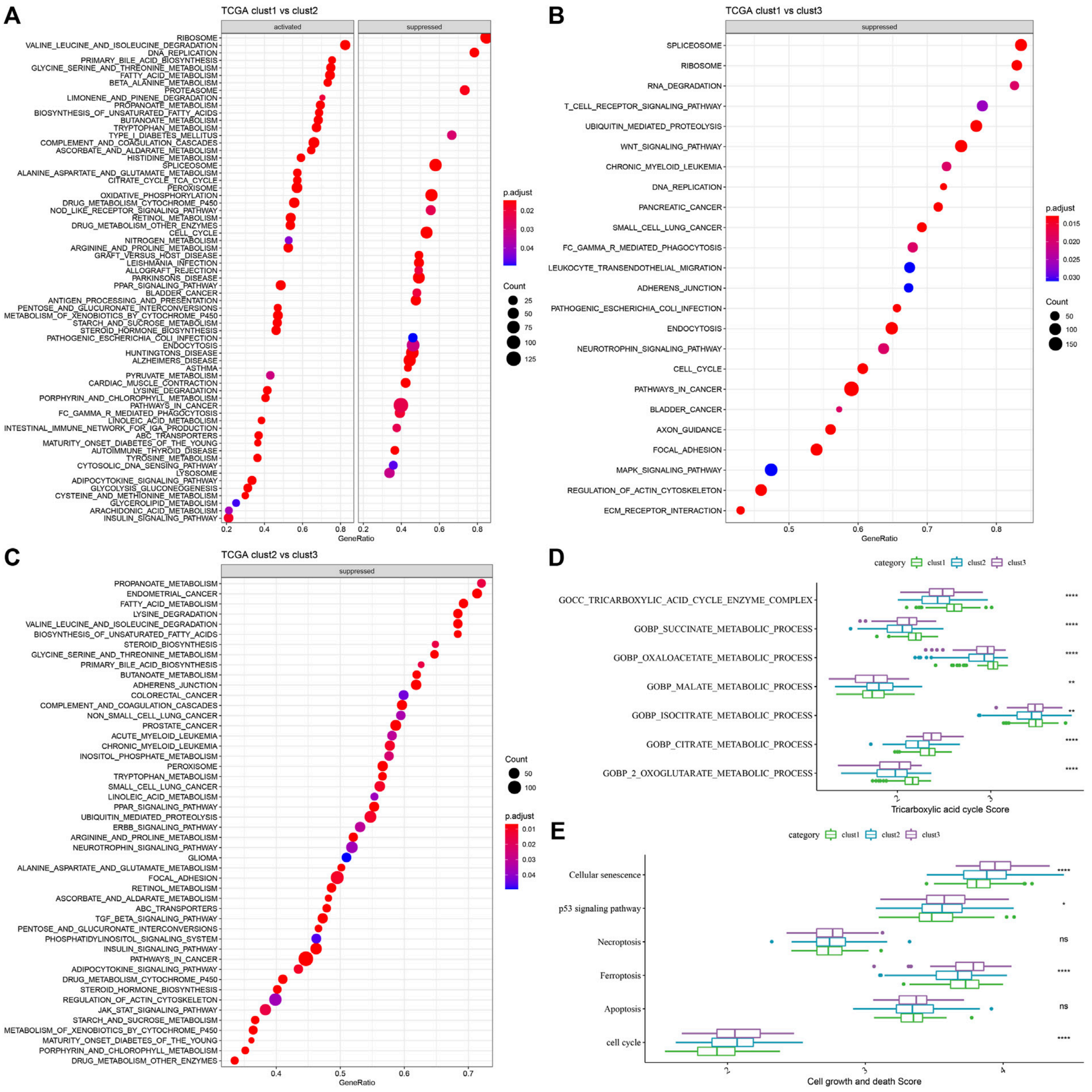


FIGURE 4
GSEA analysis of the 3 molecular subtypes. **(A)** Clust1 vs. clust2 GSEA analysis results in TCGA-LIHC dataset; **(B)** Clust1 vs. clust3 GSEA analysis results in TCGA-LIHC dataset; **(C)** Clust2 vs. clust3 GSEA analysis results in TCGA-LIHC dataset; **(D)** Comparison of TCA related pathway scores amongst the 3 molecular subtypes identified using the TCGA-LIHC dataset; **(E)** A comparison of cell growth- and cell death-related pathway scores amongst the 3 subtypes in the TCGA-LIHC dataset. (ANOVA, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; and $****p < 0.0001$).

downloaded from an earlier Pan-cancer Study (Thorsson et al., 2018). It can be seen that the clust1 subtype showed a low Aneuploidy Score and Homologous Recombination Defects (Figure 3A). In addition, a previous study divided HCC into 5 molecular subtypes according to 160 immune signatures, of which the immune molecular subtypes C1, C2, and C4 had the worst prognosis and C3 showed the best prognosis. In a comparison of the relationship between the 5 immune molecular subtypes and the proposed 3 molecular subtypes, it was noted that the C1/C2/C4 subtype of immune molecular subtype occupied more in clust3 and clust2 subtypes with poor prognosis, while the C3 subtype of immune molecular subtype occupied more in clust1 subtype with good prognosis (Supplementary Figures S2B,C). In addition, the differences in the gene mutations in various molecular subtypes were compared. TP53 was the mostly mutated gene in all three subtypes, with a total mutation rate of 29% (Figure 3B).

Pathway analysis of molecular subtypes

A comparative analysis of the pathways related to various biological processes in differing molecular subtypes was carried out. The results showed that when clust1 was compared to clust2 of the TCGA-LIHC dataset, the metabolic pathways like the ASCORBATE_AND_ALDARATE_METABOLISM were activated in clust1, while pathways like the SPLICEOSOME were activated in clust2. In clust1 vs. clust3, pathways such as SPLICEOSOME were activated in clust3, while in clust2 vs. clust3, pathways such as FATTY_ACID_METABOLISM were activated in clust3 (Figures 4A–C). Through previous studies, it was found that cuproptosis was related to TCA (Tsvetkov et al., 2022). ssGSEA analysis of the TCA score showed that the three subtypes had significant differences in TCA-linked pathways (Figure 4D). Then, the score of cell growth and death-associated pathways was calculated by the ssGSEA technique. The results showed no significant differences between the 3 subtypes except necroptosis and apoptosis, and there were significant differences among the other four pathways related to cell growth and death (Figure 4E).

Immune properties displayed by the various molecular subtypes

For determining the differences present in the immune microenvironment of the patients belonging to various molecular subtypes, the expression level of genes in the immune cells was utilized for assessing the level of infiltration of the immune cells in the TCGA-LIHC dataset. First, the relative quantity of 22 immune cells was determined using the CIBERSORT algorithm. Figure 5A revealed substantial disparities across the different subtypes, associated with 8 immune cell types, including the memory B cells, naive B cells, regulatory T cells, and macrophages (M0, M1, and M2), etc. The immune cell infiltration was also assessed simultaneously using ESTIMATE. The outcomes demonstrated that the three "ImmuneScore" subtypes differed significantly from one another. The "ImmuneScore" of the clust3 subtype having a poor prognosis was lower than that of other subtypes, with low immune cell infiltration (Figure 5B). Further, the sensitivity differences of different molecular subtypes in the TCGA-LIHC dataset to immunotherapy were analyzed. Firstly, the variation in the expression of various immune checkpoints in the different subtypes was compared. The findings showed that these molecular subtypes differentially expressed 34 immune checkpoint genes (Figure 5C). Figure 5A showed that the molecular subtypes particularly showed a differential expression of the macrophages, which are cells that play a vital role in immune regulation, such as Toll-like receptor signaling pathway, and macrophage antigen processing and presentation. And there are FC receptors on the surface of macrophages, which can kill tumor cells through specific antibodies, like the Antibody-Dependent Cell-mediated Cytotoxicity (ADCC) effect (NK-cell mediated cytotoxicity). Therefore, the ssGSEA was used to calculate the immune scores like NK Cytotoxicity Scores, Toll-like Receptor Score, and Antigen Processing and Presentation Score, for every sample. Simultaneously, the ANOVA test found that there were significant differences in macrophages in antigen processing and presentation (Figure 5D). Finally, the TIDE software was employed for analyzing the differences between the different subtypes with regard to immunotherapy. Figure 5E showed that the TIDE score of clust2 and clust3 subtypes in the TCGA queue was higher than the clust1, suggesting that the clust1 subtype had a lower probability of immune escape and showed a higher probability of benefitting from immunotherapy.

Establishment and validation of clinical prognosis model

Then, the cuproptosis-linked genes, which differed between various subtypes, were identified. Finally, a total of 499 differential genes were chosen for additional analysis, and the results of the differential analysis were shown in the volcanic map (Supplementary Figures S3A–C). Univariate Cox analysis was carried out on 499 differential genes, and a total of 15 genes showing a significant impact on prognosis were identified ($p < 0.001$), including 11 "Risk" and 4 "Protective" genes (Supplementary Figure S3D). Supplementary Figure S3E shows the forest map of univariate Cox analysis of 15 prognosis-related genes.

Then, 15 genes were further compressed using the stepwise regression technique, and 7 genes were derived, i.e., KIF2C, PTTG1, CENPM, CDC20, CYP2C9, SFN, CFHR3. The RSs of each sample were calculated through 7 gene expression levels with TCGA-LIHC data as the training data set. Then receiver operating characteristic (ROC) analysis was used for determining the classification efficiency of the prognosis prediction for 1–5 years. Area under the curve (AUC) for 1-, 2-, 3-, 4- and 5-year OS were seen to be 0.72, 0.66, 0.65, 0.67, and 0.75, respectively, wherein the AUC values for 1- and 5-years were >0.7 (Figure 6A). Simultaneously, Z-score conversion was performed on RS. Samples with RS > 0 were classified into the "high-risk" group, while samples with RS < 0 were categorized into the "Low-risk" group, and KM curves were drawn. Results revealed significant differences between both the groups ($p < 0.0001$), and "high" group showed a worse prognosis compared to the "low" group (Figure 6B).

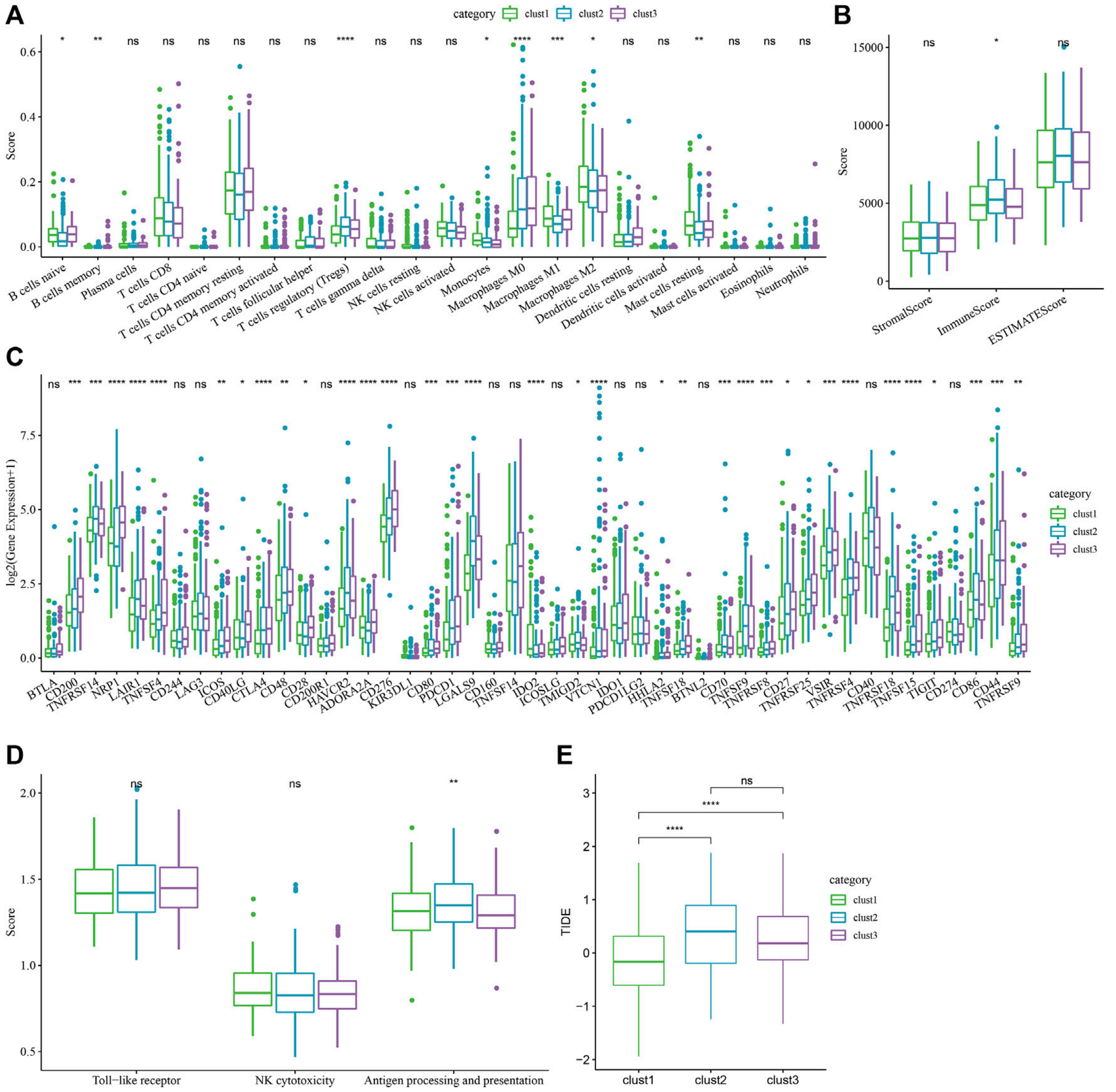


FIGURE 5

Immune-related characteristics of each cuproptosis subtype. **(A)** The variations in the 22 immune cell scores displayed by the 3 molecular subtypes identified using the TCGA-LIHC dataset; **(B)** The difference of ESTIMATE immune infiltration amongst the 3 molecular subtypes identified using the TCGA-LIHC dataset; **(C)** Immune checkpoints that were expressed differentially by the different groups in the TCGA-LIHC dataset; **(D)** Difference analysis of macrophage participation in related pathways between different groups in the TCGA-LIHC dataset; (ANOVA, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; and $****p < 0.0001$) **(E)** The difference of TIDE analysis results between different groups in the TCGA-LIHC dataset (Wilcox Test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; and $****p < 0.0001$).

To verify the robustness of the model, the GSE14520 dataset was used to verify by the same method, and ROC curves were used for analyzing the classification efficiency of predicting the prognosis of 1–5 years of OS. The results revealed that the risk model could be effectively developed using the 7 genes. The AUC values for 1–5-year OS were seen to be 0.68, 0.7, 0.69, 0.67, and 0.62, respectively, wherein the AUC values for 2-years were >0.7 (Figure 6C). The same method was used to draw the KM curve, and both the groups showed significant differences ($p < 0.05$). Furthermore, the prognosis of the "low" group was significantly better compared to the "high" group (Figure 6D).

Performance of the RiskScore in different clinicopathological features and different molecular subtypes

For testing the correlation between the RS scores and the clinical characteristics of liver cancer, the difference in RS scoring

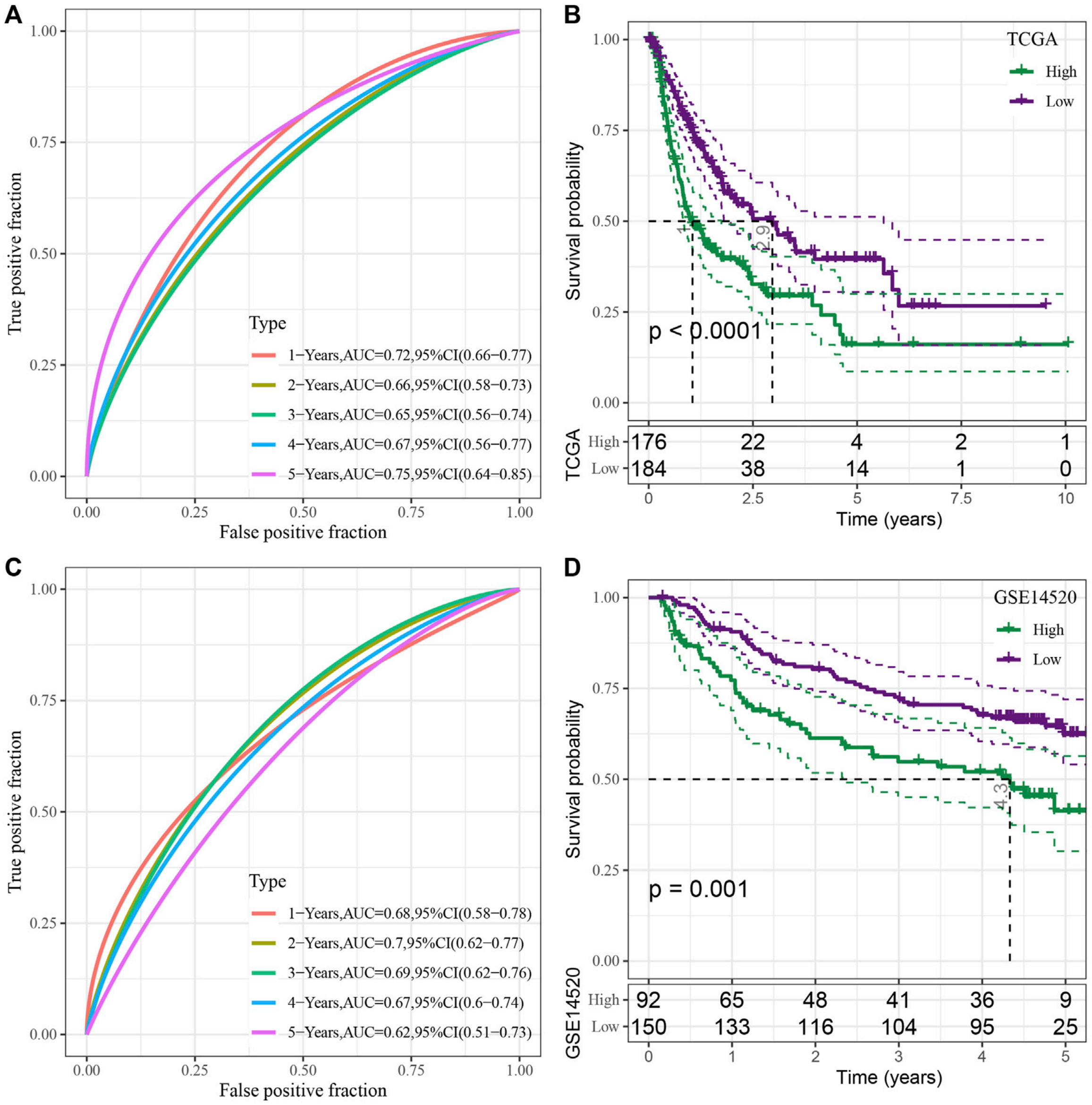


FIGURE 6
Construction and assessment of the RS model based on the seven cuproptosis-linked genes. **(A)** ROC curve of risk model constructed by seven genes in the TCGA-LIHC dataset; **(B)** KM curve of risk model constructed by seven genes in the TCGA-LIHC dataset; **(C)** ROC curve of risk model constructed by seven genes in the GSE14520 dataset; **(D)** KM curve of the risk model constructed by seven genes in the GSE14520 dataset.

between different TNM grades and Stage clinical grades in the TCGA-LIHC dataset was evaluated. The results implied that samples with higher clinical grades had higher RS (Figure 7A). Simultaneously, the clinicopathological differences between the RS groups in the TCGA-LIHC dataset were compared and significant differences were noted in the distribution of T-stage, Stage, Grade, Age, and Status between both the groups. “High” showed a higher clinical grade, and a greater number of patients died in the “high” group, which was in agreement with poor prognosis (Figure 7B).

Differences in immune characteristics and immunotherapy among RiskScore groups

The changes in the relative abundance of 22 different immune cell types in the high-RS and low-RS groups were examined in order to better understand the differences in the immunological microenvironment of patients in the RS group. The results showed that there were notable differences between the RS-high and -low groups in 12 different immune cell types, including plasma cells, macrophages (M0, M1, M2),

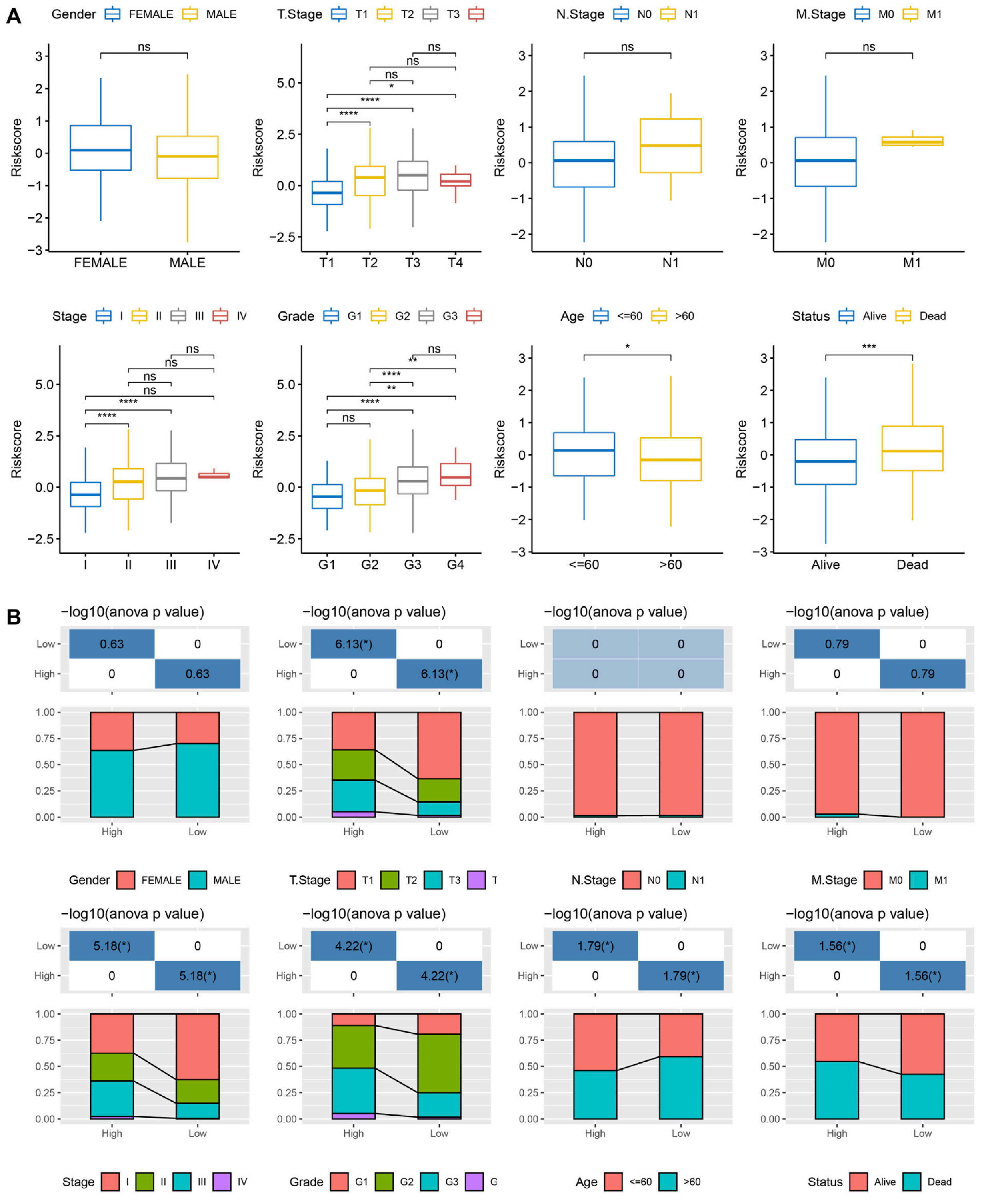


FIGURE 7
Correlation between RS and clinicopathological features. **(A)** Differences in RSs between different clinicopathological groups in the TCGA-LIHC dataset (Wilcox Test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; and ****$p < 0.0001$); **(B)** Clinicopathological features between RS groups in the TCGA-LIHC dataset.

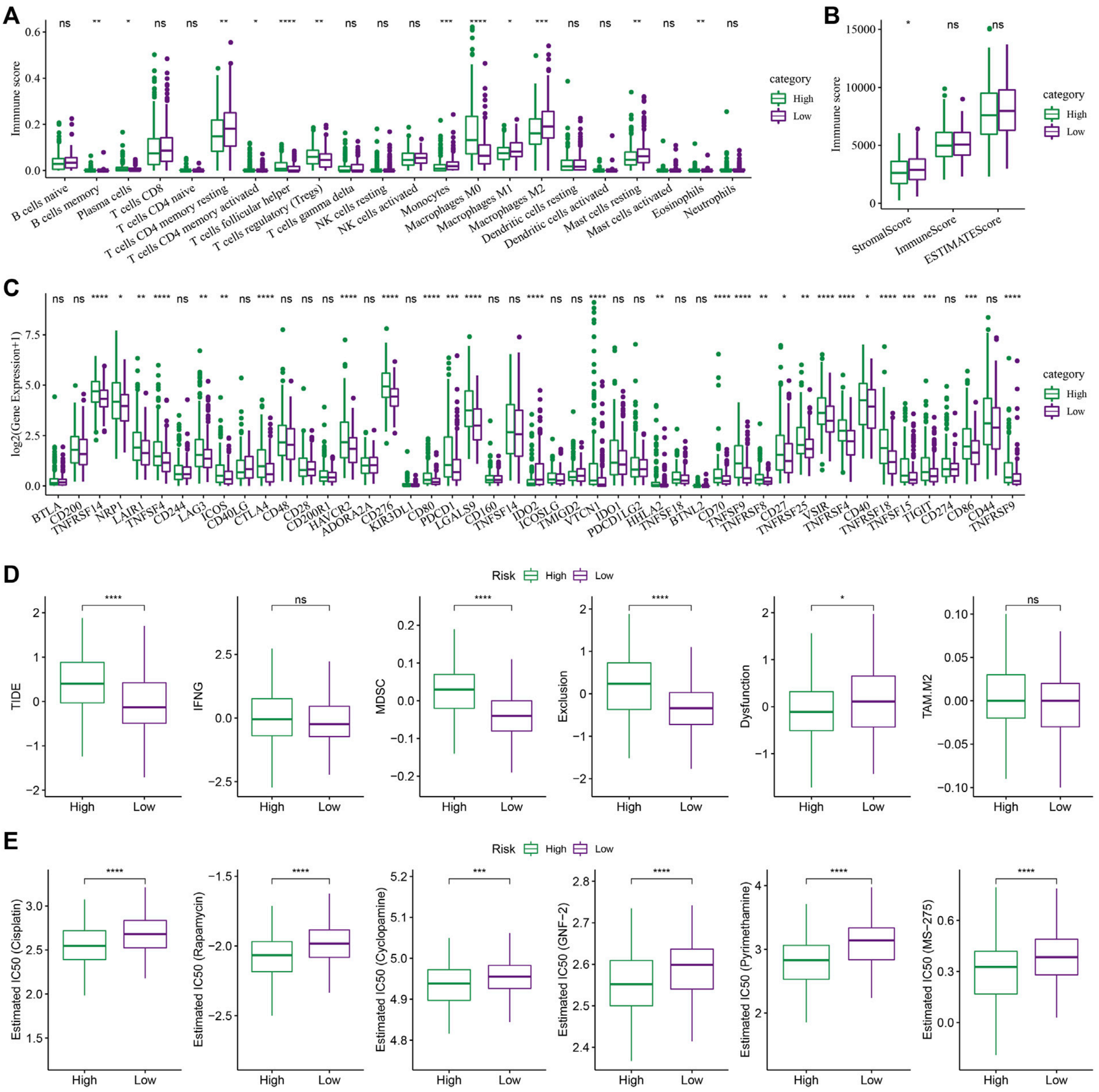


FIGURE 8
The role of the RS model in predicting the benefits of immunization/chemotherapy. **(A)** Variations in the 22 immune cell scores amongst the different risk groups identified using the TCGA-LIHC dataset; **(B)** Differences in immune and matrix scores amongst the different risk groups in the TCGA-LIHC dataset; **(C)** Immune checkpoints differentially expressed between various groups in the TCGA-LIHC dataset; **(D)** The results of TIDE analysis among different groups in TCGA-LIHC dataset were different; **(E)** Box plots of the calculated IC50 for the drug in TCGA-LIHC dataset (Wilcox test, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; and ****$p < 0.0001$).

memory B cells, etc. (Figure 8A, Wilcox.test). Additionally, the immune cell invasion was also evaluated using ESTIMATE (Wilcox.test). The findings demonstrated a statistically significant difference in the "ImmuneScore" values between the 2 groups. With a higher immune cell infiltration, the "ImmuneScore" in the "low" group was seen to be higher compared to that in the "high" group (Figure 8B).

Then the sensitivity difference of immunotherapy between the low and high-risk groups in the TCGA-LIHC dataset was analyzed. Firstly, the differences in the expression of the immune checkpoints between various checkpoints were compared. The results indicated that 28 immune checkpoint genes were expressed differentially between both the groups (Figure 8C, Wilcox.test).

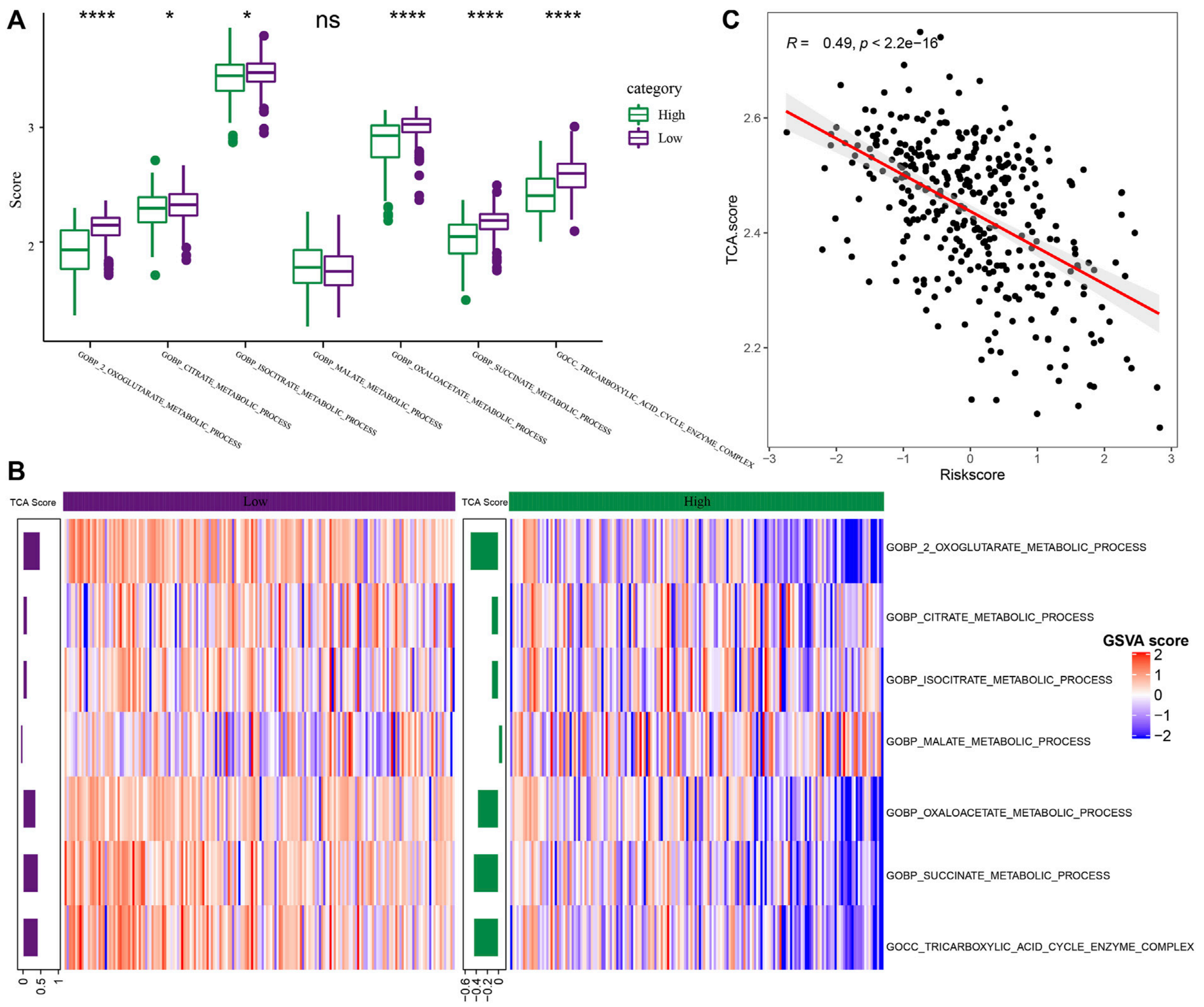


FIGURE 9
Differences in pathway characteristics among different RS groups. **(A)** The box plot of TCA-related pathway scores in high-risk and low-risk groups in the TCGA dataset (Wilcox.test, $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; and $****p < 0.0001$); **(B)** The heat map of TCA related pathway scores in the high-risk and low-risk groups in the TCGA-LIHC dataset; **(C)** The scatter diagram of correlation analysis between RS and TCA scores in TCGA-LIHC dataset.

Furthermore, the potential clinical effects of immunotherapy in the high- and low-RS groups in the TCGA-LIHC dataset, were analyzed, using the TIDE software. The analysis revealed significant differences in the MDSC, TIDE, Exclusion, and Dysfunction scores. It was concluded that the High-RS group showed higher scores than those shown by the Low-RS group (Figure 8, Wilcox.Test). In addition, the response degree of the high-risk and low-risk groups to traditional chemotherapy drugs was analyzed. It was found that there were significant differences among five traditional drugs, cisplatin, rapamycin, cyclopamine, GNF-2, and pyrimethamine, and the “high” group was more sensitive to these traditional drugs (Figure 8E, Wilcox.test).

Abnormal performance of RiskScore in tricarboxylic acid pathway

Further, the performance of RS in TCA-related pathways was compared. As shown in Figure 9A, it was noted that the score of the TCA-linked pathway in the “low” group was higher. Both the groups showed significant differences with regards to different pathways like the COBP_CITRATE_METABLIC_PROCESS, COBP_2_OXOGLUTARATE_METABLIC_PROCESS, COBP_ OXOGLUTARATE_

METABLIC_PROCESS, COBP_SUCCINATE_METABLIC _PROCESS, and COBP_TRICARBOXYLIC_ACID_CYCLE_ ENZYME_COMPLEX pathways. Then, the score of TCA-related pathways for every patient included in the TCGA-LIHC dataset

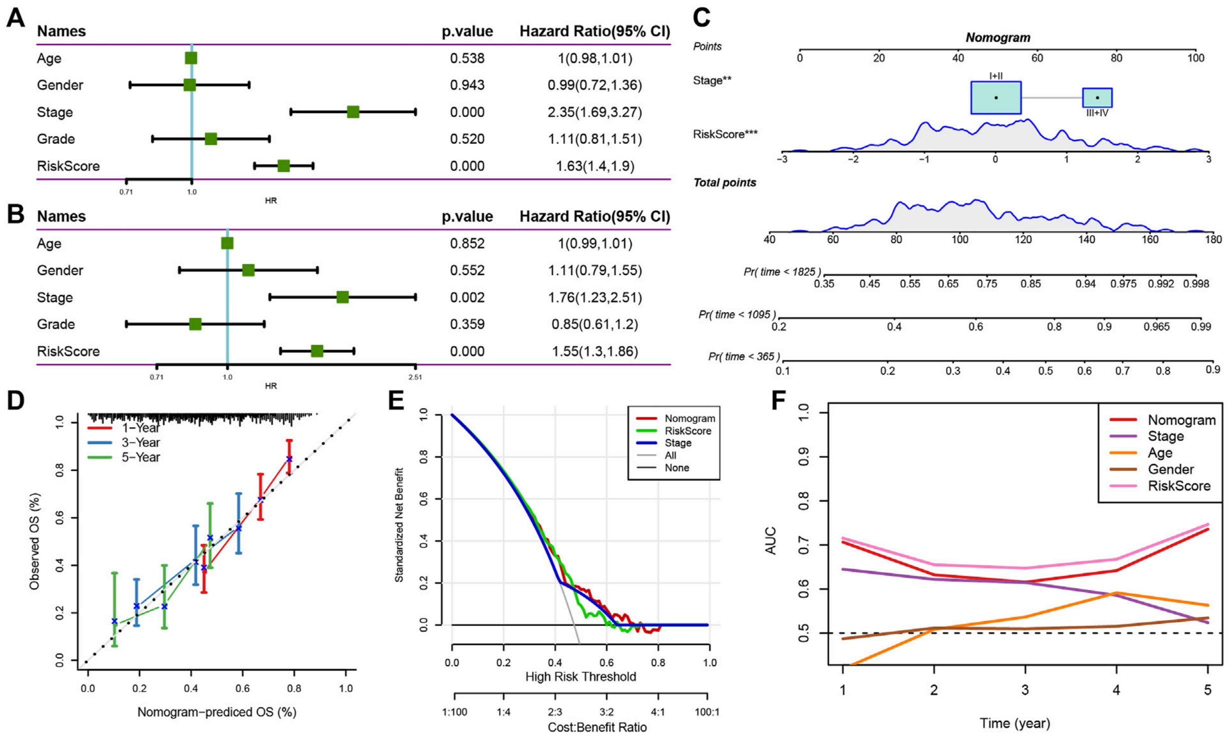


FIGURE 10
Nomogram of RS combined with clinical pathological characteristics. **(A,B)** Univariate and Multivariate Cox regression analysis of the RS and clinicopathological features; **(C)** Nomograph model; **(D)** Calibration curve of nomograph in 1, 3, and 5 years; **(E)** Decision curve of nomograph; **(F)** When compared to a few other clinicopathological characteristics, the nomogram displayed a good capacity for OS prediction.

was z-scored based on the samples in the heat map for determining the difference in the scores of the related pathways in the expression pathway of the high-risk and low-risk groups, and the TCA score of "low" group was higher (Figure 9B). The average value was calculated as the TCA score through the score of TCA-related pathways. Through analysis, it was found that the TCA score was significantly and negatively related to RS ($R = 0.49$, $p < 2.2e-16$) (Figure 9C, Spearman).

RiskScore combined with clinicopathological features to further improve the prognostic model and survival prediction

Clinicopathological characteristics and RS were analyzed using the Univariate and Multivariate Cox regression analysis, and the results revealed that Stage and RS were the most important predictive markers (Figures 10A,B). RS and Stage were coupled to create a nomogram in order to evaluate the risk assessment and the survival probability of the liver cancer patients (Figure 10C). According to the model's findings, RS had the biggest influence on the survival rate prediction. Then, the calibration curve was utilized to assess the model's predictability, as illustrated in Figure 10D. It can be seen that the three calibration points for the prediction calibration curves for 1, 3, and 5 years were close to the reference curve, indicating that the nomogram performed well in terms of prediction. Additionally, decision curve analysis (DCA) was utilized to investigate the model's dependability. It is evident that the advantages of RS and nomogram were much greater than those of the extreme curve. The nomogram and RS demonstrated the highest capacity to predict survival when compared to other clinicopathological characteristics (Figures 10E,F).

Discussion

The long-term efficacy of HCC treatment is still a significant problem in clinical practice because of the complexity of the TME and heterogeneity of HCC. The selection of the best course of treatment and action must be categorized and improved. In HCC, many transcriptome-based classifications are extensively used. Li et al. identified 2 novel molecular subgroups in liver cancer based on genes associated with ferroptosis and created a ferroptosis-associated prognostic RS model made up of six genes that can be used to predict outcomes and identify the cancer patients who would respond well to immunotherapy. Three molecular subtypes of HCC were discovered by Wang et al., each with a unique prognosis and metabolic profile. Lin et al. found 18 lncRNAs and 2 molecular subtypes in HCC with specific immune dysfunction that present distinct prognostic characteristics and immunological characteristics, which aids in understanding the function of lncRNA and motivates the discovery of immunotherapy targets. In this study, the HCC molecular subtypes were detected, from the cuproptosis perspective, since one cannot ignore the regulatory impact of cuproptosis.

From the literature review, 13 genes associated with cuproptosis were obtained. After analyzing the differences in the mRNA expression of cuproptosis-related genes between the primary tumor samples and adjoining normal tissue samples, it was noted that a majority of these variations were statistically significant. HCC patients were categorized into three categories depending on the consistent clustering of 13 cuproptosis-related gene expression profiles. The three subtypes had distinct prognosis characteristics, according to the prognostic analysis. The prognosis for Clust3 was the worst, followed by Clust2 and Clust1, while Clust1 showed the best prognosis. Additional examination of the clinicopathological traits of the various subtypes revealed that the Clust3 subtype showed a late clinical T stage and Stage, in addition to a higher Grade. All these findings were in agreement with its poor prognosis.

Further, the genomic variations displayed by the 3 molecular subtypes, identified using the TCGA-LIHC dataset were analyzed. The results implied that the clust1 subtype showed a lower Aneuploidy Score and Homologous Recombination Defect. Additionally, after comparing the correlation between the 5 existing immune molecular subtypes and the 3 molecular subtypes defined in this study, it was noted that among clust3 and clust2 subtypes with a poor prognosis had a low proportion of C4 subtype (lymphocyte depleted subtype) and a high percentage of C1 subtype (wound healing subtype), which was consistent with the prognosis of immune molecular subtypes. Then, the variations in the gene mutations existing between the different molecular subtypes were also identified, and significant differences were noted in the mutation frequencies of TP53, TTN, MUC16, and other genes among the 3 molecular subtypes, and clust1 had a higher mutation frequency of TP53.

Recent reports have stated that the concentration of the copper ions in the tumor tissues and serum of cancer patients was significantly higher than those of healthy patients (Blockhuys et al., 2017). Previous studies have shown that cuproptosis takes place by the direct combination of copper ions and the fatty acylated components present in the TCA cycle. This leads to the accumulation of the fatty acylated proteins and subsequent loss of the iron-sulfur cluster proteins, which leads to protein toxic stress and eventually cell death. Therefore, cuproptosis is closely related to the TCA cycle (Tsvetkov et al., 2022). TCA is seen to play a significant role in cellular energy metabolism and it is also responsible for the onset of numerous diseases, like tumors. At present, mutations and abnormal expression of TCA key genes have been found in tumors, which are significantly related to tumorigenesis and progression. The liver is an important digestive organ in the human body. The metabolic process of three major nutrients occurs actively in the liver. HCC is usually accompanied by the TCA cycle reprogramming, which regulates energy production through the TCA cycle, which ensures the survival of the tumor cells even in difficult conditions like hypoxia, nutrient deficiency, and finally, escaping the immune system (Ferrarini et al., 2019; Du et al., 2022). The score of TCA related pathway was calculated by the ssGSEA method. Significant differences were noted in the 3 molecular subtypes, with regard to the TCA cycle. Cell death is an essential and fine-tuning process, which is crucial to eliminating damaged and redundant cells. Many forms of programmed and non-programmed cell death have been identified, including apoptosis, ferroptosis, and necroptosis (Moujalled et al., 2021). This study could not detect any significant differences between the 3 molecular subtypes, except in necroptosis and apoptosis, and there were significant differences in the other four pathways related to cell growth and death.

The metabolic environment can change the immune response in the liver and make tumor cells immune escape. In addition, metabolic rearrangement of immune cells can cause abnormal self-function (Li et al., 2021). TME is a crucial intrinsic factor in the emergence, growth, invasion, and metastasis of liver cancer. The findings of this study showed that different molecular subtypes exhibited varying degrees of immune cell infiltration in the immunological microenvironment of different patients. Additionally, the "ImmuneScore" of the clust3 subtype, which has a poor prognosis due to the relatively low immune cell infiltration, was lower compared to that of other subtypes. Additionally, the immunotherapy sensitivity variations of several molecular subtypes in the TCGA-LIHC dataset were examined. The majority of immunological checkpoint genes were discovered to express differently in various subtypes. In each of the three molecular categories, the macrophages showed a significant difference. Macrophages are crucial for immunological regulatory processes such as the processing and presentation of antigens and the Toll-like receptor signaling pathway. Additionally, macrophages have FC receptors on their surface, which when activated by a specific antibody can cause an ADCC effect (NK cell-mediated cytotoxicity) that kills tumor cells (Xing et al., 2020). Through our analysis, macrophages had significant differences in the antigen processing and presentation pathway. Lu et al. (2022) showed that PD-L1 positive host macrophages, representing the main cell source of PD-L1 in HCC, showed HLA-DR^{high}CD86^{high} glycolysis phenotype, significantly produced anti-tumor IL-

12p70, and polarized through internal glycolysis metabolism. These results implied that the macrophages play a vital role in the onset of liver tumors, and the distribution differences between the three subtypes can provide a basis for tumor treatment. Further, by analyzing the sensitivity differences among different subtypes in immunotherapy, it was found that the clust2 and clust3 subtypes in the TCGA-LIHC dataset showed higher TIDE scores than those presented by clust1, suggesting that the clust1 subtype had a lower probability of immune escape and greater likelihood of benefiting from immunotherapy.

Then, a total of 15 genes among the three subtypes that had a significant impact on prognosis were identified, and the prognostic risk model was constructed by KIF2C, PTTG1, CENPM, CDC20, CYP2C9, SFN, CFHR3, and seven genes was obtained by Lasso regression and AIC algorithm. Kinesin Family member 2C (KIF2C) belongs to the kinesin 13 family, and is an M-kinesin, which is overexpressed in many human tumors. In their study, Wei et al. (2021) observed that KIF2C was overexpressed in HCC and was related to several aggressive malignancies that activate the Wnt/β-catenin signaling pathway and was also involved in the HCC progression as it interacted with TBC1D7 in mTORC1 signaling. Cho-Rok et al. (2006) found that Pituitary Tumor Transforming Gene 1 (PTTG1) was overexpressed in many types of human cancers. Furthermore, results indicated that when the PTTG1 gene was silenced, it inhibited the growth of the liver cells *in vivo* and *in vitro*. Studies have shown that centromeric protein M (CENPM) is closely related to the development of HCC. The up-regulation of CENPM promotes hepatocarcinogenesis through a variety of mechanisms and could be considered a new probable biomarker and a clinical therapeutic target for HCC (Xiao et al., 2019). Studies have found that CDC20 regulates the process of the cell cycle mainly by targeting the destruction of key substrates. In HCC, CDC20 binds to the Destruction box (D-box) motif in oxygen-dependent Prolyl Hydroxylase 3 (PHD3) to promote its polyubiquitination and degradation and is seen to play a vital role in HCC development by controlling PHD3 (Shi et al., 2021). Nizamuddin et al. found that cytochrome-P450-2C9 (CYP2C9) has genetic diversity. This gene metabolizes many drugs and is overexpressed in the human liver (Nizamuddin et al., 2021). Sulforaphane (SFN) plays an epigenetic regulatory role by inhibiting histone deacetylase (HDAC) and affects the activity of carcinogenic transcription factors through the methylation of its binding site motif, which provides insights into the chemopreventive molecular effects of SFN in HepG2 cells. It is a valuable natural cancer treatment method (Dos Santos et al., 2020). Complement factor H-related 3 (CFHR3) is a protein-coding gene that plays a role in various diseases. Liu et al. (2020) found through bioinformatics analysis that CFHR3 is a novel prognostic biomarker and therapeutic target for determining HCC.

Further, the relationship between RS scoring and clinical characteristics of liver cancer was analyzed. It was found that the samples with higher clinical grades had higher RSs. A comparison of the different immune microenvironments in the patients belonging to differing RS groups showed that the "low" group presented a high infiltration of immune cells, and a majority of the Immune checkpoint genes were expressed differentially in both groups. In addition, by analyzing the sensitivity difference between RS group to treatment, it is noted that the "high" group showed a higher TIDE score compared to the "low" group, indicating that the likelihood of immune escape in the high-risk group was higher compared to the low-risk group, and the high-risk group patients were less likely to be benefitted from immunotherapy. However, the "high" group was more sensitive to these traditional drugs. This result can provide a reference for personalized treatment of patients. Simultaneously, the performance of RS in TCA-related pathways was compared. The results in this study showed that the low-risk group showed higher TCA-related pathways scores, and the TCA scores were seen to be significantly negatively related to the RS, which was consistent with the results of subtype typing, and TCA scores with poor prognosis were higher. Finally, the clinical characteristics that showed significant differences during the Univariate and Multivariate Cox regression analysis, Stage, and RS were used for constructing a novel nomogram. Analyzing the calibration and the decision curves indicated that the model showed a higher prediction accuracy and survival prediction capacity. Additionally, the cuproptosis-linked genes were chosen as the target gene, which was essential for the onset, development, diagnosis, and treatment of HCC. The nomogram model constructed in this study could be used as the basis for deriving an individualized treatment plan for HCC patients.

This study provides novel insights into the personalized clinical treatment planning for HCC patients, however, it does have a few limitations. First of all, our research only includes bioinformatics analysis and lacks the verification of experimental clinical samples. In addition, the study was carried out using a retrospective design instead of using a prospective design. However, this analysis was carried out using 2 independent datasets, so the results are still acceptable and reliable. It can be concluded that prospective clinical trials and an investigation into the mechanisms involved need to be carried out for verifying the results noted in the study.

Conclusion

To conclude, this study presented 3 molecular subtypes that were associated with cuproptosis in liver cancer. These 3 molecular subtypes showed a heterogeneity in their pathological features, prognosis, pathway, and immune characteristics. Thereafter, a classifier known as the prognostic risk model associated with cuproptosis was constructed and verified. The model has strong stability, is independent of the clinical and pathological characteristics, and plays a stable prediction efficiency in independent data sets. The model has high prediction accuracy and survival prediction ability, which could be used for predicting prognosis and selecting the immunotherapy that was best suited for the patients. These results could help in developing a precise and individualized treatment strategy for clinical HCC patients.

Data availability statement

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

Author contributions

All authors contributed to this present work: ZL designed the study, YQ acquiresd the data. HW and QZ drafted the manuscript, ZW and WW revised the manuscript. All authors read and approved the manuscript.

Funding

This study was supported by the funding of School level natural science research project of Clinical Medicine School of Anhui Medical University (2020XJ017).

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.2022.1000652/full#supplementary-material

References

Blockhuys, S., Celauro, E., HILDESJö, C., Feizi, A., STåL, O., Fierro-GONZáLEZ, 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

Cho-Rok, J., Yoo, J., Jang, Y. J., Kim, S., Chu, I. S., Yeom, Y. I., et al. (2006). Adenovirus-mediated transfer of siRNA against PTTG1 inhibits liver cancer cell growth *in vitro* and *in vivo*. *Hepatology* 43, 1042–1052. doi:10.1002/hep.21137

Dos Santos, P., Machado, A. R. T., de Grandis, R. A., Ribeiro, D. L., Tuttis, K., Morselli, M., et al. (2020). Transcriptome and DNA methylation changes modulated by sulforaphane induce cell cycle arrest, apoptosis, DNA damage, and suppression of proliferation in human liver cancer cells. *Food Chem. Toxicol.* 136, 111047. doi:10.1016/j.fct.2019.111047

Du, D., Liu, C., Qin, M., Zhang, X., XI, T., Yuan, S., et al. (2022). Metabolic dysregulation and emerging therapeutical targets for hepatocellular carcinoma. *Acta Pharm. Sin. B* 12, 558–580. doi:10.1016/j.apsb.2021.09.019

Ferrarini, A., di Poto, C., He, S., Tu, C., Varghese, R. S., Kara Balla, A., et al. (2019). Metabolomic analysis of liver tissues for characterization of hepatocellular carcinoma. *J. Proteome Res.* 18, 3067–3076. doi:10.1021/acs.jproteome.9b00185

Guo, X., Sun, M., Gao, R., Qu, A., Chen, C., Xu, C., et al. (2021). Ultrasmall copper (I) sulfide nanoparticles prevent hepatitis B virus infection. *Angew. Chem. Int. Ed. Engl.* 60, 13073–13080. doi:10.1002/anie.202103717

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

Li, J., Tao, H., Wang, W., Li, J., and Zhang, E. (2022). The detection and verification of two heterogeneous subgroups and a risk model based on ferroptosis-related genes in hepatocellular carcinoma. *J. Oncol.* 2022, 1182383. doi:10.1155/2022/1182383

Li, X., Ramadori, P., Pfister, D., Seehawer, M., Zender, L., and Heikenwalder, M. (2021). The immunological and metabolic landscape in primary and metastatic liver cancer. *Nat. Rev. Cancer* 21, 541–557. doi:10.1038/s41568-021-00383-9

Liang, Q., Shen, X., and Sun, G. (2018). Precision medicine: Update on diagnosis and therapeutic strategies of hepatocellular carcinoma. *Curr. Med. Chem.* 25, 1999–2008. doi:10.2174/0929867325666180117101532

Lin, L., Yan, L., Liu, Y., Qu, C., Ni, J., and Li, H. (2020). The burden and trends of primary liver cancer caused by specific etiologies from 1990 to 2017 at the global, regional, national, Age, and sex level results from the global burden of disease study 2017. *Liver Cancer* 9, 563–582. doi:10.1159/000508568

Liu, H., Guo, H., Jian, Z., Cui, H., Fang, J., Zuo, Z., et al. (2020a). Copper induces oxidative stress and apoptosis in the mouse liver. *Oxid. Med. Cell. Longev.* 2020, 1359164. doi:10.1155/2020/1359164

Liu, J., Li, W., and Zhao, H. (2020b). CFHR3 is a potential novel biomarker for hepatocellular carcinoma. *J. Cell. Biochem.* 121, 2970–2980. doi:10.1002/jcb.29551

Lu, L. G., Zhou, Z. L., Wang, X. Y., Liu, B. Y., Lu, J. Y., Liu, S., et al. (2022). PD-L1 blockade liberates intrinsic antitumourigenic properties of glycolytic macrophages in hepatocellular carcinoma. *Gut* 2021, 326350. doi:10.1136/gutjnl-2021-326350

Moujalled, D., Strasser, A., and Liddell, J. R. (2021). Molecular mechanisms of cell death in neurological diseases. *Cell. Death Differ.* 28, 2029–2044. doi:10.1038/s41418-021-00814-y

Nakaichi, M., Iseri, T., Horikirizono, H., Komine, M., Itoh, H., Sunahara, H., et al. (2021). Copper-associated hepatitis in a young Dalmatian dog in Japan. *J. Vet. Med. Sci.* 83, 911–915. doi:10.1292/jvms.21-0061

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

Nizamuddin, S., Dubey, S., Singh, S., Sharma, S., Machha, P., and Thangaraj, K. (2021). CYP2C9 variations and their pharmacogenetic implications among diverse south asian populations. *Pharmgenomics. Pers. Med.* 14, 135–147. doi:10.2147/PGPM.S272015

Pham, L., Kyritsi, K., Zhou, T., Ceci, L., Baiocchi, L., Kennedy, L., et al. (2022). The functional roles of immune cells in primary liver cancer. *Am. J. Pathol.* 192, 826–836. doi:10.1016/j.ajpath.2022.02.004

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

Runa, F., Hamalian, S., Meade, K., Shisgal, P., Gray, P. C., and Kelber, J. A. (2017). Tumor microenvironment heterogeneity: Challenges and opportunities. *Curr. Mol. Biol. Rep.* 3, 218–229. doi:10.1007/s40610-017-0073-7

Shen, W., Song, Z., Xiao, Z., Huang, M., Shen, D., Gao, P., et al. (2022). Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. *iMeta* 1. doi:10.1002/imt2.36

Shi, M., Dai, W. Q., Jia, R. R., Zhang, Q. H., Wei, J., Wang, Y. G., et al. (2021). APC(CDC20)-mediated degradation of PHD3 stabilizes HIF-1a and promotes tumorigenesis in hepatocellular carcinoma. *Cancer Lett.* 496, 144–155. doi:10.1016/j.canlet.2020.10.011

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

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830. e14. doi:10.1016/j.immuni.2018.03.023

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

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

Wei, S., Dai, M., Zhang, C., Teng, K., Wang, F., Li, H., et al. (2021). KIF2C: A novel link between wnt/β-catenin and mTORC1 signaling in the pathogenesis of hepatocellular carcinoma. *Protein Cell.* 12, 788–809. doi:10.1007/s13238-020-00766-y

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

Xiao, Y., Najeeb, R. M., Ma, D., Yang, K., Zhong, Q., and Liu, Q. (2019). Upregulation of CENPM promotes hepatocarcinogenesis through mutiple mechanisms. *J. Exp. Clin. Cancer Res.* 38, 458. doi:10.1186/s13046-019-1444-0

Xing, Z., Afkhami, S., Bavananthasivam, J., Fritz, D. K., D'Agostino, M. R., Vaseghi-Shanjani, M., et al. (2020). Innate immune memory of tissue-resident macrophages and trained innate immunity: Re-Vamping vaccine concept and strategies. *J. Leukoc. Biol.* 108, 825–834. doi:10.1002/JLB.4MR0220-446R

Xu, R., Liang, J., Cheng, M., Wu, H., Wu, H., Cao, S., et al. (2021). Liver and urine metabolomics reveal the protective effect of Gandou decoction in copper-laden Hepatolenticular degeneration model rats. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1179, 122844. doi:10.1016/j.jchromb.2021.122844

Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Ann. Transl. Med.* 4, 136. doi:10.21037/atm.2016.03.35

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fgene.2022.1007108

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Ying Guo,
Maternity and Child Health Hospital of Qinhuangdao, China
Wenjie Shi,
Otto von Guericke University Magdeburg, Germany

*CORRESPONDENCE
Jianjun Wang,
jjwang0524@sina.com
Chunyan Dong,
cy_dong@tongji.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 29 July 2022
ACCEPTED 15 September 2022
PUBLISHED 13 October 2022

CITATION
Qiang S, Fu F, Wang J and Dong C (2022), Definition of immune molecular subtypes with distinct immune microenvironment, recurrence, and PANoptosis features to aid clinical therapeutic decision-making.
Front. Genet. 13:1007108.
doi: 10.3389/fgene.2022.1007108

Definition of immune molecular subtypes with distinct immune microenvironment, recurrence, and PANoptosis features to aid clinical therapeutic decision-making

Sufeng Qiang[1,2†], Fei Fu[2†], Jianjun Wang[2]* and Chunyan Dong[3,4]*

[1]Department of Gynaecology and Obstetrics, Shanghai East Hospital, Nanjing Medical University, Shanghai, China, [2]Department of Gynaecology and Obstetrics, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, [3]Breast Cancer Center, Shanghai East Hospital, Nanjing Medical University, Shanghai, China, [4]Breast Cancer Center, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China

Objective: Cervical cancer poses a remarkable health burden to females globally. Despite major advances in early detection and treatment modalities, some patients still relapse. The present study proposed a novel immune molecular classification that reflected distinct recurrent risk and therapeutic responses in cervical cancer.

Methods: We retrospectively collected two cervical cancer cohorts: TCGA and GSE44001. Consensus clustering approach was conducted based on expression profiling of recurrence- and immune-related genes. The abundance of immune cells was inferred *via* five algorithms. Immune functions and signatures were quantified through ssGSEA. Genetic mutations were analyzed by maftools package. Immunotherapeutic response was inferred *via* tumor mutation burden (TMB), Tumor Immune Dysfunction and Exclusion (TIDE), and Submap methods. Finally, we developed a LASSO model for recurrence prediction.

Results: Cervical cancer samples were categorized into two immune subtypes (IC1, and IC2). IC2 exhibited better disease free survival (DFS), increased immune cell infiltration within the immune microenvironment, higher expression of immune checkpoints, higher activity of immune-relevant pathways (APC co-inhibition and co-stimulation, inflammation-promoting, MHC class I, IFN

Abbreviations: IRGs, immune-related genes; TCGA, the Cancer Genome Atlas; GEO, Gene Expression Omnibus; DFS, disease free survival; HR, hazard ratio; CDF, cumulative distribution function; DEGs, differential expressed genes; FC, fold change; FDR, false discovery rate; GO, Gene Ontology; KEGG, functional annotation and Kyoto Encyclopedia of Genes and Genomes; TMB, tumor mutation burden; TIDE, Tumor Immune Dysfunction and Exclusion; CTL, cytotoxic T lymphocyte; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; IC50, half maximal inhibitory concentration; LASSO, least absolute shrinkage and selection operator; AUROC, area under the time-dependent receiver-operating characteristic; DCA, decision curve analysis.

response, leukocyte and stromal fractions, macrophage regulation, and TCR Shannon), and higher frequencies of genetic mutations. This molecular classification exhibited a remarkable difference with existing immune subtypes, with diverse PANoptosis (pyroptosis, apoptosis and necroptosis) features. Patients in IC2 were more likely to respond to immunotherapy and targeted, and chemotherapeutic agents. The immune subtype-relevant signature was quantified to predict patients' recurrence risk.

Conclusion: Altogether, we developed an immune molecular classification, which can be utilized in clinical practice to aid decision-making on recurrence management.

KEYWORDS

cervical cancer, recurrence, immune subtypes, immune microenvironment, therapeutic response, PANoptosis

Introduction

Cervical cancer poses a significant health burden to females worldwide (Castle et al., 2021). Although this malignancy is preventable and treatable, it remains the fourth most diagnosed cancer as well as the fourth major cause of cancer-related deaths in females (Sung et al., 2021). It is estimated that around 600,000 females are diagnosed with cervical cancer, and over 300,000 females die from this malignancy globally each year (Sung et al., 2021). Most cervical cancer cases correlate to infection with high-risk HPV. Primary treatment of cervical cancer that typically includes surgery, chemoradiation, or their combination has a cure rate of about 95% for early-stage disease as well as 40%–60% for locally advanced disease (Miccò et al., 2022). Recurrent cervical cancer is defined as local tumor regrowth, lymph node or distant metastasis following the primary tumor has regressed for at least 6 months (Miccò et al., 2022). Management of recurrent cervical cancer depends upon treatment history, location, and degree of recurrence. Up to 26% of patients with early-stage cervical cancer relapse following initial surgery (Cibula et al., 2022). Furthermore, many patients with treatment history experience recurrence without symptoms.

Accumulated evidence demonstrates that the complex cellular compositions within the immune microenvironment result in intratumoral heterogeneity (Qiu et al., 2020). Immune-related genes (IRGs) correlate to the immune microenvironment, clinical outcomes as well as treatment response in cervical cancer (Xu et al., 2021). PANoptosis is an inflammatory programmed cell death, which can be activated by components that are simultaneously involved in pyroptosis, apoptosis and/or necroptosis (Karki et al., 2020). If one or more programmed cell death pathways are hindered in tumor cells, PANoptosis is beneficial for tumor suppression *via* inducing mechanisms by which the host activates alternative cell death defense mechanisms (Lee et al., 2021). Currently, the use of immunotherapy to treat cervical cancer is being actively researched, though several immunotherapy drugs (pembrolizumab, etc.), have gained the approval of the FDA (Colombo et al., 2021). Because immunotherapies have established a new standard of care in cervical cancer treatment, novel biomarkers to recognize ideal patient populations for these therapies are of importance. In the present study, in accordance with the expression profiling of recurrence-related IRGs, we proposed a novel immune molecular classification, and classified cervical cancer patients into two immune subtypes with distinct recurrence risk, immune microenvironment as well as immuno-, targeted- and chemotherapeutic responses, thus aiding clinical therapeutic decision-making.

Materials and methods

Data collection and processing

RNA sequencing, somatic mutation data (MAF format) and clinical follow-up information of cervical cancer patients were downloaded from the Cancer Genome Atlas (TCGA; https://portal.gdc.cancer.gov/). After thoroughly querying the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds/), an eligible cervical cancer dataset (GSE44001) with disease recurrence information and genetic profiling was enrolled (Lee et al., 2013). All data were downloaded on 15 April 2022. After removal of samples without recurrence time and status (i.e. disease free survival (DFS)), we included 174 samples in the TCGA dataset and 300 samples in the GSE44001 dataset for subsequent analysis. The demographics and follow-up data were listed in Supplementary Table S1. A total of 782 IRGs were collected from previously published literature, as listed in Supplementary Table S2 (Charoentong et al., 2017). In addition, we collected the gene sets of PANoptosis (pyroptosis, apoptosis and necroptosis) were acquired from previously published literature (Supplementary Table S3) (Pan et al., 2022).

Unsupervised consensus clustering analysis

Univariate cox regression analysis of IRGs was implemented based on recurrence time and status both in the TCGA and GSE44001 datasets. In accordance with $p < 0.05$, recurrence-related IRGs were obtained, with hazard ratio (HR)>1 as a risk factor and HR < 1 as a protective factor. Then, protective and risk factors of the two datasets were separately intersected for subsequent consensus clustering analysis. Unsupervised consensus clustering was implemented for constructing an immune molecular classification using ConsensusClusterPlus package (version 1.52.0) based on recurrence-related IRGs (Wilkerson and Hayes, 2010). The clustering procedure, with 100 iterations, was carried out by subsampling 80% of the data in each iteration. The optimal number of clusters was identified based on consensus cumulative distribution function (CDF), relative change in area under CDF curve, and consensus heatmap.

Differential expression and functional enrichment analyses

By limma package (Ritchie et al., 2015), differential expressed genes (DEGs) between subtypes were selected according to |fold change (FC)|>1.5 and false discovery rate (FDR)<0.05. Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of up-regulated and down-regulated genes were carried out using clusterProfiler package (Yu et al., 2012), respectively, with FDR<0.05 as the threshold value. The GSEA function of clusterProfiler package was also employed to conduct functional enrichment analysis between subtypes. The "c2.cp.kegg.v7.4.symbols.gmt" and "h.all.v7.4.symbols.gmt" gene sets in the Molecular Signatures database (http://www.gsea-msigdb.org/) (Liberzon et al., 2015) were utilized as reference sets.

Immune cell infiltrations

EPIC (Racle et al., 2017), Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) (Yoshihara et al., 2013), MCP-counter (Becht et al., 2016), single-sample gene set enrichment analysis (ssGSEA) (Hänzelmann et al., 2013), and Tumor Immune Estimation Resource (TIMER) (Li et al., 2017) algorithms were implemented for inferring the relative proportions of infiltrating immune cells. The activity of known immune functions or signatures was estimated with ssGSEA approach.

Mutational analysis

The mutation frequencies of genes were calculated and visualized utilizing maftools package (version 2.4.05) (Mayakonda et al., 2018). Tumor mutation burden (TMB), a promising biomarker for immunotherapeutic response, was computed as the total number of nonsynonymous mutations in the coding region per megabase (Wang et al., 2021).

Prediction of immunotherapeutic response

Tumor Immune Dysfunction and Exclusion (TIDE) was computed in accordance with two main mechanisms of tumor immune escape: inducing T cell dysfunction in tumor tissue with increased cytotoxic T lymphocyte (CTL) infiltration as well as preventing T cell infiltration within tumor tissue with reduced CTL level (Jiang et al., 2018; Fu et al., 2020). The immunotherapeutic response was predicted with TIDE algorithm on the basis of gene expression profiling. Submap approach was also utilized to infer the responses (complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD)) to immune checkpoint blockade of PD-1 and CTLA4 from the IMvigor210 cohort (Yang et al., 2020). The demographics and follow-up information of the IMvigor210 cohort was shown in Supplementary Table S4. FDR < 0.05 was regarded as the threshold for a significant response or nonresponse to anti-PD1 or anti-CTLA4 treatment.

Evaluation of sensitivity to targeted and chemotherapeutic agents

Half maximal inhibitory concentration (IC50) values of AKT inhibitor VIII, Cisplatin, Erlotinib, Lapatinib, Paclitaxel, and Temozolomide were estimated to reflect therapeutic response with pRRophetic package (Geeleher et al., 2014) on the basis of the Genomics of Drug Sensitivity in Cancer database (www.cancerRxgene.org) (Yang et al., 2013), and the prediction accuracy was assessed with 10-fold cross-validation.

Establishment of an immune subtype-relevant signature

Cervical cancer patients in the GSE44001 dataset were randomly classified into the training and testing sets at a ratio of 1:1. Firstly, in the training set, univariate cox regression analysis of DEGs was conducted by survival package (version 3.1-12), with $p < 0.05$ as recurrence-related DEGs. Afterwards, the least absolute shrinkage and selection operator (LASSO) algorithm was implemented utilizing glmnet package (Engebretsen and Bohlin, 2019). Genes with regression coefficient equal to zero after the shrinkage process were removed. The optimal tuning parameter lambda (λ) was selected when the partial likelihood deviance reached the lowest on the basis of 10-fold cross-validation. The final identified genes were utilized to establish a multivariate cox regression model. The formula of the recurrence model was as follows:

$$\text{RiskScore} = \sum_{i=1}^{n} coef\,(i) * gene\,(i),$$

where coef(i) represents the coefficient of the ith gene, and gene(i) represents the expression level of the ith gene. RiskScore of each cervical cancer patient was computed. The optimal cutoff was determined *via* surv_cutpoint function of survminer package, and patients were classified into high- and low-risk subgroups. Kaplan-Meier approach was utilized to assess DFS, whereas log-rank test was implemented for assessing recurrence risk. The area under the receiver-operating characteristic curve (AUROC) was used to appraise the predictive capacity of the immune subtype-relevant signature. Above analyses were validated in the GSE44001-testing set, GSE44001-entire set, and independent TCGA dataset. Associations of the signature and clinicopathological parameters with prognosis were estimated through uni- and multivariate cox regression analyses in the TCGA dataset.

Establishment of a nomogram

A nomogram that integrated the immune subtype-relevant signature and clinicopathological parameters for recurrence prediction was constructed using rms package. The calibration curve was plotted for evaluating the predictive accuracy of the nomogram *via* rms package. Decision curve analysis (DCA) was conducted for determining the clinical application value of the nomogram *via* computing the net benefits at distinct risk thresholds.

Statistical analysis

Statistical analysis was executed *via* R software (version 4.1.0; https://www.r-project.org/). To compare continuous variables between groups, Student's t-test was implemented, whereas Wilcoxon test was applied to compare non-normally distributed variables. Through chi-squared test, categorical variables between groups were compared. The two-sided $p \leq 0.05$ was regarded as statistical significance.

Results

Construction of immune subtypes for cervical cancer with different recurrence outcomes based on recurrence-related immune-related genes

Both in the TCGA and GSE44001 datasets, we determined 7 protective factors and 2 risk factors for cervical cancer recurrence among IRGs by implementing univariate cox regression analysis (Figure 1A), which were included for consensus clustering analysis. Consensus matrix heatmap, CDF and relative change in area under CDF curve showed that cervical cancer samples in the TCGA dataset were clearly classified as two immune subtypes (namely IC1 and IC2) (Figure 1B; Supplementary Figures S1A,B). This classification was confirmed in the GSE44001 dataset (Supplementary Figure S2A), indicating that the clustering of samples was stable and reliable. There was a remarkably difference in DFS outcome between immune subtypes both in the TCGA (Figure 1C) and GSE44001 (Supplementary Figure S2B) datasets, with a more favorable DFS outcome for IC2. Clinicopathological features were also compared between immune subtypes. The proportion of non-recurrent patients was remarkably higher in IC2 than that of IC1 both in the TCGA (Figures 1D,E) and GSE44001 (Supplementary Figures S2C–E) datasets.

Immune subtypes-relevant genes and their biological significance

To unveil the molecular mechanisms underlying immune subtypes, DEGs between IC1 and IC2 were selected with |FC| >1.5 and FDR<0.05. In the TCGA dataset, 1223 DEGs with up-regulation and 976 with down-regulation were determined in IC1 compared with IC2 (Supplementary Table S5). Moreover, 138 DEGs with up-regulation and 539 with down-regulation were investigated in IC1 compared with IC2 in the GSE44001 dataset (Supplementary Table S6). DEGs with down-regulation were remarkably linked with immune-relevant pathways (cytokine, chemokine and their receptor signaling, etc.; Figure 2A; Supplementary Figure S3A). Moreover, DEGs with up-regulation were mainly correlated to cervical cancer-relevant pathways (Hippo signaling pathway, etc.; Figure 2B; Supplementary Figure S3B). GSEA was also employed for pathways activated in IC1 and IC2. Both in the TCGA and GSE44001 datasets, epithelial mesenchymal transition, pancreas beta cells and O-glycan biosynthesis were remarkably activated in IC1 (Figure 2C; Supplementary Figure S3C). Moreover, B cell receptor, chemokine, and T cell receptor signaling pathways and primary immunodeficiency displayed remarkable activation in IC2.

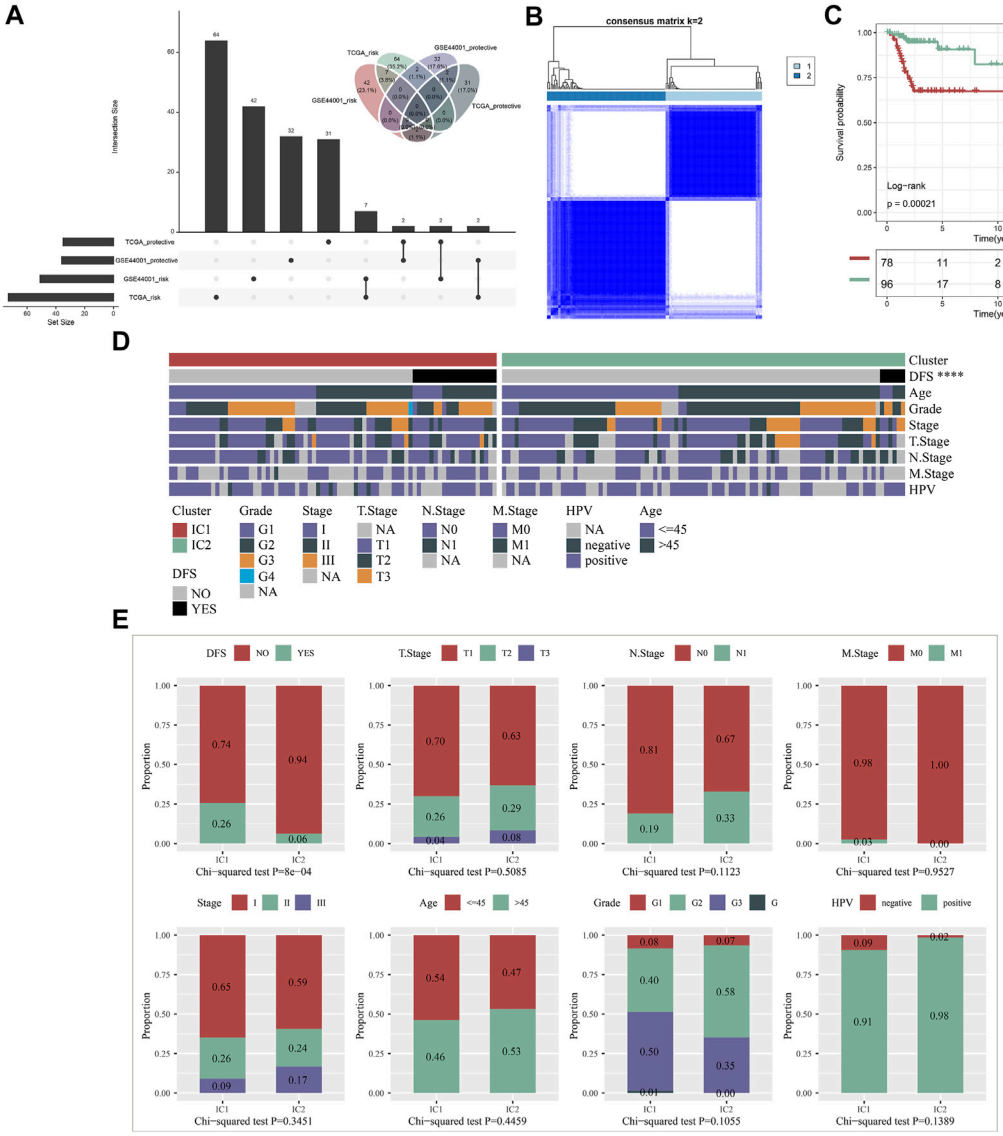


FIGURE 1

Construction of immune subtypes for cervical cancer with different recurrence outcomes based on recurrence-related IRGs in the TCGA dataset. **(A)** Univariate cox regression analysis of protective factors and risk factors for cervical cancer recurrence among IRGs in the TCGA and GSE44001 datasets. **(B)** Consensus matrix heatmap of cervical cancer samples based on the expression profiling of recurrence-related IRGs when k = 2. **(C)** Kaplan-Meier curves of DFS between IC1 and IC2 in the TCGA dataset. **(D)** Landscape of clinicopathological parameters in two immune subtypes (*****p* < 0.0001). **(E)** Comparison of DFS, T, N, M, histological stage, age, grade, and HPV between immune subtypes.

Immune subtypes with distinct immune microenvironment and immune checkpoints

Five approaches (MCP-counter, ESTIMATE, ssGSEA, EPIC, and TIMER) were applied for inferring the abundance of immune cells across cervical cancer from the TCGA and GSE44001 datasets. The consistent results from distinct approaches showed that IC2 exhibited the higher abundance of immune cells in comparison to IC1 both in two datasets (Figures 3A,B). Moreover, we acquired known immune checkpoints from previously published literature (Danilova et al., 2019; Li et al., 2019; Zheng et al., 2020). Both in the TCGA and GSE44001 datasets, BTLA, CD244, CD274, CD28, CD40, CTLA4, ICOS, PDCD1, PDCD1LG2, and TIGIT displayed higher expression in IC2 in comparison to IC1 (Figures 3C,D).

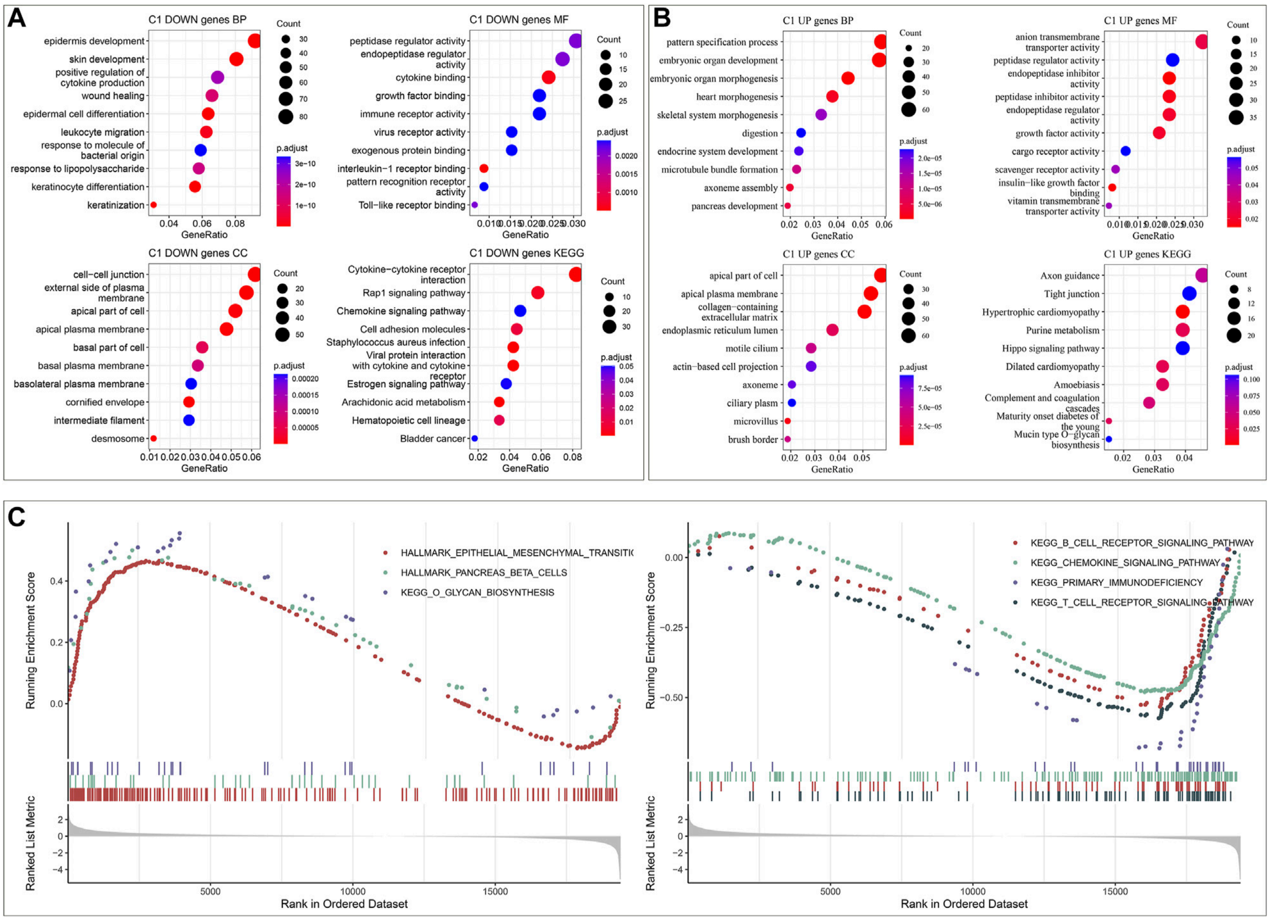


FIGURE 2
Immune subtypes-relevant genes and their biological significance in the TCGA dataset. **(A)** The main biological process (BP), molecular function (MF), cellular component (CC) and KEGG enrichment results of genes with down-regulation in IC1. **(B)** The main BP, MF, CC and KEGG enrichment results of genes with up-regulation in IC1. **(C)** GSEA for the main Hallmark and KEGG pathways activated in IC1 or IC2.

Immune subtypes with distinct immune functions

By applying ssGSEA approach, we inferred the activity of immune-relevant pathways across cervical cancer. In comparison to IC1, we observed that APC co-inhibition and co-stimulation, inflammation-promoting, MHC class I, type I and II IFN responses exhibited increased activity in IC2 both in the TCGA and GSE44001 datasets (Figures 4A,B).

Associations of our molecular classification with existing immune subtypes

As depicted in Figure 5A, our molecular classification exhibited a remarkable difference with existing immune subtypes (Ceccarelli et al., 2016). Compared with IC1, higher proportion of C2 and lower proportion of C1 were observed in IC2 (Figure 5B). However, there was no remarkable difference in DFS among existing immune subtypes (C1, C2, and C4) (Figure 5C). Thus, immune subtypes we proposed was a novel molecular classification of cervical cancer different from existing immune subtypes. The activity of immune signatures was compared between IC1 and IC2. Compared with IC1, IFN gamma response, leukocyte fraction, macrophage regulation, stromal fraction, and TCR Shannon exhibited higher activity in IC2 (Figure 5D).

Immune subtypes with distinct genetic mutations

Through maftools approach, we analyzed and visualized genetic mutations of cervical cancer. The mutation frequencies of the top 500 mutated genes were compared between IC1 and

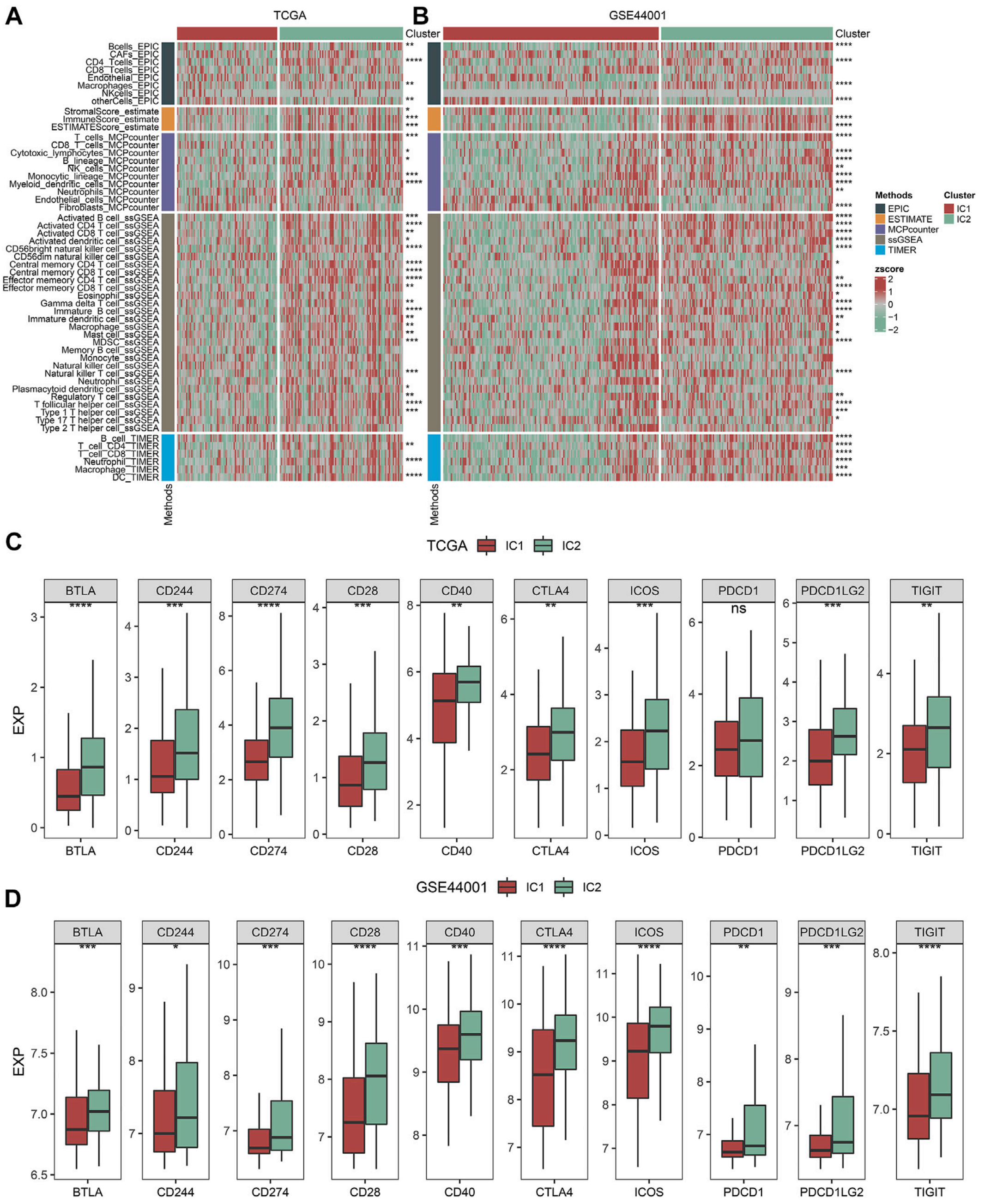


FIGURE 3
Immune subtypes with distinct immune microenvironment and immune checkpoints. **(A,B)** Heatmaps of the abundance of immune cells inferred by several approaches in IC1 and IC2 immune subtypes in the TCGA and GSE44001 datasets. **(C,D)** Expression of known immune checkpoints in two immune subtypes in the TCGA and GSE44001 datasets (ns: no significance; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$).

IC2 immune subtypes. With $p < 0.05$, mutations in 18 genes differed significantly between immune clustered (Figure 6A). Compared with IC1, MUC17, SYNE1, MYH15, PRKDC, RALGAPA1, ZNF91, CDK12, DGKK, MAPK1, ANAPC1, EPG5, FRYL, MIA3, WDR44, COL15A1, KAT6A, and LAMA3 displayed higher mutation frequencies in IC2. However, lower mutation frequency of TP73 was observed in IC2. We also computed TMB of each cervical cancer, and analyzed the difference in TMB between immune subtypes. In Figure 6B, compared with IC1, higher TMB score was observed in IC2. Especially, we compared mutation frequencies of SYNE1 and MAPK1 between immune subtypes. Higher

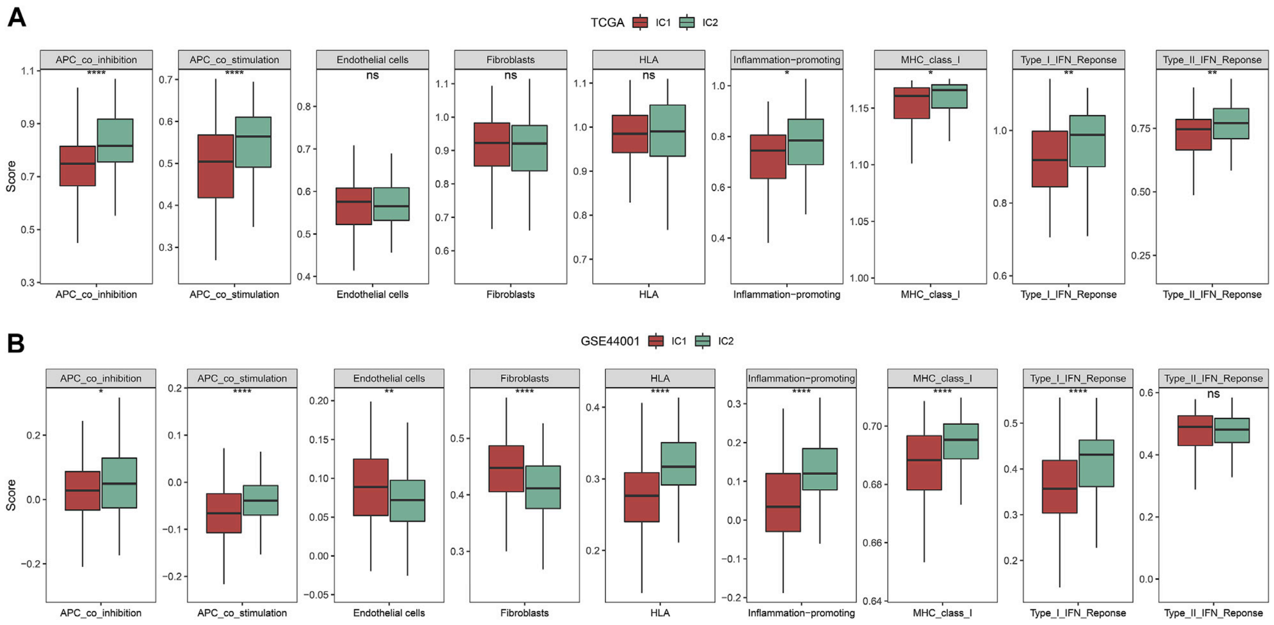


FIGURE 4
Immune subtypes with distinct immune-relevant pathways. **(A,B)** The activity of immune-relevant pathways in IC1 and IC2 immune subtypes in the TCGA and GSE44001 datasets (ns, no significance; **p* < 0.05; ***p* < 0.01; *****p* < 0.0001).

mutation frequencies of SYNE1 and MAPK1 were found in IC2 compared with IC1 (Figures 6C,D). Altogether, our data demonstrated that IC2 exhibited higher genetic mutations than IC1.

Immune subtypes with distinct immunotherapeutic responses

TIDE is a reliable biomarker to predict response to immunotherapy (Jiang et al., 2018; Fu et al., 2020). The lower TIDE score, the greater the likelihood of immunotherapeutic response. Here, we computed TIDE in cervical cancer from the TCGA and GSE44001 datasets. Higher dysfunction, and lower exclusion score were observed in IC2 than IC1 (Figures 6E,F) both in two datasets. There was no remarkable difference in TIDE score between immune subtypes in the TCGA dataset. Differently, IC2 exhibited lower TIDE score in comparison to IC1. Altogether, our data demonstrated that patients in IC2 were more likely to respond to immunotherapy. Furthermore, we employed Submap approach to compare the expression profiling of two immune subtypes with that of the IMvigor210 anti-PD-L1 immunotherapy dataset. Both in the TCGA and GSE44001 datasets, the expression profiling of IC2 was similar to that of patients who responded to anti-PD-L1 immunotherapy (Figures 7A,B). Thus, patients in IC2 were more suitable for immunotherapy, which were similar to TIDE prediction.

Immune subtypes with distinct targeted, and chemotherapeutic responses

Then, we analyzed the difference in targeted, and chemotherapeutic responses between IC1 and IC2 immune subtypes. Both in the TCGA and GSE44001 datasets, lower IC50 values of AKT inhibitor VIII, Cisplatin, Erlotinib, Lapatinib, Paclitaxel, and Temozolomide were observed in IC2 compared with IC1 (Figures 7C,D). This indicated that patients in IC2 were more likely to respond to above targeted, and chemotherapeutic agents.

Construction of an immune subtype-relevant recurrence model for cervical cancer

A total of 212 DEGs between immune subtypes were shared by the TCGA and GSE44001 datasets (Figure 8A). Among them, 26 DEGs were remarkably linked with recurrence of cervical cancer, which were used for subsequent recurrence model establishment (Supplementary

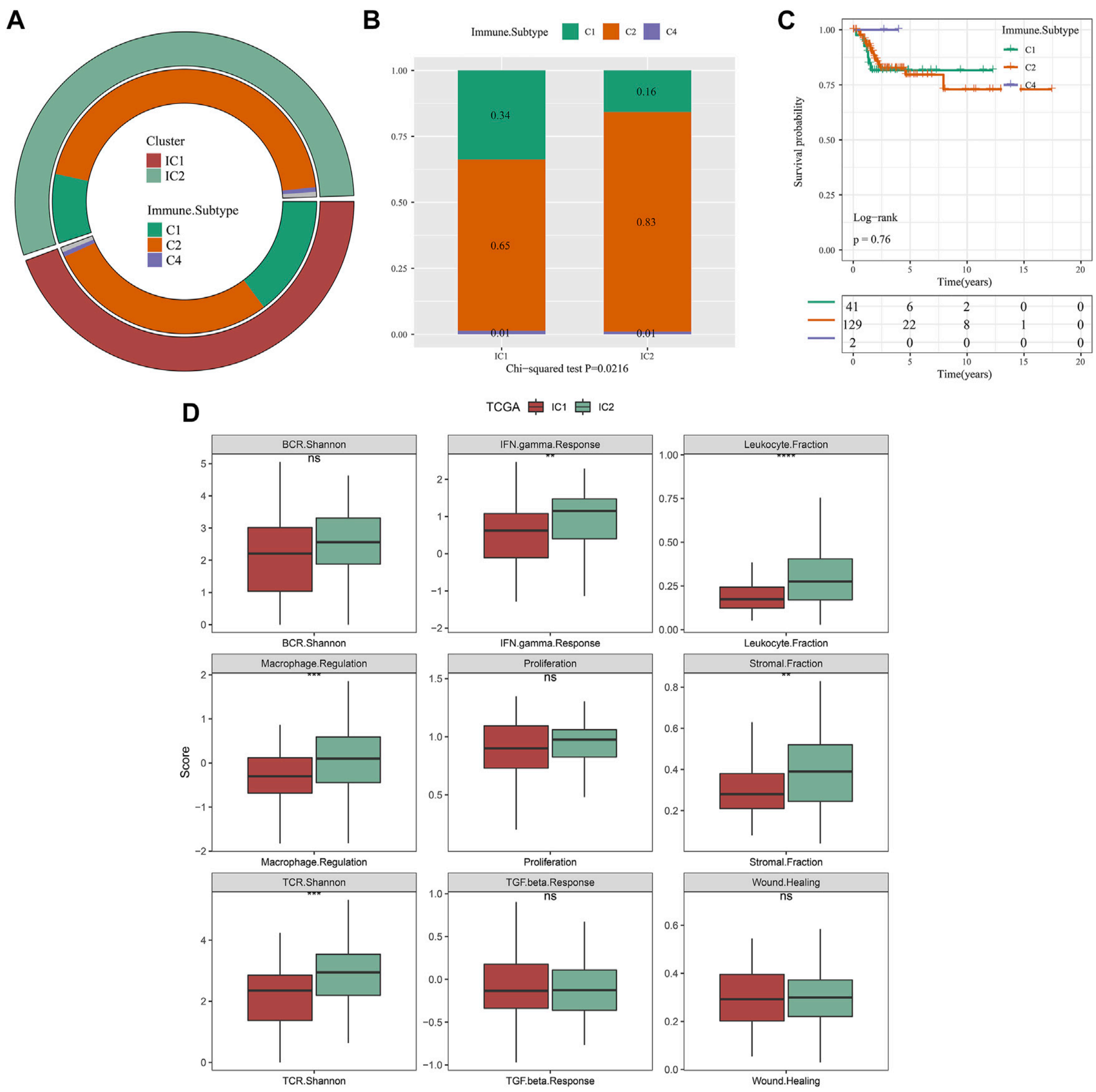


FIGURE 5
Associations of our molecular classification with existing immune subtypes in the TCGA dataset. **(A)** The distribution of existing immune subtypes (C1, C2 and C4) across our molecular classification. **(B)** Comparison of proportion of existing immune subtypes between IC1 and IC2 immune subtypes. **(C)** Kaplan-Meier curves of DFS among existing immune subtypes (C1, C2 and C4). **(D)** The activity of immune signatures between IC1 and IC2 immune subtypes (ns, no significance; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$).

Table S7). The GSE44001 dataset was randomly classified into the training and testing sets. In the training set, we applied LASSO algorithm to remove DEGs with regression coefficient equal to zero (Figure 8B). Moreover, based on 10-fold cross-validation, the optimal tuning parameter λ was 0.0271963 when the partial \likelihood deviance reached the lowest (Figure 8C). Finally, TMEM125, TFF1, DECR2, LONRF3, DAPL1, and ANKRD35 were included for establishing a multivariate cox regression model. By combining regression coefficients and expression values of above DEGs, we computed risk score of cervical cancer cases (Table 1). With the optimal cutoff, patients were classified into high- and low-risk subgroups. In Figure 8D, we observed that high-risk cases exhibited worse DFS in comparison to low-risk cases in the training set. AUROC values at 1-, 3- and 5-year DFS were all >0.60, demonstrating the excellent performance of this model in predicting recurrence (Figure 8E). The similar results were observed in the testing set (Figures 8F,G), the GSE44001 (Figures 8H,I) and TCGA (Figures 8J,K) datasets. Thus, the immune subtype-relevant recurrence model exhibited the favorable efficiency in predicting cervical cancer recurrence.

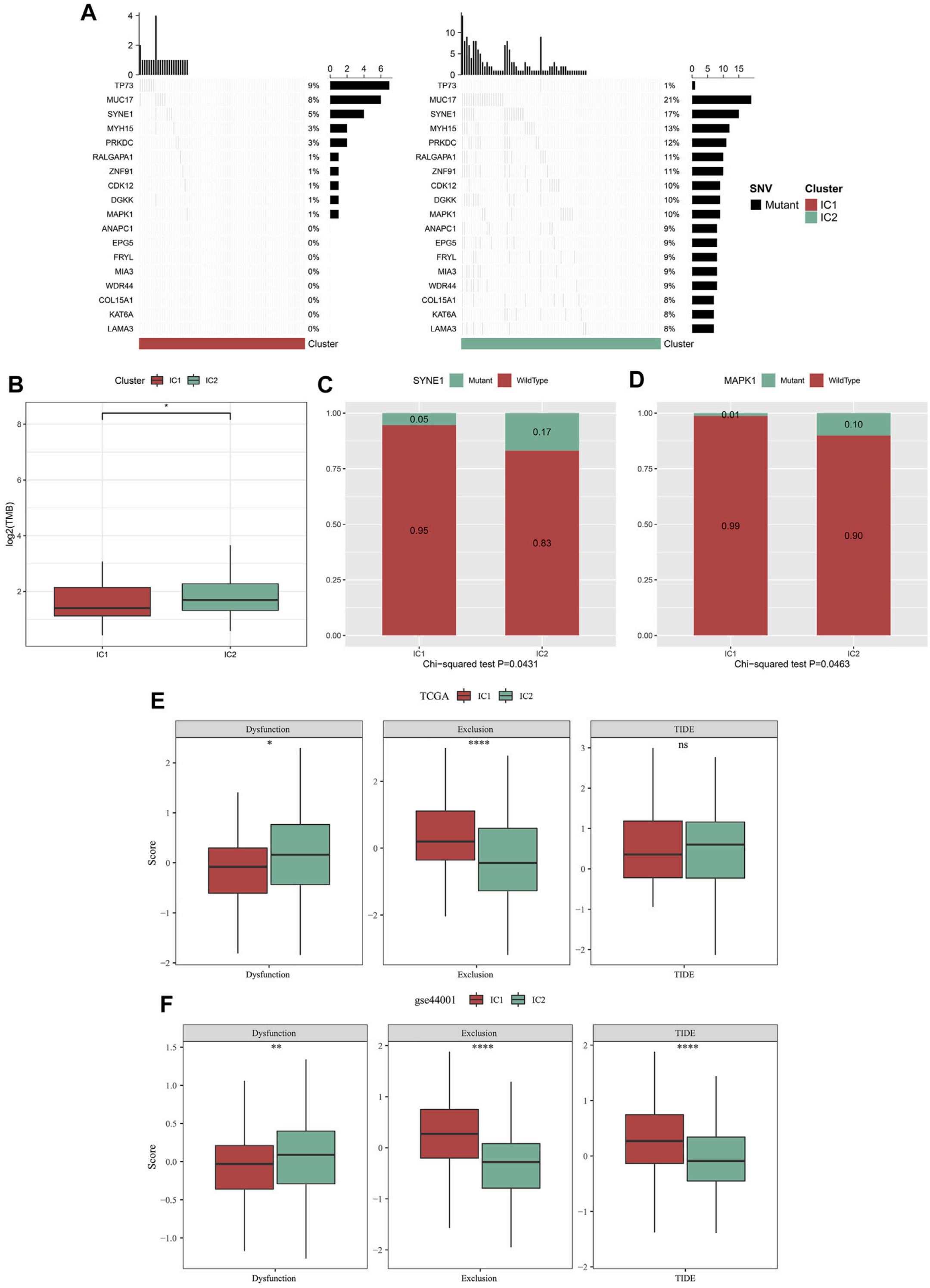


FIGURE 6
Immune subtypes with distinct genetic mutations and immunotherapeutic responses. **(A)** Waterfall diagrams of the mutational frequencies of genes with remarkable differences between IC1 and IC2 immune subtypes in the TCGA dataset. **(B)** Comparison of TMB between two immune subtypes in the TCGA dataset. **(C,D)** Comparison of mutation frequencies of SYNE1 and MAPK1 between immune subtypes in the TCGA dataset. **(E,F)** Comparison of dysfunction, exclusion, and TIDE score between immune subtypes in the TCGA and GSE44001 datasets (ns, no significance; $*p < 0.05$; $**p < 0.01$; $****p < 0.0001$).

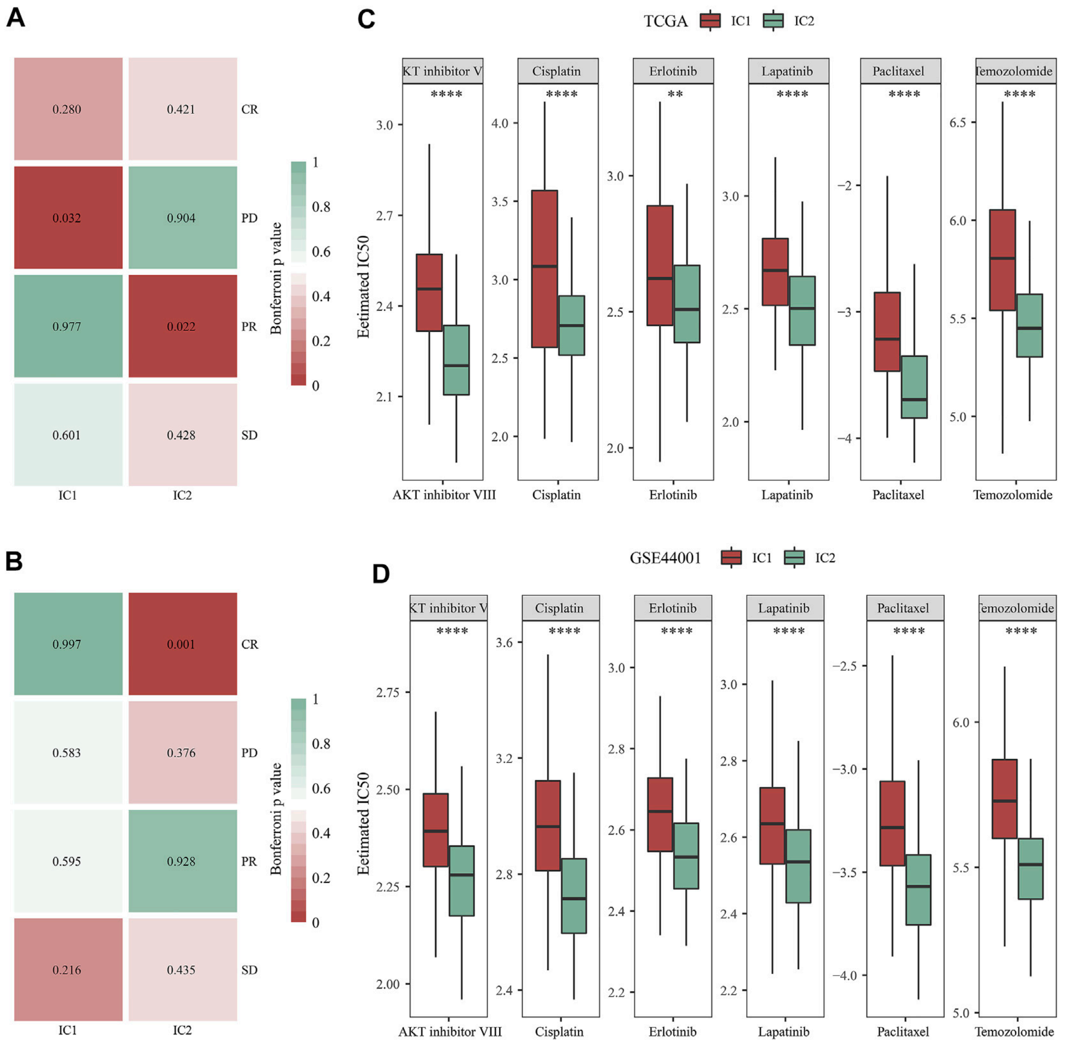


FIGURE 7
Immune subtypes with distinct immuno-, targeted, and chemotherapeutic responses. **(A,B)** Submap for comparing the expression profiling of two immune subtypes with that of IMvigor210 cohort (including four subsets with different responses to anti-PD-L1 immunotherapy in the TCGA and GSE44001 datasets. **(C,D)** Comparison of the IC50 values of targeted, and chemotherapeutic agents (AKT inhibitor VIII, Cisplatin, Erlotinib, Lapatinib, Paclitaxel, and Temozolomide) between two immune subtypes in the TCGA and GSE44001 datasets (**$p < 0.01$; ****$p < 0.0001$).

The immune subtype-relevant recurrence model independently predicts cervical cancer recurrence

Then, we observed that concordance index (C-index) was 0.72, indicating the high prediction accuracy of the recurrence model (Figure 9A). In the TCGA dataset, patients with ≤45 had the relatively higher risk score than those with >45 (Figure 9B). Both in the TCGA and GSE44001 datasets, IC1 exhibited the higher risk score than IC2 (Figures 9B,C). Uni- and multivariate cox regression analysis demonstrated that the model independently predicted cervical cancer recurrence (Figures 9D,E).

Establishment of a nomogram for cervical cancer recurrence

To facilitate clinical application of the immune subtype-relevant recurrence model, we established a nomogram comprising the risk score and other clinicopathological parameters (Figure 9F). As demonstrated by calibration curves, the nomogram-predicted DFS exhibited the relatively high consistency to actual outcomes (Figure 9G). The net benefits of the nomogram were better than other clinicopathological parameters (Figure 9H), indicating the excellent clinical usefulness.

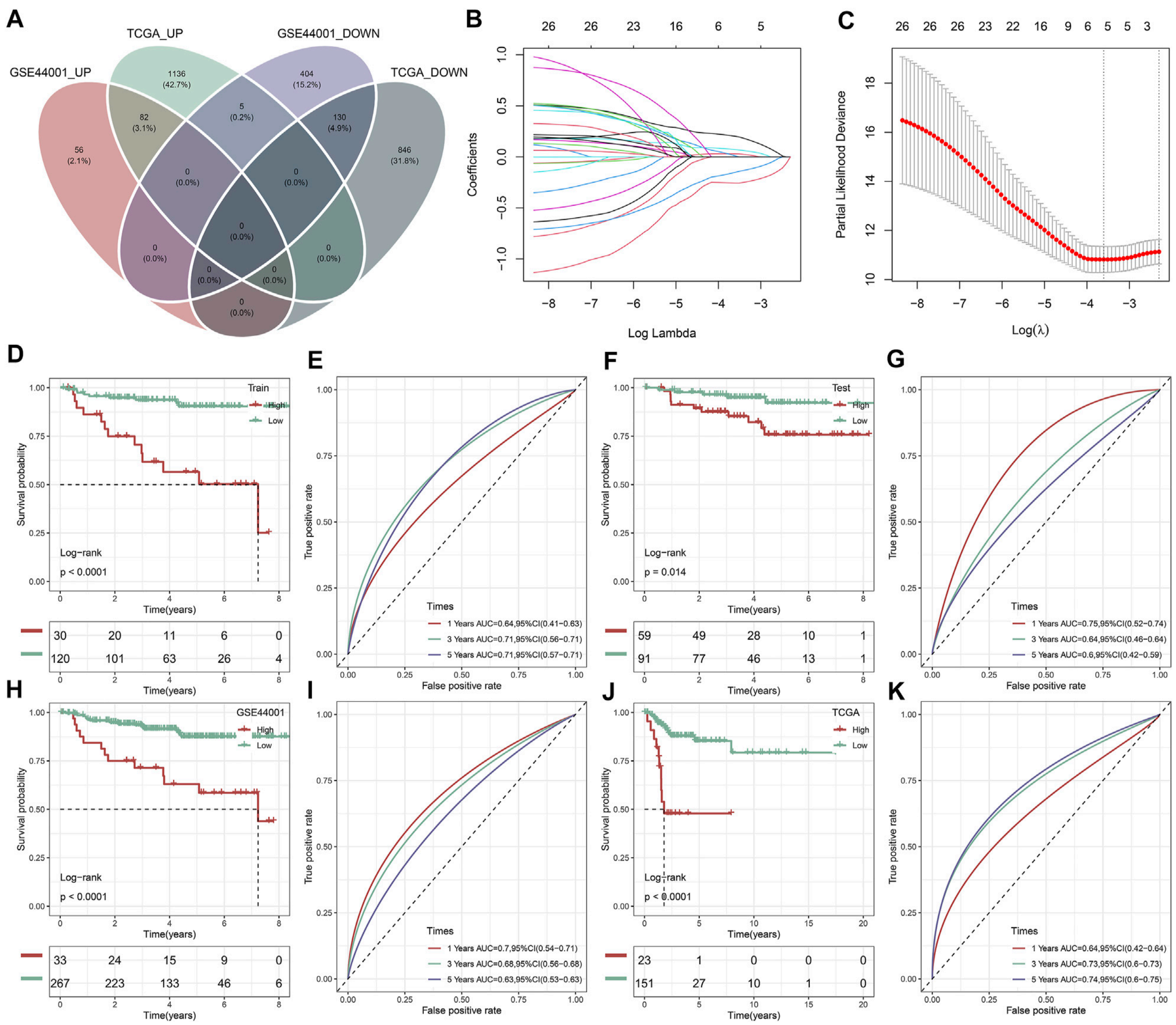


FIGURE 8
Construction and verification of an immune subtype-relevant recurrence model for cervical cancer. **(A)** Venn diagram of shared DEGs between IC1 and IC2 immune subtypes both in the TCGA and GSE44001 datasets. **(B)** Landscape of regression coefficients of 26 recurrence-related DEGs in the training set. **(C)** Selection of the optimal tuning parameter λ *via* 10-fold cross-validation. Kaplan-Meier curves of DFS and ROC curves at 1-, 3-, and 5-year DFS in the **(D,E)** GSE44001 training set, **(F,G)** GSE44001 testing set, **(H,I)** GSE44001 dataset and **(J,K)** TCGA dataset based on the risk score.

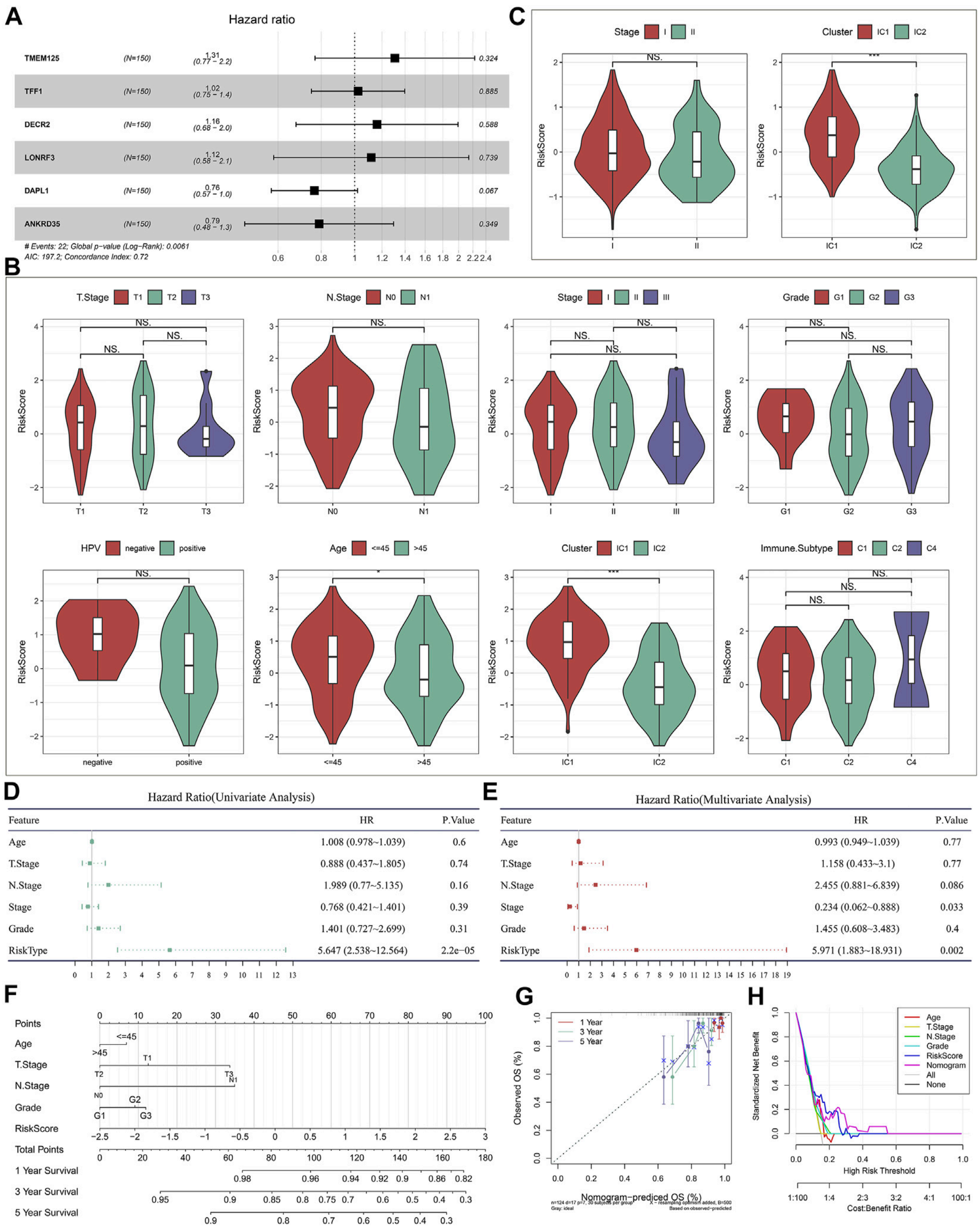
TABLE 1 Multivariate cox regression results in the training set.

Gene	coef	HR	HR (lower, 0.95)	HR (upper, 0.95)	P
TMEM125	0.26809	1.3075	0.7677	2.227	0.3237
TFF1	0.02287	1.0231	0.7495	1.397	0.8855
DECR2	0.14912	1.1608	0.6763	1.992	0.5885
LONRF3	0.11064	1.117	0.5829	2.141	0.7388
DAPL1	−0.26837	0.7646	0.5736	1.019	0.0673
ANKRD35	−0.23696	0.789	0.4803	1.296	0.3494

Abbreviations: coef, coefficient; HR, hazard ratio.

PANoptosis features of immune subtypes and immune subtype-relevant recurrence model

Accumulated evidence demonstrates that pyroptosis, apoptosis and necroptosis (PANoptosis) participate in cancer immunity (Pan et al., 2022). Most of PANoptosis genes were significantly linked to cervical cancer prognosis (Figures 10A,B). In addition, there were notable interactions between PANoptosis genes. Both in TCGA and GSE44001 datasets, most pyroptosis, apoptosis and necroptosis genes displayed notable differential



Hazard Ratio(Univariate Analysis)

Feature	HR	P.Value
Age	1.008 (0.978~1.039)	0.6
T.Stage	0.888 (0.437~1.805)	0.74
N.Stage	1.989 (0.77~5.135)	0.16
Stage	0.768 (0.421~1.401)	0.39
Grade	1.401 (0.727~2.699)	0.31
RiskType	5.647 (2.538~12.564)	2.2e−05

Hazard Ratio(Multivariate Analysis)

Feature	HR	P.Value
Age	0.993 (0.949~1.039)	0.77
T.Stage	1.158 (0.433~3.1)	0.77
N.Stage	2.455 (0.881~6.839)	0.086
Stage	0.234 (0.062~0.888)	0.033
Grade	1.455 (0.608~3.483)	0.4
RiskType	5.971 (1.883~18.931)	0.002

FIGURE 9
Evaluation of the clinical significance of the immune subtype-relevant recurrence model and establishment of a nomogram for cervical cancer. **(A)** Forest diagram of multivariate cox regression analysis of each variable in the recurrence model in the TCGA dataset. **(B)** Distribution of the risk score across distinct clinicopathological parameters in the TCGA dataset. **(C)** Distribution of the risk score across distinct clinicopathological parameters in the GSE44001 dataset. **(D,E)** Uni- and multivariate cox regression analysis for the associations of the risk score and clinicopathological parameters with DFS in the TCGA dataset. **(F)** Establishment of a nomogram including independent variables in the TCGA dataset. **(G)** Calibration curves for assessing the nomogram-predicted and actual survival outcome. **(H)** Decision curve analysis for evaluation of the net benefit.

expression between immune subtypes (Figures 10C,D). In addition, the immune subtype-relevant recurrence model-derived risk score was significantly correlated to PANoptosis genes in two datasets (Figures 10E,F).

Discussion

Cervical cancer remains a major gynecological issue globally (Mittal et al., 2022). Despite major advances in early detection

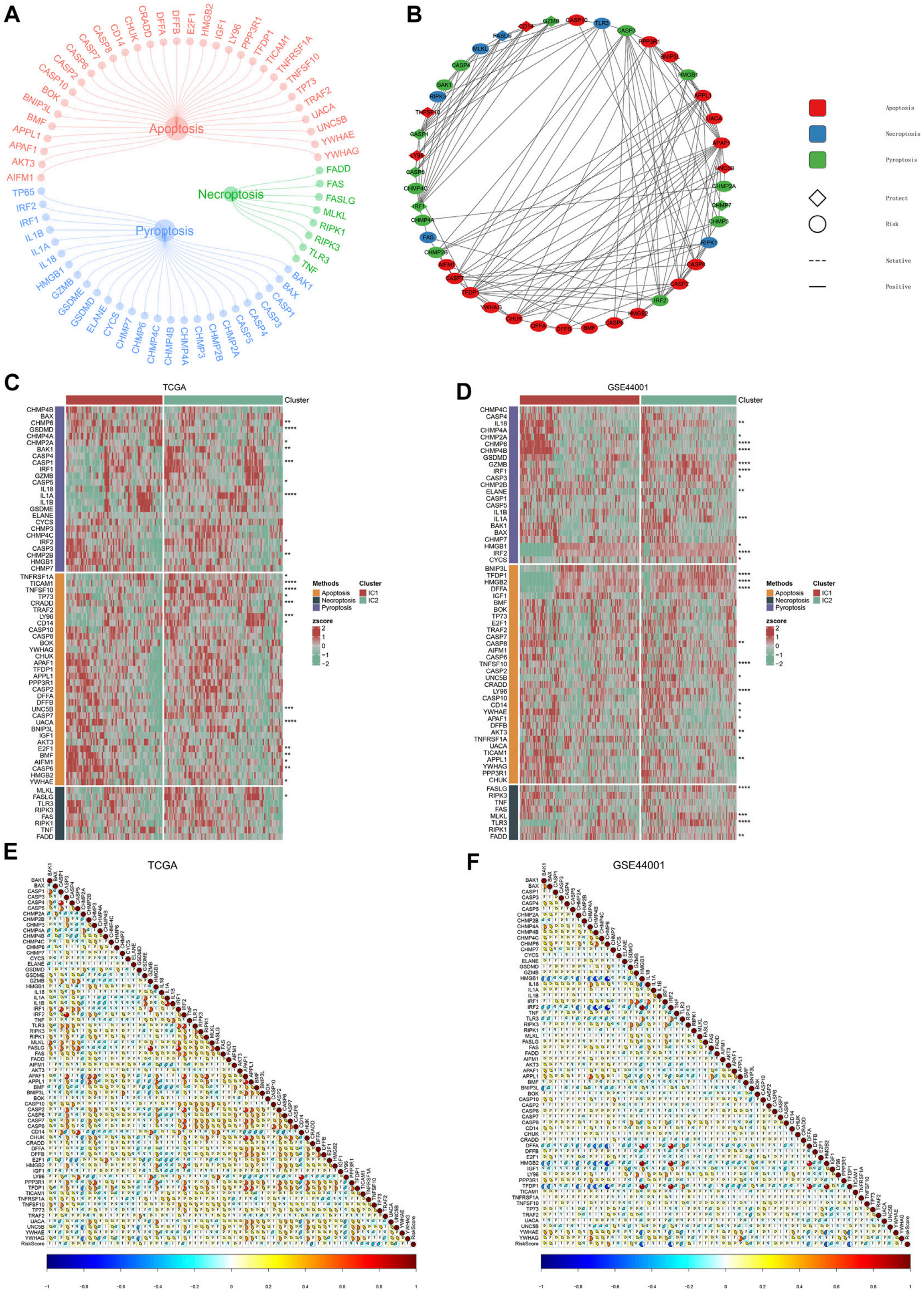


FIGURE 10
PANoptosis features of immune subtypes and immune subtype-relevant recurrence model. **(A)** The network of PANoptosis (pyroptosis, apoptosis, and necroptosis) genes. **(B)** Prognostic significance of PANoptosis genes and their interactions. **(C,D)** Heatmaps of the expression patterns of PANoptosis genes across two subtypes. **(E,F)** Correlations between immune subtype-relevant recurrence model-derived risk score and PANoptosis genes (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$).

and treatment modalities, some patients still relapse. Clinical management of recurrent cervical cancer depends upon treatment history, location as well as degree of recurrence (Zhang et al., 2022). Recurrent cervical cancer usually exhibits challenges for clinicians due to undesirable survival outcomes and limited therapeutic options (Grau et al., 2020). Here, cervical cancer samples were categorized as two immune subtypes with distinct recurrence risk. The novel immune molecular classification was different from existing immune subtypes (Ceccarelli et al., 2016).

Immunoregulators experience immune editing when tumor cells enable to escape immunological surveillance, permit unchecked growth as well as spread (O'Donnell et al., 2019). Also, tumor cells usually apply naturally occurring immunoregulators to escape immune surveillance as well as construct an immunosuppressive microenvironment, whereas lowering anti-tumor activity *via* effector T cells. The results from five algorithms revealed that IC2 exhibited the higher abundance levels of immune cells within the immune microenvironment than IC1. Immune checkpoints can be regulated *via* agonist or antagonist monoclonal antibodies for enhancing T cell activation as well as eliminating suppression of T cell activation, thereby reactivating T cells to attack tumor cells (van der Leun et al., 2020). Recent clinical trials showed that survival outcomes were remarkably longer with cemiplimab, anti-PD-1 inhibitor, compared with single-agent chemotherapy for patients with recurrent cervical cancer following the first-line platinum treatment (Tewari et al., 2022). Moreover, dual PD-1 and CTLA-4 blockage combination displayed durable clinical activity and favorable tolerability as the second-line therapeutic regimen for advanced cervical cancer (O'Malley et al., 2022). However, how to predetermine which patients will respond to immunotherapy remains an issue. Here, we observed that IC2 exhibited higher expression of immune checkpoints (BTLA, CD244, CD274, CD28, CD40, CTLA4, ICOS, PDCD1, PDCD1LG2, and TIGIT) and higher activity of immune-relevant pathways (APC co-inhibition and co-stimulation, inflammation-promoting, MHC class I, IFN response, leukocyte and stromal fractions, macrophage regulation, and TCR Shannon). Evidence indicates that PANoptosis may open an additional avenue for developing promising novel strategies cancer GC immunotherapy. Herein, two immune subtypes exhibited distinct PANoptosis features, and immune subtype-relevant recurrence model-derived risk score correlated to PANoptosis. In accordance with higher TMB, lower TIDE and higher similarity to the expression profiling of patients who well responded to immunotherapy, patients in IC2 were more suitable for immune checkpoint blockade.

Concurrent chemoradiotherapy remains the standard of care for patients with FIGO stage IB 2 or higher (Mittal et al., 2022). Among them, cisplatin is the best-studied and most active single chemotherapeutic drug. Additionally, targeted therapy (anti-angiogenic agent) as well as tyrosine kinase inhibitors have been applied for treating recurrent or metastatic patients. IC2 patients were more likely to respond to targeted, and chemotherapeutic agents (comprising AKT inhibitor VIII, Cisplatin, Erlotinib, Lapatinib, Paclitaxel, and Temozolomide). Currently, CEA, CA125 and SCC remain three major biomarkers of cervical cancer for early screening, treatment monitoring as well as prognostic evaluation (Cao et al., 2022). However, because they exhibit low sensitivity and specificity as expected, novel biomarkers with high reliability, sensitivity and specificity are needed. In the present study, the immune subtype-relevant signature (covering TMEM125, TFF1, DECR2, LONRF3, DAPL1, and ANKRD35) was quantified, which predicted cervical cancer recurrence accurately and independently. Nonetheless, no studies have reported the roles of the genes derived from the signature in cervical cancer. Also, to facilitate clinical practice, we established a nomogram that comprising the immune subtype-relevant signature and known clinicopathological parameters. Despite this, this is a retrospective analysis based on two large cohorts. We will verify above findings in a prospective, and larger cohort in our future research.

Conclusion

Collectively, our findings proposed a novel immune molecular classification for cervical cancer, which classified cervical cancer patients into two immune subtypes with distinct recurrence risk, immune microenvironment, PANoptosis features as well as immuno-, targeted- and chemotherapeutic responses. Altogether, our findings might aid clinicians to make clinical therapeutic regimens for cervical cancer patients and facilitate personalized precision medicine.

Data availability statement

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

Author contributions

CD and JW conceived and designed the study. SQ and JW conducted most of the experiments and data analysis, and wrote the manuscript. FF participated in collecting data and helped to draft the manuscript. All authors reviewed and approved the 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.2022.1007108/full#supplementary-material

SUPPLEMENTARY FIGURE S1

Identification of the optimal number of clusters. **(A)** Consensus CDF. **(B)** Relative change in area under CDF curve.

SUPPLEMENTARY FIGURE S2

Validation of immune subtypes for cervical cancer with different recurrence outcomes in the GSE44001 dataset. **(A)** Consensus matrix heatmap of cervical cancer samples in accordance with the expression values of recurrence-related IRGs when k = 2. **(B)** Kaplan-Meier curves of DFS between IC1 and IC2. **(C)** Landscape of clinicopathological parameters in two immune subtypes ($^{**}p < 0.01$). **(D,E)** Comparison of DFS, and histological stage between immune subtypes.

SUPPLEMENTARY FIGURE S3

Validation of immune subtypes-relevant genes and their biological significance in the GSE44001 dataset. **(A)** The main biological process (BP), molecular function (MF), cellular component (CC) and KEGG enrichment results of genes with down-regulation in IC1. **(B)** The main BP, MF, CC and KEGG enrichment results of genes with up-regulation in IC1. **(C)** GSEA for the main Hallmark and KEGG pathways activated in IC1 or IC2.

SUPPLEMENTARY TABLE S1

The demographics and follow-up data of cervical cancer samples in the TCGA and GSE44001 datasets.

SUPPLEMENTARY TABLE S2

The list of 782 immune-related genes.

SUPPLEMENTARY TABLE S3

The gene sets of PANoptosis (pyroptosis, apoptosis and necroptosis).

SUPPLEMENTARY TABLE S4

The demographics and follow-up information of the IMvigor210 cohort.

SUPPLEMENTARY TABLE S5

The information of DEGs between IC1 and IC2 immune subtypes in the TCGA dataset.

SUPPLEMENTARY TABLE S6

The information of DEGs between IC1 and IC2 immune subtypes in the GSE44001 dataset.

SUPPLEMENTARY TABLE S7

DEGs that were significantly linked with recurrence of cervical cancer.

References

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Cao, X., Yao, J., Jia, M., Shen, X., Zhang, J., and Ju, S. (2022). Serum CCAT2 as a biomarker for adjuvant diagnosis and prognostic prediction of cervical cancer. *J. Ovarian Res.* 15 (1), 20. doi:10.1186/s13048-022-00950-0

Castle, P. E., Einstein, M. H., and Sahasrabuddhe, V. V. (2021). Cervical cancer prevention and control in women living with human immunodeficiency virus. *Ca. Cancer J. Clin.* 71 (6), 505–526. doi:10.3322/caac.21696

Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., Murray, B. A., et al. (2016). Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164 (3), 550–563. doi:10.1016/j.cell.2015.12.028

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

Cibula, D., Dostálek, L., Jarkovsky, J., Mom, C. H., Lopez, A., Falconer, H., et al. (2022). Post-recurrence survival in patients with cervical cancer. *Gynecol. Oncol.* 164 (2), 362–369. doi:10.1016/j.ygyno.2021.12.018

Colombo, N., Dubot, C., Lorusso, D., Caceres, M. V., Hasegawa, K., Shapira-Frommer, R., et al. (2021). Pembrolizumab for persistent, recurrent, or metastatic cervical cancer. *N. Engl. J. Med.* 385 (20), 1856–1867. doi:10.1056/NEJMoa2112435

Danilova, L., Ho, W. J., Zhu, Q., Vithayathil, T., De Jesus-Acosta, A., Azad, N. S., et al. (2019). Programmed cell death ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival. *Cancer Immunol. Res.* 7 (6), 886–895. doi:10.1158/2326-6066.Cir-18-0822

Engebretsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. *Clin. Epigenetics* 11 (1), 123. doi:10.1186/s13148-019-0730-1

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

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

Grau, J. F., Farinas-Madrid, L., and Oaknin, A. (2020). A randomized phase III trial of platinum chemotherapy plus paclitaxel with bevacizumab and atezolizumab versus platinum chemotherapy plus paclitaxel and bevacizumab in metastatic (stage IVB), persistent, or recurrent carcinoma of the cervix: The BEATcc study (ENGOT-Cx10/GEICO 68-C/JGOG1084/GOG-3030). *Int. J. Gynecol. Cancer* 30 (1), 139–143. doi:10.1136/ijgc-2019-000880

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

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

Karki, R., Sharma, B. R., Lee, E., Banoth, B., Malireddi, R. K. S., Samir, P., et al. (2020). Interferon regulatory factor 1 regulates PANoptosis to prevent colorectal cancer. *JCI Insight* 5 (12), 136720. doi:10.1172/jci.insight.136720

Lee, S., Karki, R., Wang, Y., Nguyen, L. N., Kalathur, R. C., and Kanneganti, T. D. (2021). AIM2 forms a complex with pyrin and ZBP1 to drive PANoptosis and host defence. *Nature* 597 (7876), 415–419. doi:10.1038/s41586-021-03875-8

Lee, Y. Y., Kim, T. J., Kim, J. Y., Choi, C. H., Do, I. G., Song, S. Y., et al. (2013). Genetic profiling to predict recurrence of early cervical cancer. *Gynecol. Oncol.* 131 (3), 650–654. doi:10.1016/j.ygyno.2013.10.003

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

Li, W., Wang, H., Ma, Z., Zhang, J., Ou-Yang, W., Qi, Y., et al. (2019). Multi-omics analysis of microenvironment characteristics and immune escape mechanisms of hepatocellular carcinoma. *Front. Oncol.* 9, 1019. doi:10.3389/fonc.2019.01019

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

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

Miccò, M., Lupinelli, M., Mangialardi, M., Gui, B., and Manfredi, R. (2022). Patterns of recurrent disease in cervical cancer. *J. Pers. Med.* 12 (5), 755. doi:10.3390/jpm12050755

Mittal, P., Chopra, S., Charnalia, M., Dora, T., Engineer, R., Mulani, J., et al. (2022). Patterns of relapse after adjuvant chemoradiation for cervical cancer in a phase 3 clinical trial (PARCER): An evaluation of updated NRG oncology/RTOG target delineation guidelines. *Int. J. Radiat. Oncol. Biol. Phys.* 113 (2), 369–378. doi:10.1016/j.ijrobp.2022.02.007

O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* 16 (3), 151–167. doi:10.1038/s41571-018-0142-8

O'Malley, D. M., Neffa, M., Monk, B. J., Melkadze, T., Huang, M., Kryzhanivska, A., et al. (2022). Dual PD-1 and CTLA-4 checkpoint blockade using balstilimab and zalifrelimab combination as second-line treatment for advanced cervical cancer: An open-label phase II study. *J. Clin. Oncol.* 40 (7), 762–771. doi:10.1200/jco.21.02067

Pan, H., Pan, J., Li, P., and Gao, J. (2022). Characterization of PANoptosis patterns predicts survival and immunotherapy response in gastric cancer. *Clin. Immunol.* 238, 109019. doi:10.1016/j.clim.2022.109019

Qiu, X. T., Song, Y. C., Liu, J., Wang, Z. M., Niu, X., and He, J. (2020). Identification of an immune-related gene-based signature to predict prognosis of patients with gastric cancer. *World J. Gastrointest. Oncol.* 12 (8), 857–876. doi:10.4251/wjgo.v12.i8.857

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

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

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

Tewari, K. S., Monk, B. J., Vergote, I., Miller, A., de Melo, A. C., Kim, H. S., et al. (2022). Survival with cemiplimab in recurrent cervical cancer. *N. Engl. J. Med.* 386 (6), 544–555. doi:10.1056/NEJMoa2112187

van der Leun, A. M., Thommen, D. S., and Schumacher, T. N. (2020). CD8(+) T cell states in human cancer: Insights from single-cell analysis. *Nat. Rev. Cancer* 20 (4), 218–232. doi:10.1038/s41568-019-0235-4

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

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

Xu, F., Shen, J., and Xu, S. (2021). Integrated bioinformatical analysis identifies GIMAP4 as an immune-related prognostic biomarker associated with remodeling in cervical cancer tumor microenvironment. *Front. Cell Dev. Biol.* 9, 637400. doi:10.3389/fcell.2021.637400

Yang, C., Huang, X., Liu, Z., Qin, W., and Wang, C. (2020). Metabolism-associated molecular classification of hepatocellular carcinoma. *Mol. Oncol.* 14 (4), 896–913. doi:10.1002/1878-0261.12639

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

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

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

Zhang, Y. F., Fan, Y., Zhang, P., Ruan, J. Y., Mu, Y., and Li, J. K. (2022). Cervical cancer recurrence and patient survival after radical hysterectomy followed by either adjuvant chemotherapy or adjuvant radiotherapy with optional concurrent chemotherapy: A systematic review and meta-analysis. *Front. Oncol.* 12, 823064. doi:10.3389/fonc.2022.823064

Zheng, M., Hu, Y., Gou, R., Liu, O., Nie, X., Li, X., et al. (2020). Identification of immune-enhanced molecular subtype associated with BRCA1 mutations, immune checkpoints and clinical outcome in ovarian carcinoma. *J. Cell. Mol. Med.* 24 (5), 2819–2831. doi:10.1111/jcmm.14830

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 14 October 2022
DOI 10.3389/fgene.2022.1016871

Identification of cuproptosis-related gene signature to predict prognosis in lung adenocarcinoma

Yanju Lv[1†], Yajie Xiao[2†], Xiaoli Cui[2], Haitao Luo[2*] and Long Xu[3*]

[1]Department of Internal Medicine, Second Affiliated College of Harbin Medical University, Harbin, China, [2]Department of Medicine, YuceBio Technology Co., Ltd., Shenzhen, China, [3]Department of Oncology, General Hospital of Northern Theater Command, Shenyang, China

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Ziheng Wang,
University of Macau, China
Yaozeng Xu,
The First Affiliated Hospital of Soochow University, China

*CORRESPONDENCE
Haitao Luo,
luohaitao@yucebio.com
Long Xu,
doctorxul@163.com

†These authors have contributed equally to this work

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

RECEIVED 11 August 2022
ACCEPTED 28 September 2022
PUBLISHED 14 October 2022

CITATION
Lv Y, Xiao Y, Cui X, Luo H and Xu L (2022), Identification of cuproptosis-related gene signature to predict prognosis in lung adenocarcinoma.
Front. Genet. 13:1016871.
doi: 10.3389/fgene.2022.1016871

Background: Studies have reported that coppers are involved in the tumorigenesis and development of tumor. In herein, we aimed to construct a prognostic classification system for lung adenocarcinoma (LUAD) associated with cuproptosis.

Methods: Samples information of LUAD were acquired from The Cancer Genome Atlas (TCGA) and GSE31210 dataset. Cuproptosis-related genes were screened from previous research. ConsensusClusterPlus was applied to determine molecular subtypes, which evaluated by genome analysis, tumor immune microenvironment analysis, immunotherapy, functional enrichment analysis. Furthermore, univariate Cox analysis combined with Lasso analysis were employed to construct a cuproptosis-related risk model for LUAD.

Results: 14 genes related to cuproptosis phenotype were identified, and 2 clusters (C1 and C2) were determined. Among which, C1 had better survival outcome, less advanced stages, enhanced immune infiltration and enriched in TCA related pathways. A 7 cuproptosis-associated genes risk model was constructed, and the performance was verified in the GSE31210 dataset. A higher RiskScore was significantly correlated with worse overall survival, advanced stages. Cox survival analysis showed that RiskScore was an independent predictor. High-risk group patients had weakened immune infiltration, less likely to benefit from immunotherapy and was more sensitived to immunotherapy.

Conclusion: The cuproptosis-related gene signature could serve as potential prognostic predictors for LUAD patients and may provide clues for the intervention of cuproptosis induced harm and targeted anti-tumor application.

KEYWORDS

lung adenocarcinoma, cuproptosis, molecular subtype, RiskScore, prognosis

Introduction

Lung adenocarcinoma (LUAD), the most common type of non-small cell lung cancer, is characterized by dense lymphocytic infiltration and early metastasis (Luo et al., 2020). Although treatment strategies for LUAD have improved greatly in recent years, the survival rate of patients with LUAD is still very low (Kleczko et al., 2019). Chemotherapy, surgical resection and radiotherapy are routine treatments for LUAD, However, due to the lack of specificity of these treatments, they can also cause damage to adjacent normal cells (Wang et al., 2021). Targeted therapy and immunotherapy are one of the main methods for the treatment of LUAD. Although both have achieved good clinical efficacy (Osmani et al., 2018; Xing et al., 2019), the clinical benefit population is still limited (Park and Jang, 2016; Testa et al., 2018). Therefore, it is of great significance to further search for new diagnostic markers and therapeutic targets for LUAD.

It is well known that copper, as a cofactor of essential enzymes, plays an important role in human life (Kim et al., 2008). The concentration of copper in normal cells is very low, which mainly prevents the harmful accumulation of free intracellular copper through the homeostasis mechanism across the concentration gradient, thus maintaining cellular copper homeostasis (Lutsenko, 2010; Ge et al., 2022). However, a recent study shows that copper death is dependent on mitochondrial respiration, which is different from the previously known programmed cell death (such as iron death and apoptosis). In this process, copper directly binds to the lipid components of the tricarboxylic acid cycle, resulting in the aggregation of lipoacylated proteins and the loss of iron-sulfur cluster proteins, leading to proteotoxic stress and cell death (Tsvetkov et al., 2022). The importance of copper homeostasis in immune infiltration has also been demonstrated in some recent correlation studies (Choi et al., 2013; Tan et al., 2021). Tan et al. (2021) found that copper chelation on macrophages can eliminate lysyl oxidase-like 4-mediated programmed death molecule ligand 1 presentation, thereby inhibiting cell immune escape. Choi et al. (2013) showed that chlorodoxyquine (a common copper chelator) can effectively reduce the infiltration of encephalitis-causing immune cells (CD4, CD8, etc.).

Based on this, this study is the first to investigate the prognosis of LUAD by combining LUAD microarray data and cuproptosis-related genes. Using the GSE31210 dataset of TCGA database and GEO database, combined with the 13 copper-death genes provided by Tsvetkov et al. (2022), we identified important cuproptosis related genes and molecular subtypes, and constructed a risk model. Finally, based on the subtypes and risk models obtained above, functional enrichment analysis, immune infiltration analysis, immunotherapy and chemotherapy drug prediction were performed, so as to provide some theoretical support for the mechanism research of cuproptosis in LUAD.

Materials and methods

Data source

In order to obtain the microarray data related to lung adenocarcinoma, TCGA and GEO databases were searched with “LUAD” as the search term. The TCGA-LUAD dataset contains 472 cancer patient samples and 59 healthy samples, and the GSE31210 dataset contains 226 LUAD samples. 472 tumor samples were classified into the training cohort randomly (n = 236), the testing cohort (n = 236). The two groups were similar in age, gender, Stage, follow-up time, and Event (Table 1).

The 13 cuproptosis related genes, SLC31A1, PDHB, PDHA1, LIPT1, FDX1, DLD, DLST, DBT, LIAS, DLAT, GCSH, ATP7A, and ATP7B, were derived from a recent report by team Tsvetkov et al. (2022).

Differentially expressed genes analysis

Based on 13 cuprotosis related genes, scores of cuprotosis related genes in each sample were calculated by single sample gene set enrichment analysis (ssGSEA), and DEGs were screened between cancer tissues and para-carcinoma tissue with FDR<0.05 and |log2FC|>2.

Then, the correlation analysis between DEGs and scores were analyzed by pearson methods with selection criteria |R|>0.2 and p.value < 0.05 to obtained genes associated with cuproptosis phenotype.

Univariate COX survival analysis

Next, Univariate COX survival analysis using coxph function of R package was used to analysis genes associated with cuproptosis phenotype with $p < 0.05$ to determine cuproptosis-related genes for LUAD prognosis, for subsequent analysis

Cluster analysis

Base on cuproptosis-related genes, Then, molecular subtypes were performed separately for TCGA-LUAD dataset samples *via* the Consensus Cluster Plus 1.52.0 (Wilkerson and Hayes, 2010). “pam” arithmetic and “pearson” distance were utilized to complete 500 bootstraps with every bootstrap having specimens (≥80%) of

TABLE 1 Sample information of TCGA training dataset and validation dataset.

Characteristics	Train (N = 236)	Test (N = 236)	Total (N = 472)	*p* value	FDR
Gender				1	1
FEMALE	127 (26.91%)	128 (27.12%)	255 (54.03%)		
MALE	109 (23.09%)	108 (22.88%)	217 (45.97%)		
T.stage				0.79	1
T1	76 (16.10%)	84 (17.80%)	160 (33.90%)		
T2	130 (27.54%)	123 (26.06%)	253 (53.60%)		
T3	22 (4.66%)	21 (4.45%)	43 (9.11%)		
T4	8 (1.69%)	7 (1.48%)	15 (3.18%)		
Ukown	0 (0.0e+0%)	1 (0.21%)	1 (0.21%)		
N.stage				0.2	1
N0	160 (33.90%)	153 (32.42%)	313 (66.31%)		
N1	40 (8.47%)	46 (9.75%)	86 (18.22%)		
N2	34 (7.20%)	28 (5.93%)	62 (13.14%)		
N3	0 (0.0e+0%)	2 (0.42%)	2 (0.42%)		
Ukown	2 (0.42%)	7 (1.48%)	9 (1.91%)		
M.stage				0.88	1
M0	159 (33.69%)	161 (34.11%)	320 (67.80%)		
M1	9 (1.91%)	7 (1.48%)	16 (3.39%)		
Ukown	68 (14.41%)	68 (14.41%)	136 (28.81%)		
Stage				0.83	1
I	135 (28.60%)	129 (27.33%)	264 (55.93%)		
II	53 (11.23%)	61 (12.92%)	114 (24.15%)		
III	36 (7.63%)	34 (7.20%)	70 (14.83%)		
IV	9 (1.91%)	7 (1.48%)	16 (3.39%)		
Ukown	3 (0.64%)	5 (1.06%)	8 (1.69%)		
Event				0.57	1
Alive	141 (29.87%)	148 (31.36%)	289 (61.23%)		
Dead	95 (20.13%)	88 (18.64%)	183 (38.77%)		
Age				0.13	0.92
<=65	69 (14.62%)	77 (16.31%)	146 (30.93%)		
>65	159 (33.69%)	157 (33.26%)	316 (66.95%)		
Ukown	8 (1.69%)	2 (0.42%)	10 (2.12%)		

TCGA-LUAD dataset. Cluster number k was between 2 and 10, and the optimum k was identified as per cumulative distribution function (CDF) and AUC. Survival curves (KM curves) between molecular subtypes were then analyzed for difference. In addition, differences in the distribution of clinical characteristics between molecular subtypes were compared and a chi-square test was completed, and $p < 0.05$ had significance on statistics.

Single-sample GSEA

The ssGSEA was used to evaluate the various pathways scores (Charoentong et al., 2017) using GSVA of R package. NES>0 indicates pathway activation, and NES<0 indicates pathway inhibition.

Estimation of STromal and immune cells in MAlignant tumours using expression data

R software ESTIMATE arithmetic (Yang et al., 2021) was utilized to compute overall stroma level (Stromal Score), the immunocyte infiltration (Immune Score) and the combination (ESTIMATE Score) of sufferers in the TCGA-LUAD cohort using Wilcox.test analysis to determine difference.

Cell-type identification by estimating relative subsets of RNA transcripts

CIBERSORT analyses were utilized to compare diversities in different immunocytes in molecular subtypes. Wilcox.test

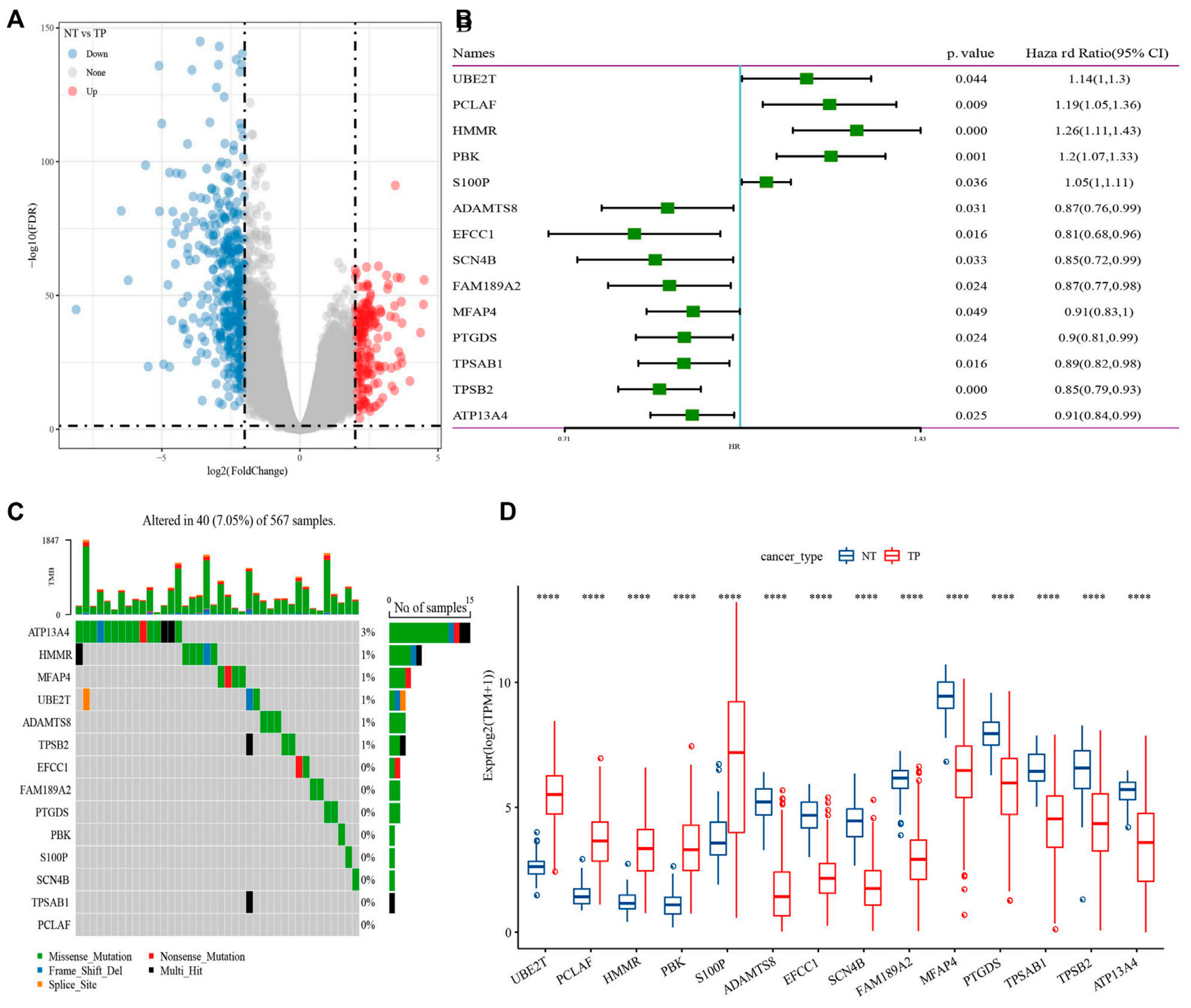


FIGURE 1
Identification of genes closely related to cuproptosis related gene pathway score. **(A)** Differentially expressed gene between cancer tissue and para-carcinoma tissue. **(B)** 14 genes closely related to cuproptosis related gene pathway score. **(C)** mutation analysis of genes in TCGA-LUAD dataset. **(D)** The expression levels of 14 genes in cancer tissue and para-carcinoma tissue. ***$p < 0.0001$.

analyses were completed to identify the difference of 22 kinds of infiltrating immunocyte score between molecular subtypes. The “ggplot2” package (Ito and Murphy, 2013) was used to realize the visualization of the distributional status of the diversities in 22 kinds of infiltration immunocytes.

Immunotherapy

The expression levels of 47 immune checkpoint genes, which from HigsAtlas (Liu et al., 2017), were determined.

Construction and evaluation a prognostic risk model for lung adenocarcinoma

Lasso-cox regression was performed using the Glnmet package in R language to select the best prognostic genes (Tibshirani, 1997). Glmnet is a software package for fitting generalized linear and similarity models by penalized maximum likelihood. The regularization path is the calculation of the lasso or elastic net penalty on the value (on a logarithmic scale) of the regularization parameter lambda (Goeman, 2010). The optimal value of the penalty

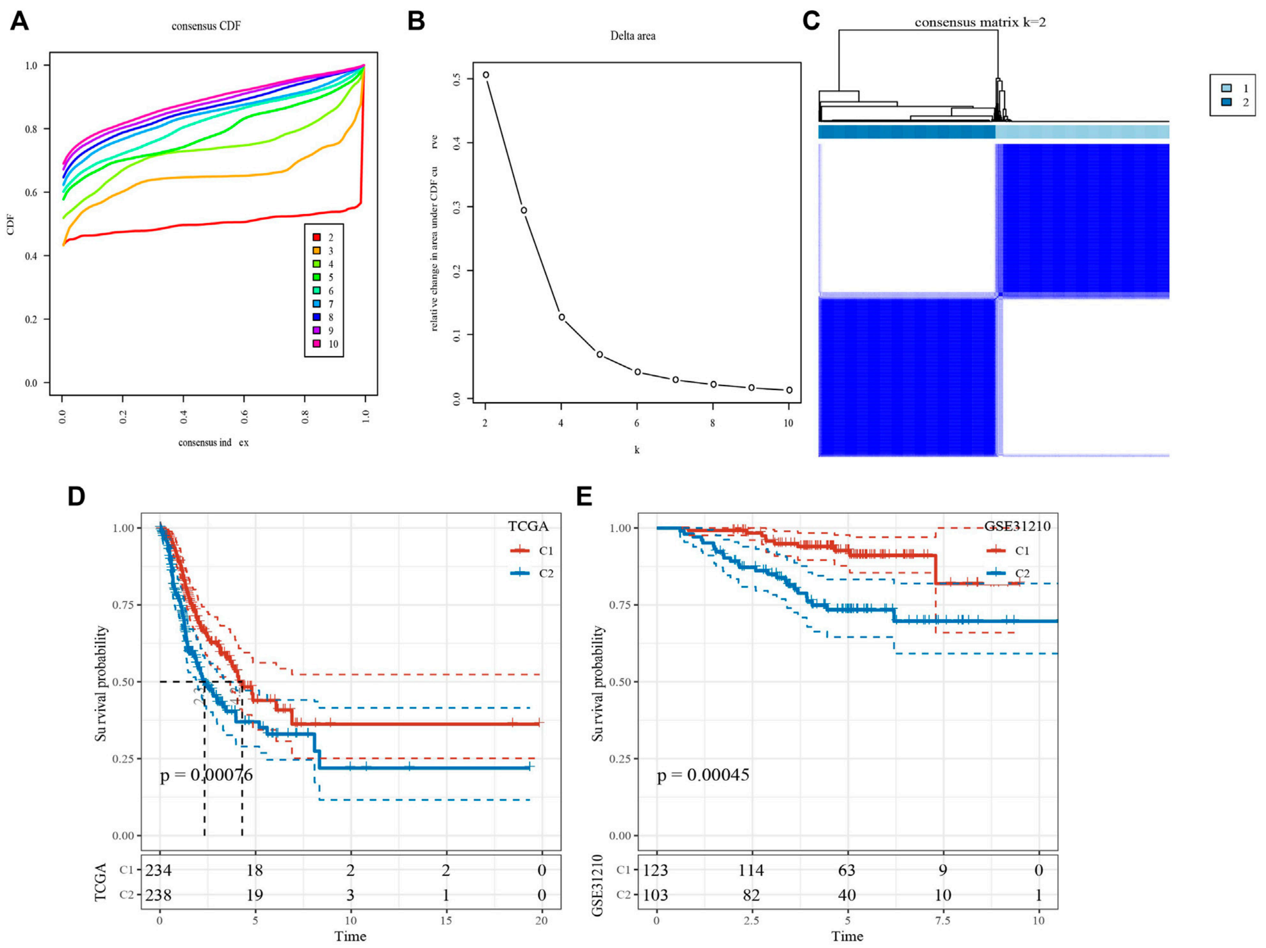


FIGURE 2
Identification of molecular subtypes. **(A)** Cumulative distribution function. **(B)** Delta area. **(C)** Heatmap of sample clustering when k = 2. **(D)** KM survival analysis of C1 and C2 in TCGA-LUAD dataset. **(E)** KM survival analysis of C1 and C2 in GSE31210 dataset.

coefficient λ and the genes to be included in the model were selected by running the 10-fold cross-validation probability 1000 times. Subsequently, Cox multivariate regression analysis coefficients of prognostic genes were extracted, and the gene expression levels were used to calculate the risk score by the following formula as the survival risk score of each patient:

$$RiskScore = \sum\nolimits_{k=0}^{n} \beta i \times Expi$$

Where, βi represents the Cox hazard ratio coefficient of mRNA, and Expri represents the gene expression level. TCGA-LUAD samples were divided into high risk (RiskScore>0) and low risk groups (RiskScore<0) according to the risk score, which was for zscore. At the same time, GSE31210 were used to evaluate the effectiveness and robustness of the prognostic risk model. Kaplan-Meier (KM) curves combined with the Logrank test were used to analyze survival differences among different risk groups. The timeROC package was used to determine the area under the receiver operating characteristic curve (AUC) to predict 1-year, 2-year, 3-year, 4-year and 5-year survival rates, respectively.

Independent prognostic power of RiskScore

Univariate and multivariate COX regression were used to examine the independent prognostic power of RiskScore.

Tumor immune dysfunction and exclusion

TIDE (Jiang et al., 2018; Fu et al., 2020) algorithm (http://tide.dfci.harvard.edu) was used to evaluate three cell types that

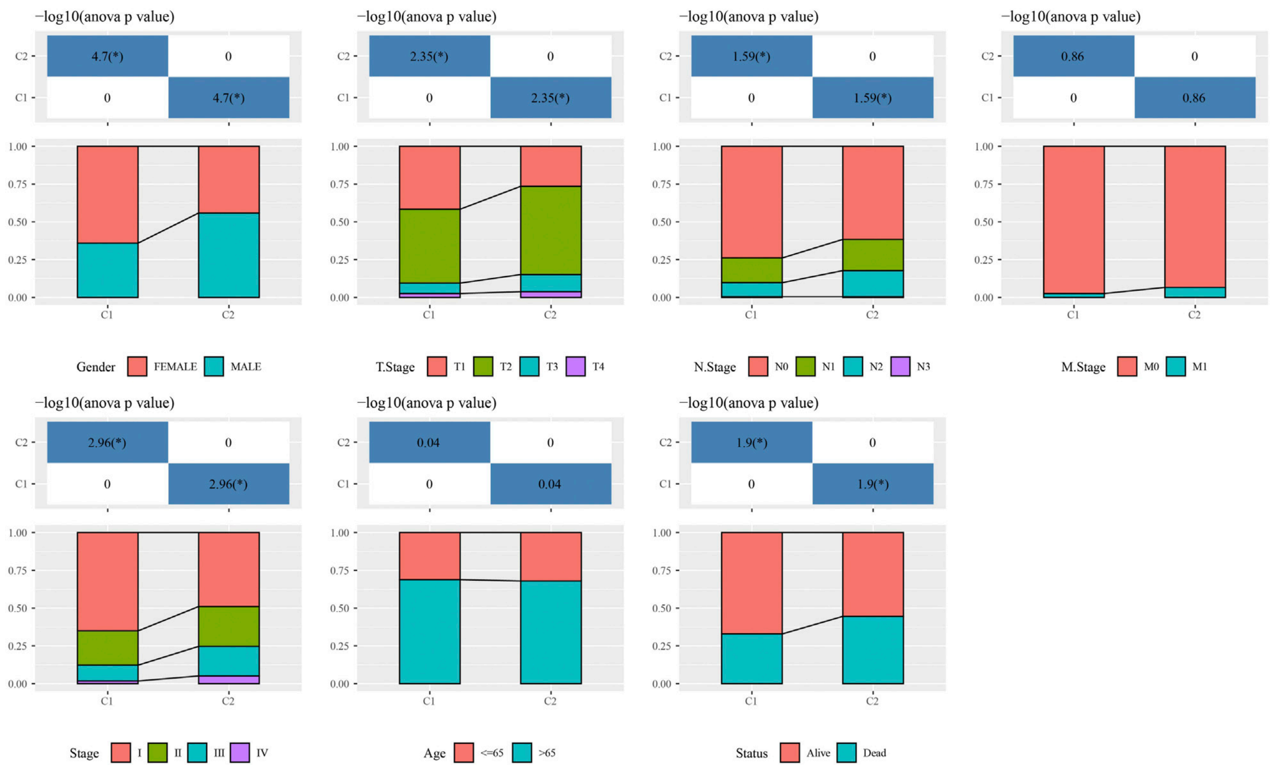


FIGURE 3
The distribution of clinical features, included Gender, T Stage, N Stage, M Stage, Stage Ages, and Status in C1 and C2. $*p < 0.05$.

limit T-cell invasion into tumors, including IFNG, myeloid suppressor cells (MDSC), and M2 subtypes of tumor-associated macrophages (TAM.M2), as well as dysfunction of tumor infiltration cytotoxic T lymphocytes (CTL) and exclusion of CTL by immunosuppressive factors.

Chemotherapy drugs sensitivity analysis

pRRophetic (Geeleher et al., 2014) was used to predict the sensitivity of Cisplatin, Salubrinal, Vinorelbine, QS11, AKT inhibitor VIII and Embelin to IC50.

Sangerbox provided assistance with this article (Shen et al., 2022).

Results

Identification of genes closely related to cuproptosis related gene pathway score

602 DEGs were screened between cancer tissue and para-carcinoma tissue in TCGA-LUAD dataset (Figure 1A), from which, 138 genes were closely with cuproptosis related gene pathway score. Next, Univariate Cox regression analysis identified 14 genes associated with prognosis in lung adenocarcinoma (Figure 1B). 40 of 567 samples (7.05%) in TCGA-LUAD had genes mutation (Figure 1C). The expression levels of 14 genes had significance between cancer tissue and para-carcinoma tissue (Figure 1D). Those data showed that cuproptosis was associated with LUDA.

Identification of molecular subtypes

Based on 14 genes, samples in TCGA-LUAD dataset were clustered with CDF and delta area (Figures 2A,B). When k = 2, 2 clusters (C1 and C2) were found (Figure 2C). KM survival analysis indicated that patients in C1 had better survival outcome in TCGA-LUAD dataset (p = 0.00076, Figure 2D) and GSE32210 dataset (p = 0.00045, Figure 2E). Distribution of clinical features in clusters showed that samples in C2 had more Male, T3/4 stage, N1/2 stage, StageIII/IV and Dead patients (Figure 3). Those analysis indicated that the 2 clusters had clinical significance.

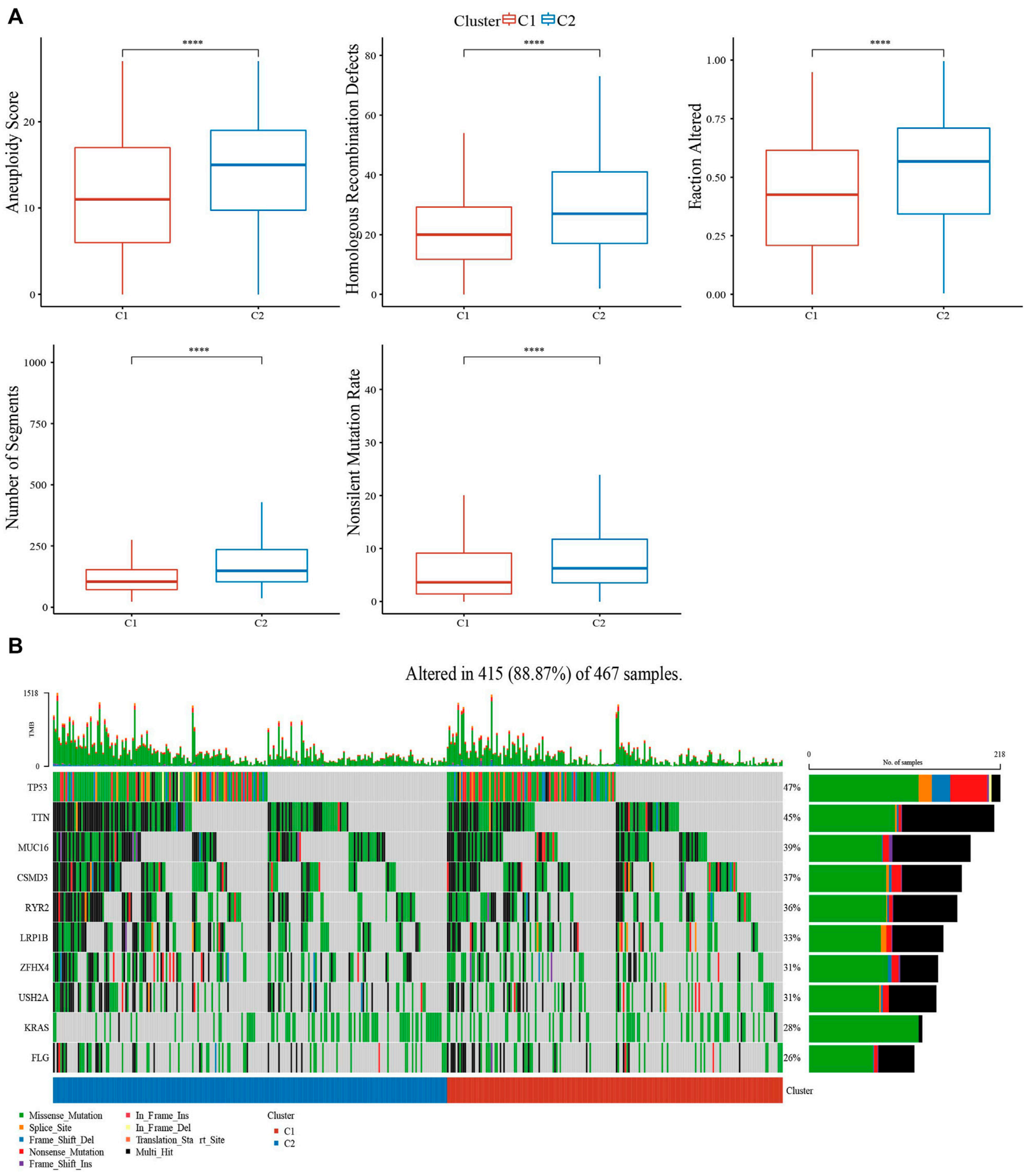


FIGURE 4
Genome analysis. **(A)** the analysis of Aneuploidy Score, Homologous Recombination Defects, Fraction Altered, number of segments, and non-silent mutation rate in C1 and C2. **(B)** Top 10 mutation genes in C1 and C2. ***$p < 0.0001$.

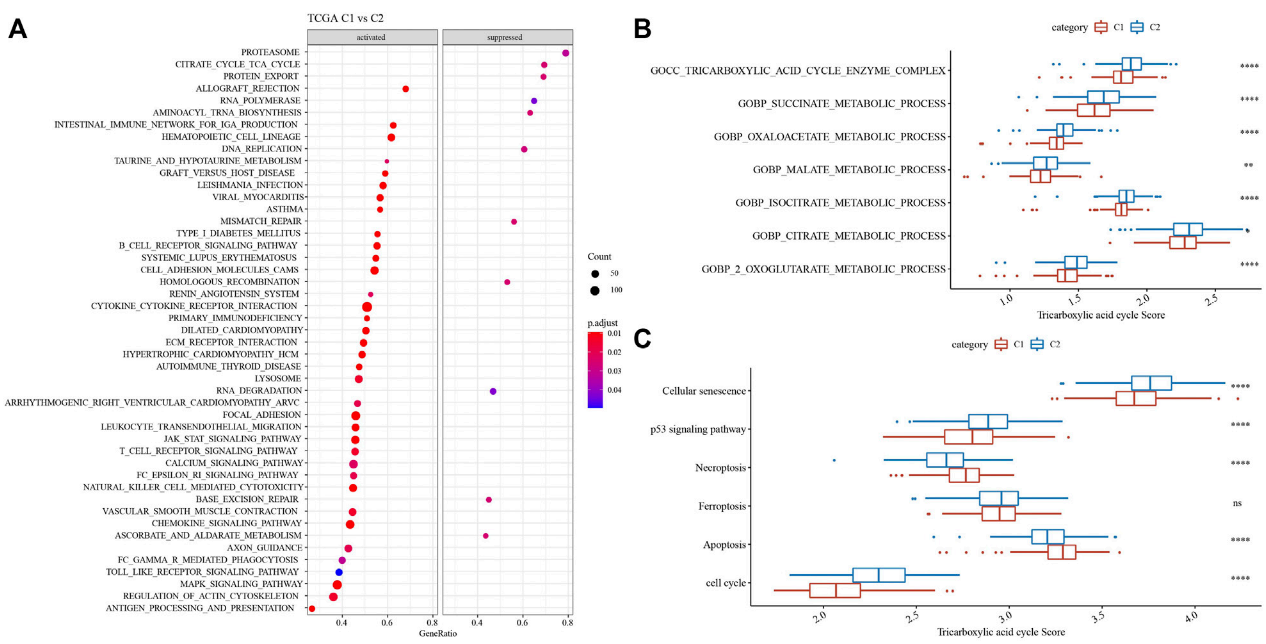


FIGURE 5

Functional enrichment analysis. **(A)** GSEA analysis demonstrated that pathways, such as, cell cycle were activated in C2. **(B)** 7 TCA pathways were activated in C2. **(C)** 6 pathways associated with tumorigenesis had differences in C1 and C2. *$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$, ns: no significance.

High gene mutation was observed in subtypes

Genome analysis between clusters showed that C1 patients presented less Aneuploidy Score, Homologous Recombination Defects, Fraction Altered, Number of Segments, Nonsilent Mutation Rate (Figure 4A). In addition, top10 genes, especially TP53, TTN, MUC16, had obviously mutation differences between C1 and C2 (Figure 4B).

Moreover, GSEA analysis showed that such as CITRATE_CYCLE_TCA_CYCLE and AMINOACYL_TRNA_BIOSYNTHESIS were activated in C2, while, TAURINE_AND_HYPOTAURINE_METABOLISM were activated in C1 (Figure 5A). Tricarboxylic acid cycle related pathways and genes were acquired to calculated TCA pathways scores using ssGSEA, and the results showed that 7 pathways were higher enriched in C2 (Figure 5B). Cell growth and death pathways, and genes were obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp/kegg/pathway.html), ssGSEA analysis indicated that cellular senescence, p53 signaling pathway and cell cycle were higher in C2, while Necroptosis and Apoptosis were activated in C1 (Figure 5C).

C1 had higher immune infiltration

16 of 22 immune cells had significantly difference using CIBERSORT analysis between 2 clusters (Figure 6A). Then, ESTIMATE analysis showed that C1 had higher score of StromalScore, ImmuneScore and ESTIMATEScore (Figure 6B). Our team afterwards evaluated the 47 immune check genes expressions, and 41 immune checkpoint genes had obviously high expressions in C1 that those in C2 (Figure 6C). Next, the scores of CYT, T cell receptor signaling pathway and B cell receptor signaling pathway, were calculated using ssGSEA, and they all were higher in C1 that those in C2 (Figures 6D–F).

Identification of hub genes and RiskScore

1687 DEGs, including 1462 upregulated genes and 422 downregulated genes, were identified in C1 vs. C2 (Figure 7A). TCGA-LUAD dataset was divided into TCGA-training dataset and TCGA-test dataset. In TCGA- training dataset, univariate Cox survival analysis determined 14 genes associated with prognosis, included 12 risk genes and 2 protective genes (Figures 7B,C). LASSO Cox regression module was conducted to build a prognostic signature based on the expression matrix of the 14 genes. Consequently, we identified a 7-genes signature module according to the optimal λ value (Figures 7D,E). RiskScore of LUAD patients base on 7 genes was calculated using the following formula: RiskScore = 0.168*ARHGEF39-0.079*EFCC1-0.124*SERPIND1+0.065*INSL4+0.11*ANLN+0.04*RHOV+0.17*CCL20.

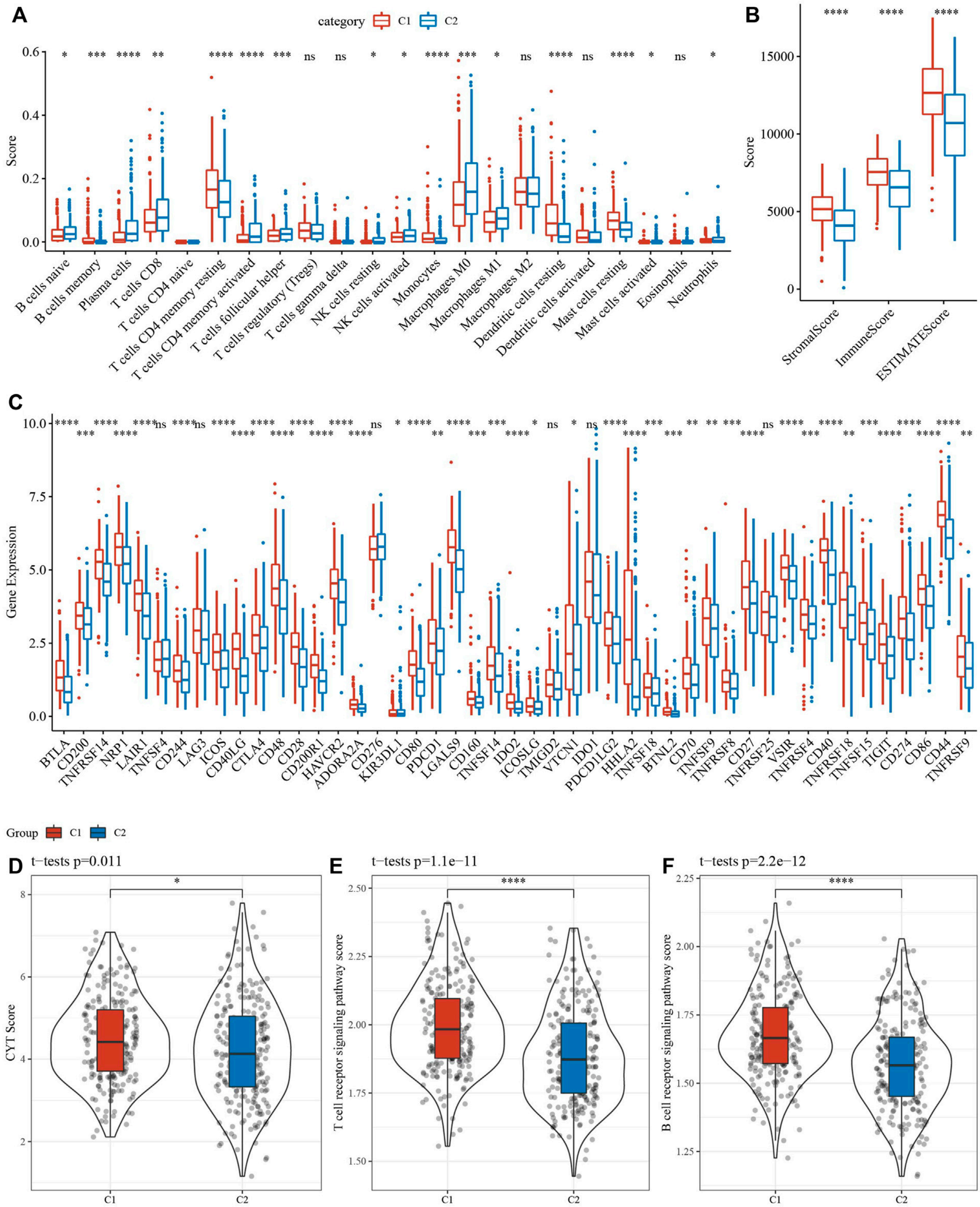


FIGURE 6

Analysis of immune infiltration. **(A)** analysis of 22 immune cells using CIBERSORT. **(B)** Analysis of immune infiltration using ESTIMATE. **(C)** The expression levels of 42 immune check genes between C1 and C2. **(D–F)** The differences of CTY score, T cell receptor signaling pathway score, B cell receptor signaling pathway score between C1 and C2. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ***$p < 0.0001$, ns: no significance.

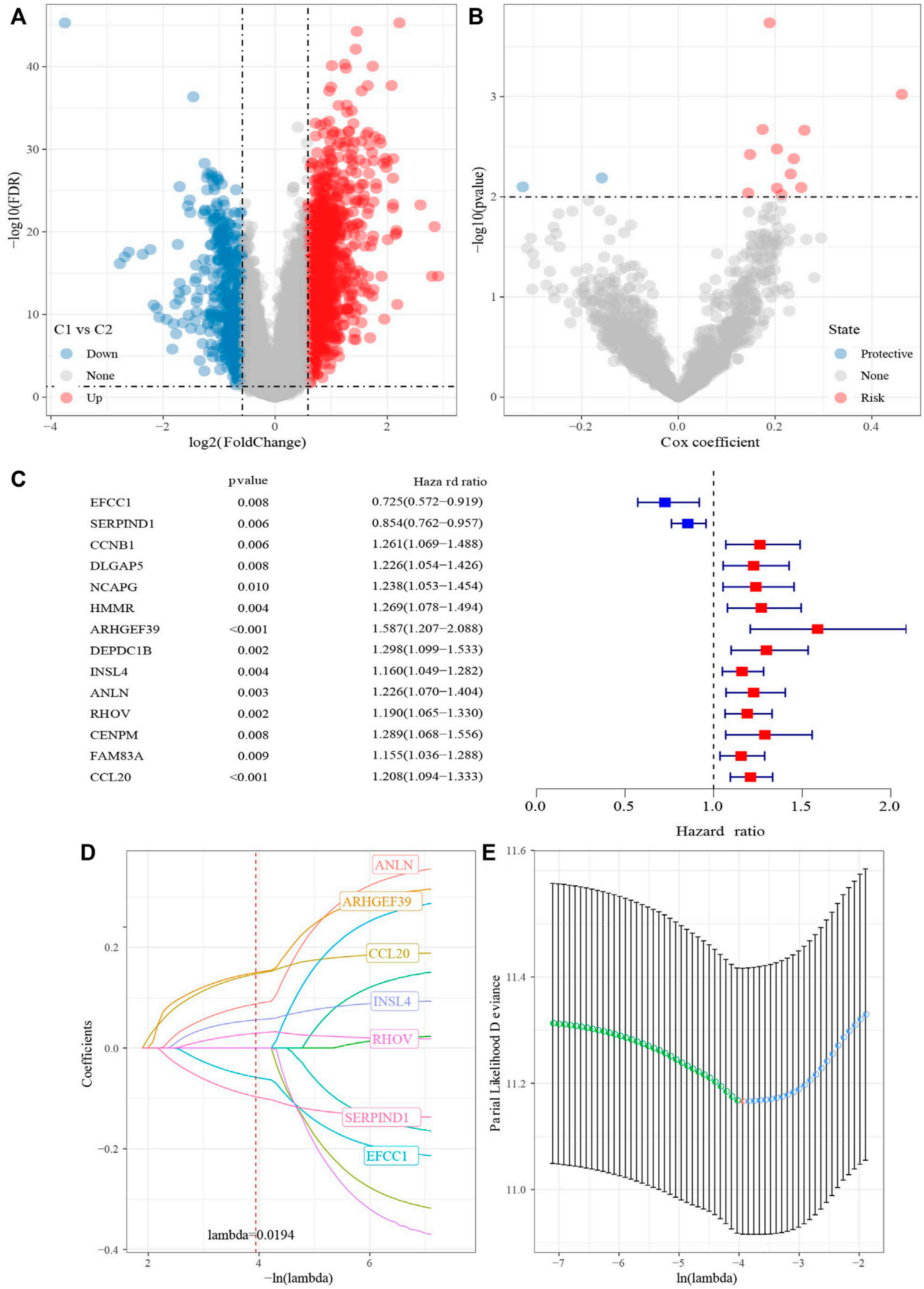


FIGURE 7
Identification hub cuproptosis related genes. **(A)** Volcano of differentially expressed genes identified from C1 and C2. **(B)** Volcano of differentially expressed genes identified using univariate Cox analysis. **(C)** Forest map of differentially expressed genes identified using univariate Cox analysis. **(D)** Lambda trajectory of differentially expressed genes. **(E)** Confidence interval under lambda.

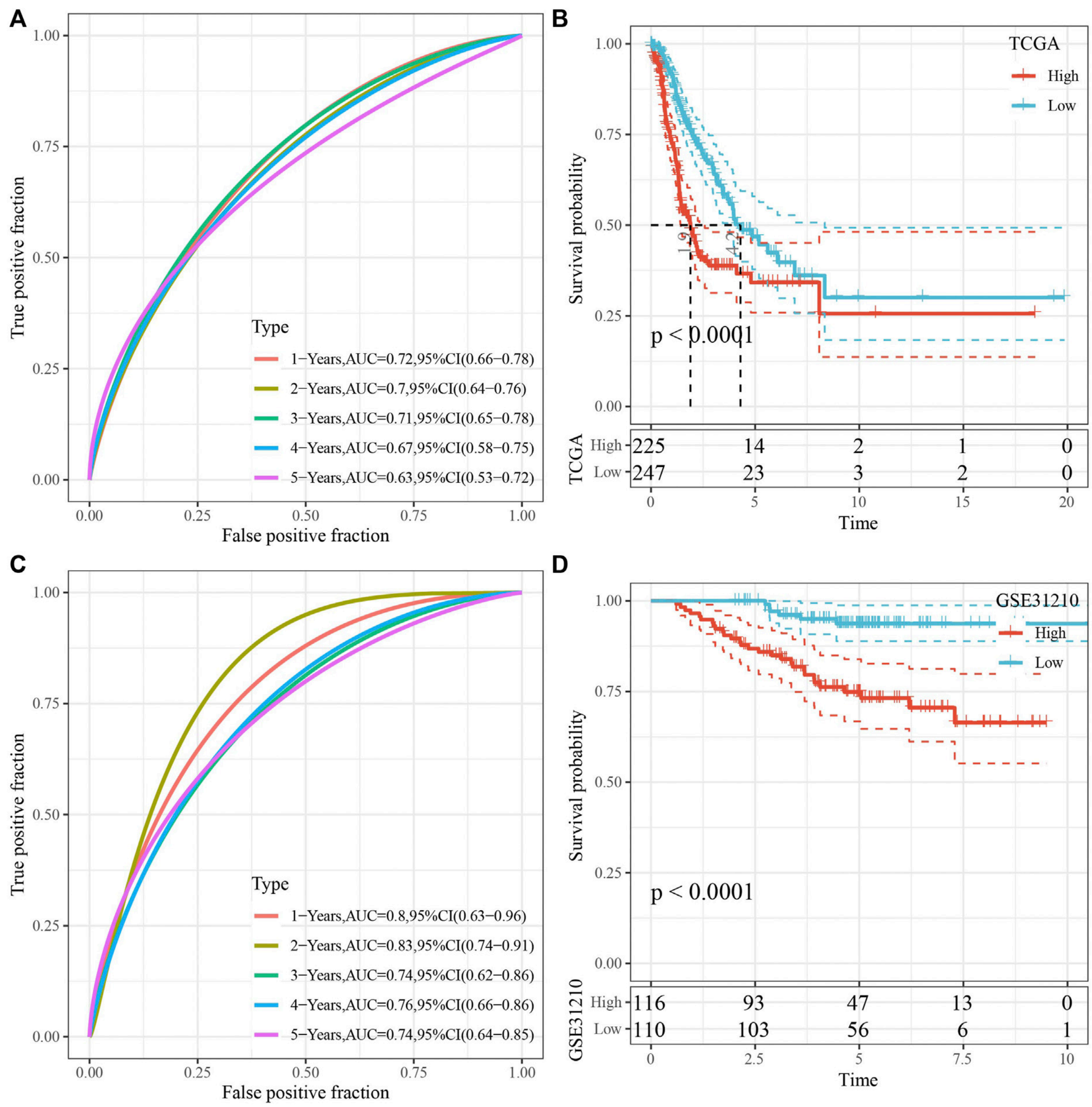


FIGURE 8
Validation of RiskScore. **(A)** ROC analysis of RiskScore in TCGA-test dataset. **(B)** KM survival analysis of RiskScore in TCGA-test dataset. **(C)** ROC analysis of RiskScore in GSE31210 dataset. **(D)** KM survival analysis of RiskScore in GSE31210 dataset.

Prognostic model has well predictive performance

RiskScore was for zscore, and the samples into high-risk (RiskScore> 0) and low-risk (RiskScore <0) groups in TCGA-test and GSE31210 dataset. ROC and survival analyses were performed in TCGA-test dataset (Figures 8A,B) and GSE31210 dataset (Figures 8C,D). The results revealed that the accuracy of the model was better in predicting the 1-, 2-, 3-, 4-, and 5-year survival rates in above datasets, as all values of the area under the curve (AUC) were greater than 0.6. Results of Kaplan-Meier survival analysis showed overall survival was higher in low-risk group than high-risk group. High group had more samples with higher clinical grade (Figure 9A), the RiskScore was higher in MALE, a higher T stage, N2 stage and clinical stage, and dead samples (Figure 9B).

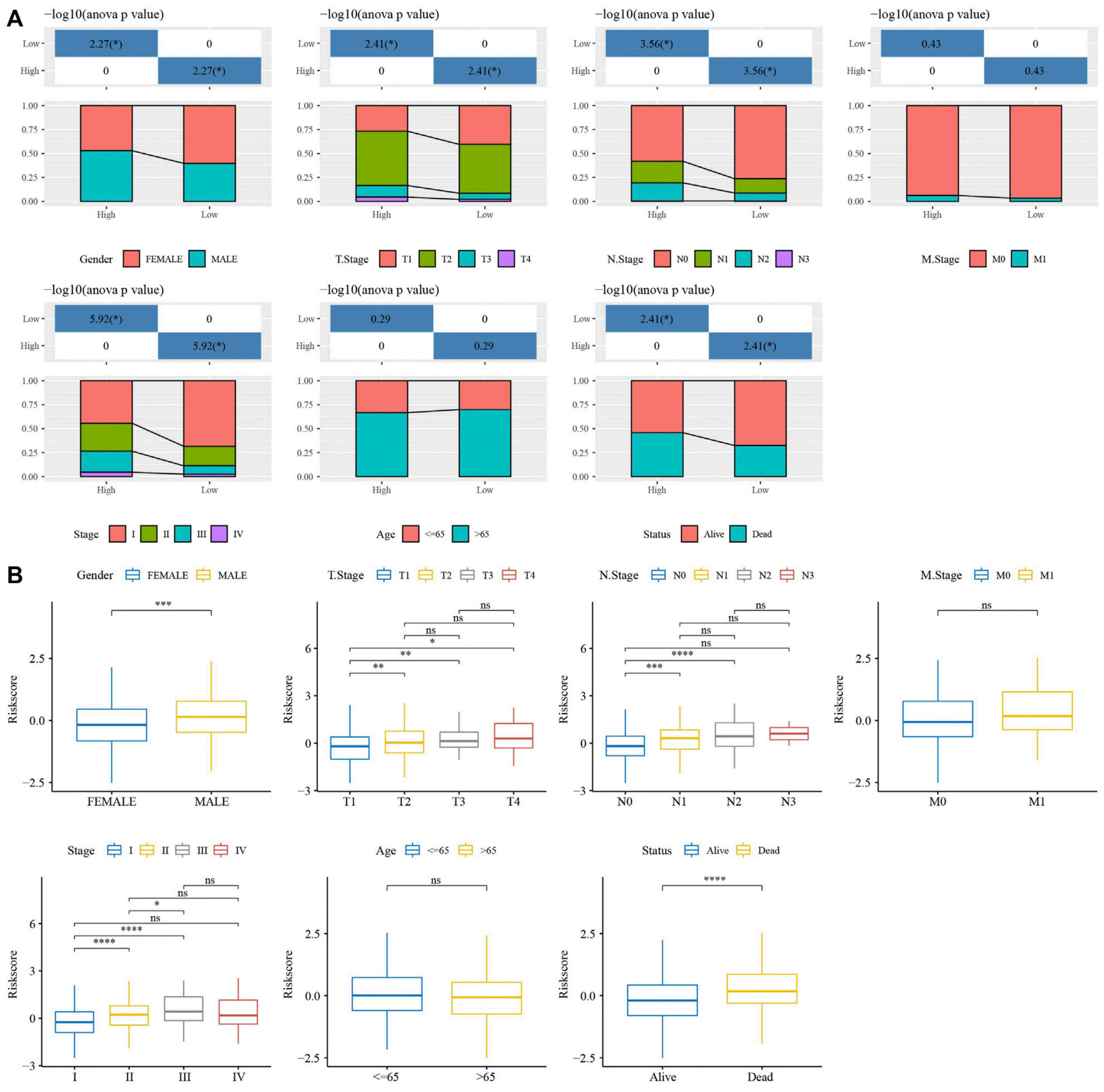


FIGURE 9
Analysis of clinical features in RiskScore. **(A)** The distribution of clinical features groups, included Gender, T Stage, N Stage, M Stage, Stage, Age and Status, in high group and low group. **(B)** The RiskScore differences analysis in clinical features groups, included Gender, T Stage, N Stage, M Stage, Stage, Age and Status. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{***}p < 0.0001$, ns: no significance.

RiskScore was an independent prognostic factor

To identify the independence of 7-gene signature model in clinical application, in TCGA-LUAD dataset, univariate and multivariate COX regression were used to analyze the HR, 95%CI of HR and Pvalue of Age, Gender, T Stage, N Stage, M Stage, Stage and RiskType. Univariate COX regression analysis showed that T Stage, N Stage, Stage and RiskType were significantly associated with survival (Figure 10A), while multivariate COX regression analysis showed that only RiskType (HR = 2.06, 95%CI = 1.43–2.99, $p < 0.001$) was still significantly associated with survival (Figure 10B). Those data imply that RiskType was an independent prognostic factor.

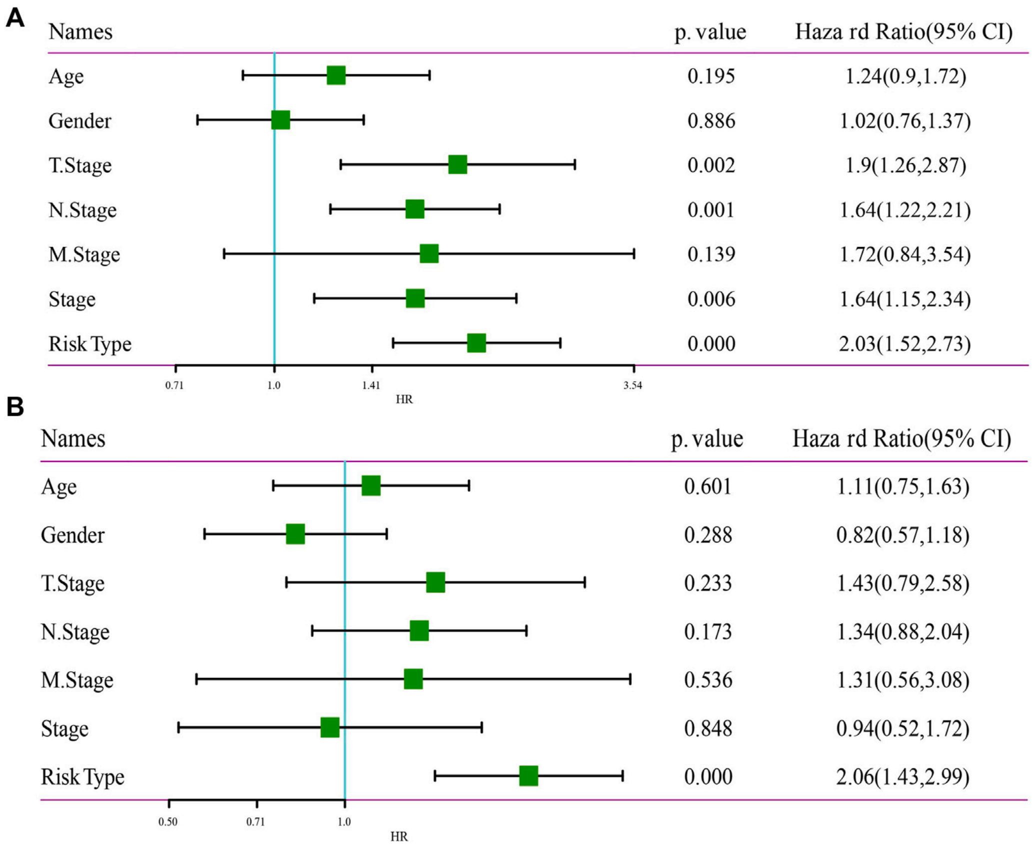


FIGURE 10
Independence of RiskScore. **(A)** Univariate Cox regression analysis. **(B)** Multivariate Cox regression analysis.

Low group had higher immune infiltration and sensitived to immunotherapy

CIBERSORT analysis indicated that 14 of 22 immune cells, most were significantly higher in low group that those in high group (Figure 11A). While, ESTIMATE analysis showed that low group had higher StromalScore, ImmuneScore and ESTIMATEScore (Figure 11B). And 24 immune checkpoint genes had obviously difference expressions between high group and low group (Figure 11C). TIDE, MDSC and Exclusion were lower in low group that in high group, while Dysfunction and TAM.M2 were higher in low group (Figure 11D), suggesting that low group was more likely to benefit from immunotherapy. IC50 of Cisplatin, Salubrinal, Vinorelbine, QS11, AKT inhibitor VIII and Embelin were higher in low group, which suggested the developed model could be used to predict chemotherapeutic drug sensitivity (Figure 11E).

Discussion

Cuproptosis is a newly discovered form of cell death, which is characterized by the accumulation of intracellular free copper and the lipidation of proteins leading to cytotoxic stress, thereby inducing cell death (Tsvetkov et al., 2022). However, the mechanism of copper death in LUAD has not been studied. Based on this, the relevant microarray was downloaded from TCGA and GEO databases, and the correlation and difference of immune infiltration were analyzed. Then, the results were integrated with cuproptosis related genes, and the risk model was constructed. Finally, seven copper death genes related to lung adenocarcinoma were screened out, including ARHGEF39, EFCC1, SERPIND1, INSL4, ANLN, RHOV and CCL20.

The overexpression of ARHGEF39 has also been identified in various human malignancies, including non-small cell lung cancer (Zhou et al., 2018), gastric cancer (Wang et al., 2018), and hepatocellular carcinoma (Wang et al., 2012). Decreased expression of EFCC1 was significantly associated with progression of LUAD (Xia et al., 2019; Yu and Zhang, 2020). SERPIND1 acts as a potential oncogene in the development of tumor, including in lung cancer (Bossé et al., 2012; Zhu et al., 2016). INSL4 as prognostic marker for proliferation and invasiveness in Non-Small-Cell Lung Cancer (Scopetti et al., 2021). ANLN participates in cell developmental processes *via* regulating nuclear division pathway in LUAD (Long et al., 2018). Overexpression of RHOV in LUAD promotes the progression (Chen et al., 2021). Production of

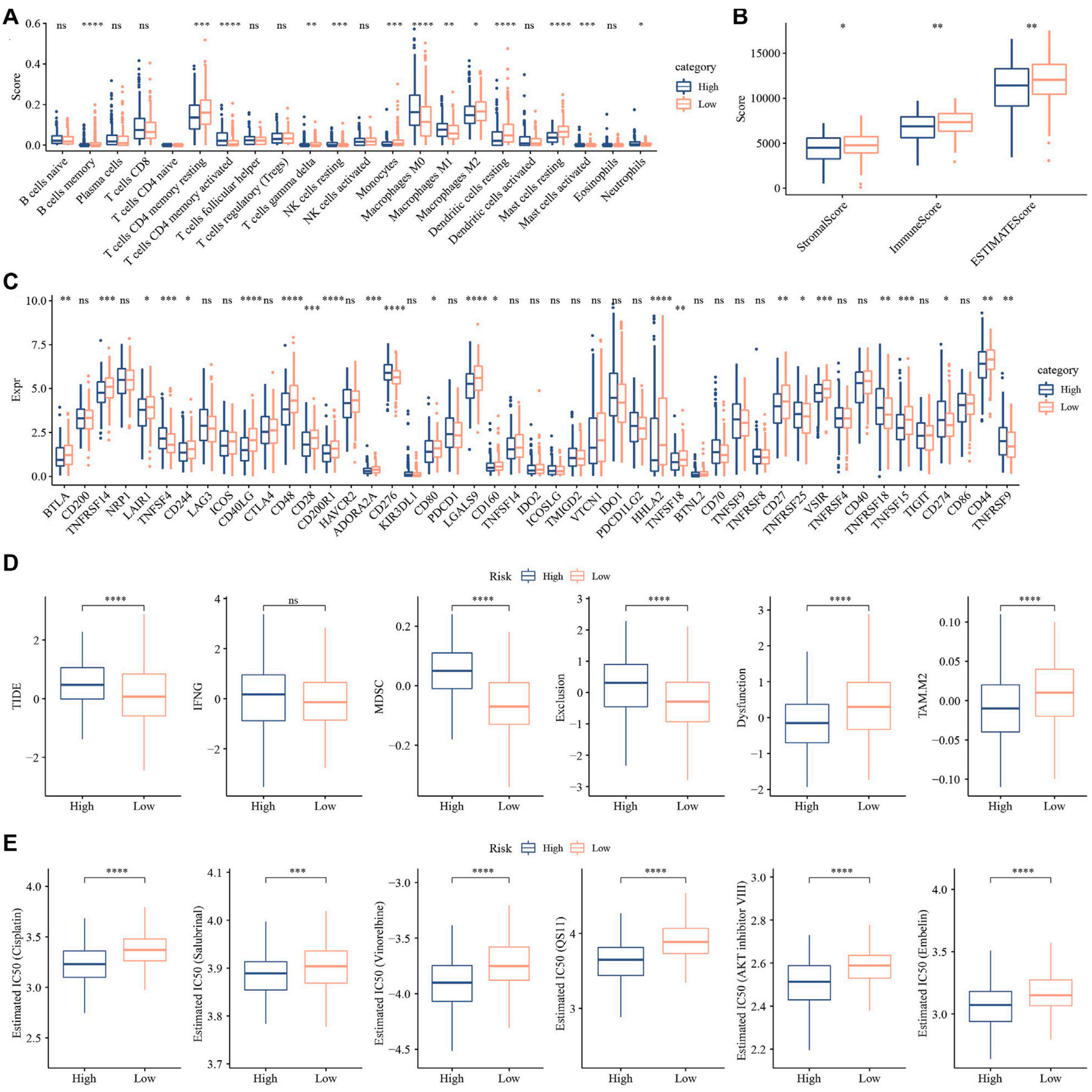


FIGURE 11
Analysis of immune infiltration. **(A)** analysis of 22 immune cells using CIBERSORT. **(B)** Analysis of immune infiltration using ESTIMATE. **(C)** The expression levels of 42 immune check genes between low group and high group. **(D)** The differences of TIDE, IFNG, MDSC, Exclusion, Dysfunction and TAM.M2 between low group and high group. **(E)** IC50 of traditional drugs in low group and high group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ***$p < 0.0001$, ns: no significance.

CCL20 from lung cancer cells induces the cell migration and proliferation (Wang et al., 2016). To sum up, although the copper death related gene in LUAD mechanism study is less, but according to previous research and the research results can be speculated that cuproptosis related genes may play an important role in LUAD progress, steady state and how to adjust the copper to prevention and treatment of LUAD, is the need for further research.

The analysis results of this study have certain reference value for the subsequent basic research of cuproptosis on LUAD, and could reduce unnecessary waste in experiments to a certain extent. However, this study still has some limitations. First, although the chip data used has met the sample size required by the research, the results may still be biased due to the small sample size. Second, although cuproptosis related genes associated with LUAD have been screened out, their specific mechanism of action has not been elucidated, which needs to be further explored in subsequent studies.

Data availability statement

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

Author contributions

All authors contributed to this present work: YL designed the study, YX acquired the data. XC drafted the manuscript, HL and LX revised the manuscript. All authors read and approved the manuscript.

Funding

The present study was supported by Liaoning Province Science and Technology Planning project (2021JH2/10300091).

Conflict of interest

YX, XC, and HL were employed by the company YuceBio Technology Co., 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1016871/full#supplementary-material

References

Bossé, Y., Postma, D. S., Sin, D. D., Lamontagne, M., Couture, C., Gaudreault, N., et al. (2012). Molecular signature of smoking in human lung tissues. *Cancer Res.* 72 (15), 3753–3763. doi:10.1158/0008-5472.CAN-12-1160

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

Chen, H., Xia, R., Jiang, L., Zhou, Y., Xu, H., Peng, W., et al. (2021). Overexpression of RhoV promotes the progression and EGFR-TKI resistance of lung adenocarcinoma. *Front. Oncol.* 11, 619013. doi:10.3389/fonc.2021.619013

Choi, B. Y., Jang, B. G., Kim, J. H., Seo, J. N., Wu, G., Sohn, M., et al. (2013). Copper/zinc chelation by clioquinol reduces spinal cord white matter damage and behavioral deficits in a murine MOG-induced multiple sclerosis model. *Neurobiol. Dis.* 54, 382–391. doi:10.1016/j.nbd.2013.01.012

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

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 (2), 102–113. doi:10.1038/s41568-021-00417-2

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

Goeman, J. J. (2010). L1 penalized estimation in the Cox proportional hazards model. *Biom. J.* 52 (1), 70–84. doi:10.1002/bimj.200900028

Ito, K., and Murphy, D. (2013). Application of ggplot2 to pharmacometric graphics. *CPT. Pharmacometrics Syst. Pharmacol.* 2 (10), e79. doi:10.1038/psp.2013.56

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

Kim, B. E., Nevitt, T., and Thiele, D. J. (2008). Mechanisms for copper acquisition, distribution and regulation. *Nat. Chem. Biol.* 4 (3), 176–185. doi:10.1038/nchembio.72

Kleczko, E. K., Kwak, J. W., Schenk, E. L., and Nemenoff, R. A. (2019). Targeting the complement pathway as a therapeutic strategy in lung cancer. *Front. Immunol.* 10, 954. doi:10.3389/fimmu.2019.00954

Liu, Y., He, M., Wang, D., Diao, L., Liu, J., Tang, L., et al. (2017). HisgAtlas 1.0: A human immunosuppression gene database. *Database.* 2017, bax094. doi:10.1093/database/bax094

Long, X., Zhou, W., Wang, Y., and Liu, S. (2018). Prognostic significance of ANLN in lung adenocarcinoma. *Oncol. Lett.* 16 (2), 1835–1840. doi:10.3892/ol.2018.8858

Luo, C., Lei, M., Zhang, Y., Zhang, Q., Li, L., Lian, J., et al. (2020). Systematic construction and validation of an immune prognostic model for lung adenocarcinoma. *J. Cell. Mol. Med.* 24 (2), 1233–1244. doi:10.1111/jcmm.14719

Lutsenko, S. (2010). Human copper homeostasis: A network of interconnected pathways. *Curr. Opin. Chem. Biol.* 14 (2), 211–217. doi:10.1016/j.cbpa.2010.01.003

Osmani, L., Askin, F., Gabrielson, E., and Li, Q. K. (2018). Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy. *Semin. Cancer Biol.* 52 (1), 103–109. doi:10.1016/j.semcancer.2017.11.019

Park, J. Y., and Jang, S. H. (2016). Epidemiology of lung cancer in korea: Recent trends. *Tuberc. Respir. Dis.* 79 (2), 58–69. doi:10.4046/trd.2016.79.2.58

Scopetti, D., Piobbico, D., Brunacci, C., Pieroni, S., Bellezza, G., Castelli, M., et al. (2021). INSL4 as prognostic marker for proliferation and invasiveness in Non-Small-Cell Lung Cancer. *J. Cancer* 12 (13), 3781–3795. doi:10.7150/jca.51332

Shen, W., Song, Z., Xiao, Z., Huang, M., Shen, D., Gao, P., et al. (2022). Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. *iMeta* 1 (3), e36. doi:10.1002/imt2.36

Tan, H. Y., Wang, N., Zhang, C., Chan, Y. T., Yuen, M. F., and Feng, Y. (2021). Lysyl oxidase-like 4 fosters an immunosuppressive microenvironment during hepatocarcinogenesis. *Hepatol. Baltim. Md)* 73 (6), 2326–2341. doi:10.1002/hep.31600

Testa, U., Castelli, G., and Pelosi, E. (2018). Lung cancers: Molecular characterization, clonal heterogeneity and evolution, and cancer stem cells. *Cancers* 10 (8), E248. doi:10.3390/cancers10080248

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

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. *Sci. (New York, NY)* 375 (6586), 1254–1261. doi:10.1126/science.abf0529

Wang, B., Shi, L., Sun, X., Wang, L., Wang, X., and Chen, C. (2016). Production of CCL20 from lung cancer cells induces the cell migration and proliferation through PI3K pathway. *J. Cell. Mol. Med.* 20 (5), 920–929. doi:10.1111/jcmm.12781

Wang, H., Li, M., Tao, X., Qian, Y., Chen, L., and Tao, G. (2018). ARHGEF39 promotes gastric cancer cell proliferation and migration via Akt signaling pathway. *Mol. Cell. Biochem.* 440 (1-2), 33–42. doi:10.1007/s11010-017-3153-3

Wang, H., Li, Y., Wang, Y., Han, Z. G., and Cai, B. (2012). C9orf100, a new member of the Dbl-family guanine nucleotide exchange factors, promotes cell proliferation and migration in hepatocellular carcinoma. *Mol. Med. Rep.* 5 (5), 1169–1174. doi:10.3892/mmr.2012.783

Wang, Y., Huang, J., Wu, Q., Zhang, J., Ma, Z., Ma, S., et al. (2021). Downregulation of breast cancer resistance protein by long-term fractionated radiotherapy sensitizes lung adenocarcinoma to SN-38. *Invest. New Drugs* 39 (2), 458–468. doi:10.1007/s10637-020-01003-3

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

Xia, L., Zhu, Y., Zhang, C., Deng, S., Deng, Y., Yang, Z., et al. (2019). Decreased expression of EFCC1 and its prognostic value in lung adenocarcinoma. *Ann. Transl. Med.* 7 (22), 672. doi:10.21037/atm.2019.10.41

Xing, P., Wang, S., Wang, Q., Ma, D., Hao, X., Wang, M., et al. (2019). Efficacy of crizotinib for advanced ALK-rearranged non-small-cell lung cancer patients with brain metastasis: A multicenter, retrospective study in China. *Target. Oncol.* 14 (3), 325–333. doi:10.1007/s11523-019-00637-5

Yang, P., Chen, W., Xu, H., Yang, J., Jiang, J., Jiang, Y., et al. (2021). Correlation of CCL8 expression with immune cell infiltration of skin cutaneous melanoma: Potential as a prognostic indicator and therapeutic pathway. *Cancer Cell Int.* 21 (1), 635. doi:10.1186/s12935-021-02350-8

Yu, S., and Zhang, G. (2020). EFCC1 as a putative prognostic biomarker in lung adenocarcinoma. *Ann. Transl. Med.* 8 (16), 981. doi:10.21037/atm-2020-93

Zhou, H., Cai, L., Zhang, X., Li, A., Miao, Y., Li, Q., et al. (2018). ARHGEF39 promotes tumor progression via activation of Rac1/P38 MAPK/ATF2 signaling and predicts poor prognosis in non-small cell lung cancer patients. *Lab. Invest.* 98 (5), 670–681. doi:10.1038/s41374-018-0022-y

Zhu, L., Guo, Q., Jin, S., Feng, H., Zhuang, H., Liu, C., et al. (2016). Analysis of the gene expression profile in response to human epididymis protein 4 in epithelial ovarian cancer cells. *Oncol. Rep.* 36 (3), 1592–1604. doi:10.3892/or.2016.4926

Frontiers in Genetics

TYPE Original Research
PUBLISHED 14 October 2022
DOI 10.3389/fgene.2022.1004163

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Weimin Zhong,
Xiamen Fifth Hospital, China
Shiming Chen,
Zhejiang University, China

*CORRESPONDENCE
Wenzheng Zhou,
1975255404@qq.com
Zhao Wang,
xjzzqwz@163.com
Peng Chao,
cp427701@163.com

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

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

RECEIVED 26 July 2022
ACCEPTED 27 September 2022
PUBLISHED 14 October 2022

CITATION
Li J, Zhang X, Ren P, Wu Y, Wang Y, Zhou W, Wang Z and Chao P (2022), Landscape of RNA-binding proteins in diagnostic utility, immune cell infiltration and PANoptosis features of heart failure.
Front. Genet. 13:1004163.
doi: 10.3389/fgene.2022.1004163

Landscape of RNA-binding proteins in diagnostic utility, immune cell infiltration and PANoptosis features of heart failure

Jie Li[1†], Xueqin Zhang[2†], Peng Ren[1], Yu Wu[3], Yaoguo Wang[4], Wenzheng Zhou[5]*, Zhao Wang[1]* and Peng Chao[1]*

[1]Department of Cardiology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China, [2]Department of Nephrology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China, [3]Department of Medical Administration, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China, [4]Department of Information Center, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China, [5]Department of Orthopaedics, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China

Objective: Heart failure remains a global public health problem linked to rising morbidity and mortality. RNA-binding proteins (RBPs) are crucial regulators in post-transcriptionally determining gene expression. Our study aimed to comprehensively elucidate the diagnostic utility and biological roles of RBPs in heart failure.

Methods: Genomic data of human failing and nonfailing left ventricular myocardium specimens were retrieved from the GEO datasets. Heart failure-specific RBPs were screened with differential expression analyses, and RBP-based subtypes were clustered with consensus clustering approach. GSEA was implemented for comparing KEGG pathways across subtypes. RBP-based subtype-related genes were screened with WGCNA. Afterwards, characteristic genes were selected through integrating LASSO and SVM-RFE approaches. A nomogram based on characteristic genes was established and verified through calibration curve, decision curve and clinical impact curve analyses. The abundance of immune cell types was estimated with CIBERSORT approach.

Results: Heart failure-specific RBPs were determined, which were remarkably linked to RNA metabolism process. Three RBP-based subtypes (namely C1, C2, C3) were established, characterized by distinct pathway activities and PANoptosis gene levels. C2 subtype presented the highest abundance of

---

Abbreviations: LVEF: left ventricular ejection fraction; RBPs: RNA binding proteins; GEO: Gene Expression Omnibus; LV: left ventricular; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PCA: principal component analyses; GSEA: Gene set enrichment analyses; WGCNA: weighted gene co-expression network analyses; TOM: topological overlap matrix; PPI: protein-protein interaction; STRING: Search Tool for the Retrieval of Interacting Genes; LASSO: least absolute shrinkage and selection operator regression; SVM-RFE: support vector machine recursive feature elimination; ROC: receiver operator characteristic; AUC: area under the curve; CMap: Connectivity Map.

immune cells, followed by C1 and C3. Afterwards, ten characteristic genes were selected, which enabled to reliably diagnose heart failure risk. The characteristic gene-based nomogram enabled to accurately predict risk of heart failure, with the excellent clinical utility. Additionally, characteristic genes correlated to immune cell infiltration and PANoptosis genes.

Conclusion: Our findings comprehensively described the roles of RBPs in heart failure. Further research is required for verifying the effectiveness of RBP-based subtypes and characteristic genes in heart failure.

KEYWORDS

heart failure, RNA binding protein, subtype, characteristic gene, nomogram, immune cells, PANoptosis

Introduction

Heart failure is a frequent complex clinical syndrome of symptoms and signs triggered by structural or functional abnormality that leads to impaired cardiac output (Packer et al., 2021b), which remains a growing public health issue affecting about 26 million individuals globally (Heidenreich et al., 2022a). Typically, in accordance with left ventricular ejection fraction (LVEF), heart failure is categorized as preserved (HFpEF) and reduced ejection fraction (HFrEF) (Heidenreich et al., 2022b). HFpEF is a filling issue because of muscle stiffness reducing left ventricular chamber size or left atrial dilation, while HFrEF is usually described as a mechanical left ventricular pump issue (Mascolo et al., 2022). Therapeutic strategies of above two types differ. Substantial evidence suggests that sequential drug treatment improves clinical outcomes in patients with HFrEF (Berg et al., 2021). Differently, no treatment options show prognostic benefits and symptom controlling as the sole management regimen of HFpEF patients (Packer et al., 2021a). Reduction of associated hospital readmission rate and clinical and economic burden remains a pivotal issue in modern cardiovascular medicine (Piepoli et al., 2022). Biomarkers that reflect the pathophysiological processes of heart failure progression can aid clinicians in the early diagnosis and management of heart failure patients.

RNA binding proteins (RBPs) control RNA fate from synthesis to decay, the expression and roles of which are highly determined by detailed networks of transcriptional, post-transcriptional as well as post-translational machinery (Liu et al., 2022). They are implicated in pathological manifestations of heart failure. For instance, RBPs have widespread translational control of human cardiac fibroblast activation (Chothani et al., 2019). RNA binding protein 24 loss enables to disrupt global alternative splicing as well as results in heart failure (Liu et al., 2019). RNA-binding protein RBM20 weakens splicing to orchestrate cardiac pre-mRNA processing, and contributes to the pathogenesis of heart failure (Maatz et al., 2014). RBP HuR-mediated SCN5A mRNA stability represses arrhythmic risk in heart failure (Zhou et al., 2018). Knockdown of RNA binding motif-20-based titin splicing system can upregulate compliant titins, and thus ameliorates diastolic function and exercise tolerance in heart failure (Methawasin et al., 2016). Despite this, comprehensive analyses of RBPs in heart failure are lacking. Accumulated evidence demonstrates that PANoptosis (apoptosis, necroptosis, together with pyroptosis) mediates heart failure progression and possesses promising therapeutic implications (Zhang et al., 2016; Zeng et al., 2019; Gao et al., 2020). Nonetheless, the mechanisms of RBPs underlying PANoptosis remain indistinct in hear failure. Our study comprehensively evaluated RBP-based molecular subtypes and relevant characteristic genes for heart failure, unveiling the crucial roles of RBPs in pathophysiological process of heart failure as well as providing reliable targets for diagnosing heart failure risk.

Materials and methods

Heart failure expression profiling

This study downloaded the expression profiling of human heart failure from the Gene Expression Omnibus (GEO) repository. The GSE5406 dataset comprised microarray expression profiles of 194 human failing left ventricular (LV) myocardium specimens and 16 human nonfailing control LV myocardium specimens on the Affymetrix platform (Hannenhalli et al., 2006). Additionally, we acquired microarray expression profiles of 177 human failing LV myocardium specimens and 136 human control specimens from the GSE57338 dataset on the basis of the Affymetrix platform (Liu et al., 2015). Above expression profiles were merged, and removal of batch effects was implemented utilizing sva package (Leek et al., 2012). Four independent datasets were utilized as external verification sets as follows: the GSE76701 dataset comprising expression profiles of 4 non-failing and 4 failing LV hearts (Kim et al., 2016); the GSE55296 dataset containing RNA-seq data of human hearts from 26 heart failure patients and 10 healthy controls (Tarazón et al., 2014); the GSE86569 dataset with expression profiling of LV hearts from 12 HFrEF and 20 HFpEF patients; the GSE3585 dataset including expression profiles of 5 non-failing and 7 failing LV hearts.

Screening heart failure-specific RBPs

Totally, 1,542 RBPs were collected from previously published literature (Supplementary Table S1) (Gerstberger et al., 2014). Expression values of RBPs were compared between human failing and nonfailing LV myocardium specimens through limma package (Ritchie et al., 2015). RBPs with adjusted $p < 0.05$ and |fold-change|>1.5 were regarded as heart failure-specific RBPs. Targets of RBPs were predicted through starBase database.

Functional enrichment analyses

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of RBPs with adjusted $p < 0.05$ were implemented utilizing clusterprofiler package (Yu et al., 2012). Terms with adjusted $p < 0.05$ were regarded as significant enrichment.

Consensus clustering analyses

Through ConsensusClusterPlus package (Wilkerson and Hayes, 2010), on the basis of RBPs with adjusted $p < 0.05$, consensus clustering analyses of heart failure specimens were run in accordance with the following parameters: item resampling, proportion of items to sample: 80%; gene resampling, proportion of features to sample: 80%; a maximum evaluated k, maximum clustering number to evaluate: 9; resampling, number of subsamples: 1,000; agglomerative hierarchical clustering algorithm): 'hc' (hclust); and distance: 'pearson' (1 - Pearson correlation). The RBP-based subtype classification was verified through principal component analyses (PCA).

Gene set enrichment analyses

GSEA methodology (Subramanian et al., 2005) was utilized for the comparisons of KEGG pathways across RBP-based subtypes. The "c2.cp.kegg.v7.4.symbols" gene set from the Molecular Signatures Database (Liberzon et al., 2015) was employed as a reference set, with the criteria of p-value<0.05.

Weighted gene co-expression network analyses

Expression profiling of the merged GSE5406 and GSE57338 datasets was extracted for conducting WGCNA utilizing WGCNA package (Langfelder and Horvath, 2008). Sample clustering was implemented for testing whether there were outlier specimens. Soft threshold power value was determined for constructing a scale-free topology network. Afterwards, the adjacency matrix was converted to the topological overlap matrix (TOM). In accordance with the TOM-based dissimilarity, distinct co-expression modules were clustered. Associations of modules and RBP-based subtypes were then evaluated. The module with the strongest correlation to RBP-based subtypes was screened, and the genes in this module were regarded as RBP-based subtype-related genes.

Protein-protein interaction analyses

RBP-based subtype-related genes were imported into the Search Tool for the Retrieval of Interacting Genes (STRING) online platform (https://www.string-db.org) (Szklarczyk et al., 2021). The interactions between their protein products were retrieved according to the default criteria. Utilizing MCODE plug-in of Cytoscape software (Doncheva et al., 2019), PPI subnetwork and hub genes were obtained following the selection criteria of degree cutoff = 2, node score cutoff = 0.2, haircut = true, Fluff = false, K-core = 2 Max, and depth from seed = 100.

Selection of characteristic genes with two machine learning approaches

Characteristic RBP-based subtype-related genes were selected utilizing two machine learning approaches: least absolute shrinkage and selection operator regression (LASSO) as well as support vector machine recursive feature elimination (SVM-RFE). Through glmnet package (Engebretsen and Bohlin, 2019), LASSO was run and penalty parameter λ tuning was implemented using ten-fold cross-validation. Additionally, the best variables were selected with ten-fold cross-validation utilizing SVM-RFE algorithm. Afterwards, characteristic genes were determined through intersection of LASSO- and SVM-RFE-derived results.

Establishment of a nomogram

A nomogram was established on the basis of characteristic genes *via* rms package. The accuracy of the nomogram in

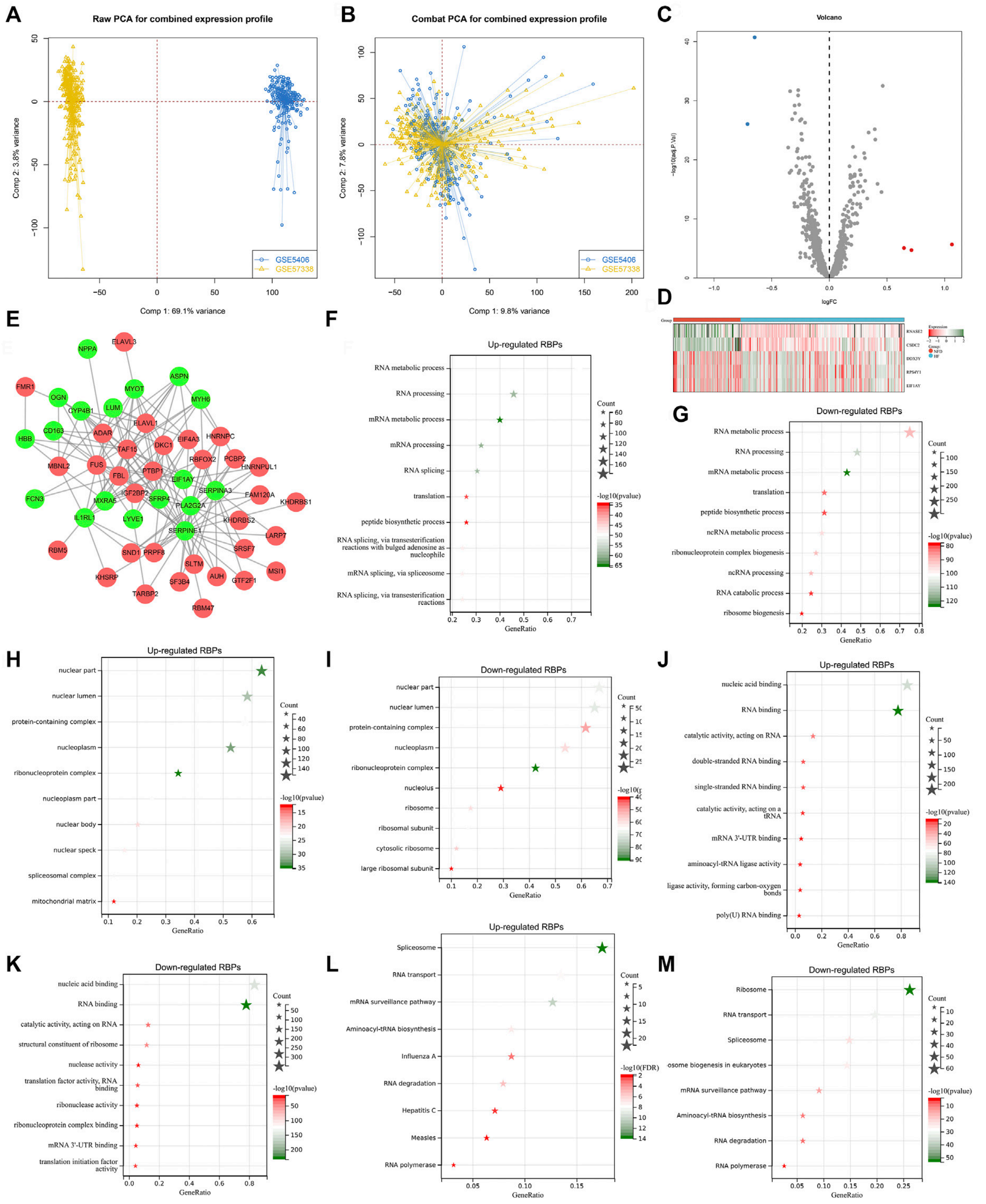


FIGURE 1
Analyses of heart failure-specific RBPs. **(A)** Integration of two heart failure expression profiling datasets (GSE5406, and GSE57338). **(B)** Removal of batch effects of the merged datasets. **(C)** Volcano plots of RBPs with differential expression between human failing and nonfailing control LV myocardium specimens following adjusted $p < 0.05$ and |fold-change|>1.5. **(D)** Heatmap of expression values of heart failure-specific RBPs in human failing (blue) and nonfailing control (red) LV myocardium specimens. **(E)** Potential downstream targets of RBPs with adjusted $p < 0.05$ that were differentially expressed in heart failure LV myocardium specimens than nonfailing controls. **(F–K)** The first ten biological processes, cellular components, and molecular functions of up- and down-regulated RBPs. **(L,M)** KEGG pathways enriched by up- and down-regulated RBPs.

predicting risk probabilities was evaluated with calibration curve. Decision curve analyses represent a novel approach for assessing clinical usefulness, which were utilized to evaluate the clinical utility of the nomogram. Clinical impact curves were drawn for assessing the clinical usefulness and applicability net benefits of the nomogram with the optimal diagnostic value.

Immune cell estimations

CIBERSORT (Newman et al., 2015) is an approach on the basis of the gene expression matrix for reliably estimating the relative abundance of 24 immune cell types in tissue specimens. CIBERSORT analyses were utilized for comparing differences in distinct immune cell types between groups. Spearman correlation analyses were implemented for exploring interactions between infiltrating immune cell types and characteristic genes.

Connectivity map (CMap) analysis

Heart failure-specific RBPs were utilized to query the CMap database (https://clue.io/). Compounds with $p < 0.05$ were selected as potential therapeutic drugs for heart failure based upon transcriptome data. The mode of action (MoA) of these compounds was then analyzed.

Statistical analyses

R software (www.r-project.org; version 3.6.1) was employed for all statistics analysis processes. Comparison analyses between groups were implemented utilizing Wilcoxon or Kruskal-Wallis test. The diagnostic efficacy of characteristic genes or nomogram was evaluated with receiver operator characteristic (ROC) curves along with area under the curve (AUC) calculation. Pearson or Spearmon correlation test was implemented for interactions between variables. $p < 0.05$ was considered statistically significant for all analysis process.

Results

Analyses of heart failure-specific RBPs

Our study collected and merged two heart failure expression profiling datasets (GSE5406, and GSE57338), and batch effects were corrected for subsequent analyses (Figures 1A,B). To determine heart failure-specific RBPs, differential expression analyses were implemented. In accordance with adjusted $p < 0.05$ and $|\text{fold-change}|>1.5$, five RBPs (EIF1AY, RPS4Y1, DDX3Y, RNASE2, and CSDC2) were found in heart failure LV myocardium specimens in comparison to nonfailing controls (Figures 1C,D). Afterwards, we predicted potential downstream targets of RBPs with adjusted $p < 0.05$ through starBase database. As depicted in Figure 1E, 18 targets (SERPINA3, FCN3, LUM, ASPN, IL1RL1, SFRP4, CD163, MYOT, OGN, MXRA5, LYVE1, MYH6, PLA2G2A, CYP4B1, SERPINE1, HBB, NPPA, and EIF1AY) had the potential binding sites of RBPs, which were differentially expressed in heart failure LV myocardium specimens in contrast to nonfailing controls. Biological functions and pathways of RBPs with adjusted $p < 0.05$ were then probed. Biological processes such as RNA/mRNA/peptide/ncRNA metabolic processes, RNA processing, and translation were both remarkably linked to up- and down-regulated RBPs (Figures 1F,G), indicating their essential roles in modulating gene expression. Additionally, RBPs with adjusted $p < 0.05$ closely correlated to multiple cellular components (nuclear part and lumen, protein-containing complex, nucleoplasm, etc.), as illustrated in Figures 1H,I. Up- and down-regulated RBPs also possessed the molecular functions of nucleic acid binding, RNA binding, catalytic activity acting on RNA, etc. (Figures 1J,K). To probe signaling pathways involved in RBPs with adjusted $p < 0.05$, KEGG enrichment analyses were implemented. In Figures 1L,M, RNA transport/degradation/polymerase, spliceosome, mRNA surveillance pathway, etc. were remarkably enriched by up- and down-regulated RBPs.

Establishment of RBP-based subtypes for heart failure

Consensus clustering analyses were employed for identifying RBP-based subtypes across heart failure specimens in accordance with RBPs with adjusted $p < 0.05$. Figure 2A illustrated the consensus matrix heatmap at $k = 3$. We found that heart failure specimens could be clearly categorized as three RBP-based subtypes, namely C1, C2 and C3. Consistent cumulative distribution (CDF) plot showed that when $k = 3$, CDF reached an approximate maximum (Figure 2B). As depicted in delta area plot, when $k = 4$, the area under the CDF curve increased only slightly, and thus 3 was an appropriate k value (Figure 2C). Tracking plot was also established for visualizing the sample classification. When $k = 3$, this classification had relatively high stability (Figure 2D). By reason of the foregoing, three RBP-based subtypes were finally identified across heart failure samples. The accuracy of this classification was verified through PCA plot. In Figure 2E, heart failure samples were clearly classified as three subtypes. Additionally, RBPs with adjusted $p < 0.05$ presented different expression values across three RBP-based subtypes (Figure 2F). Differentially expressed genes among three RBP-based subtypes were analyzed, and we

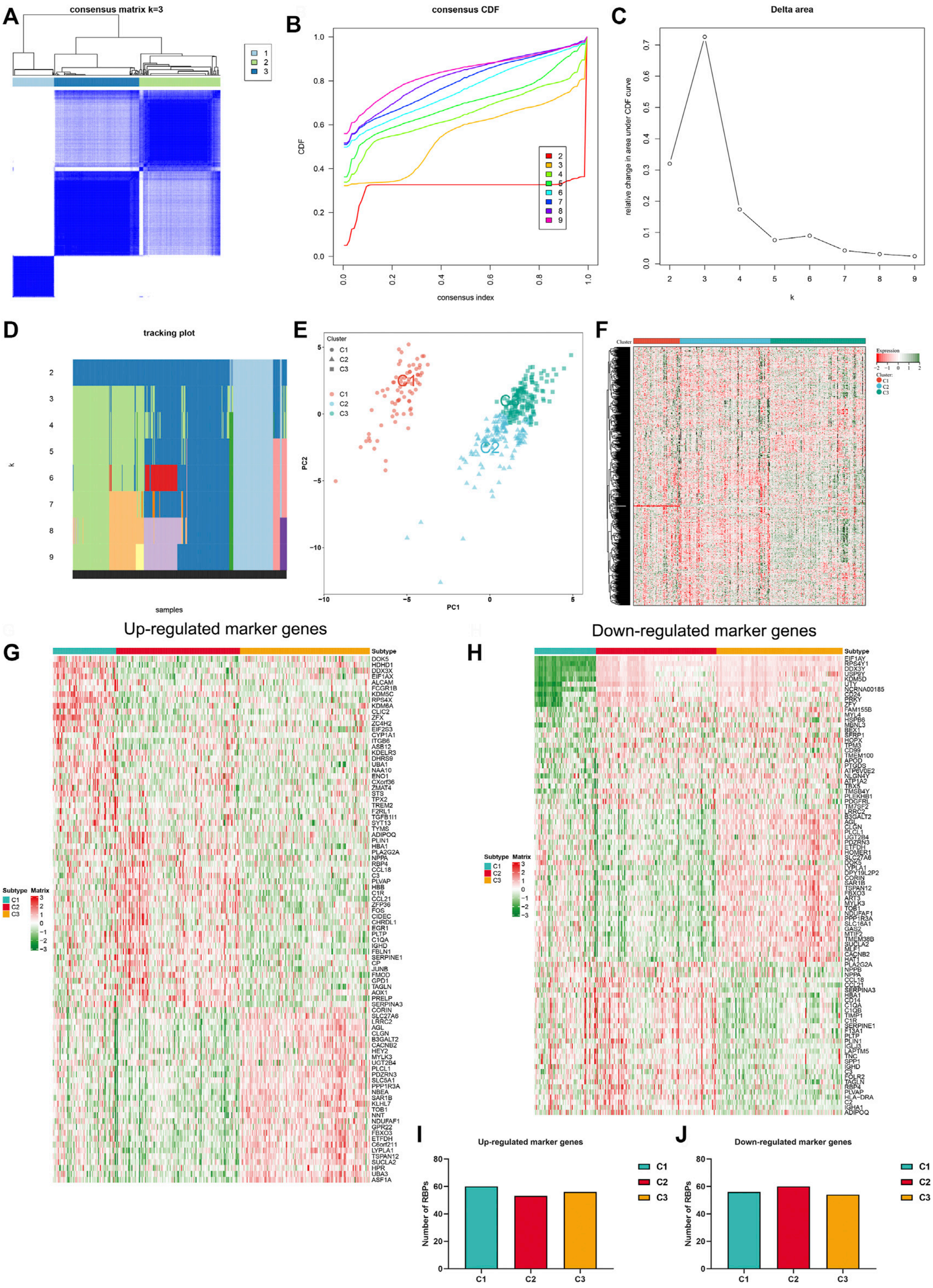


FIGURE 2
Establishment of RBP-based subtypes for heart failure. **(A)** Consensus matrix heatmap at $k = 3$. The rows and columns are samples, with consensus values on a white to blue color scale. **(B)** Consensus CDF curves at $k = 2–9$. **(C)** Delta area plot. **(D)** Tracking plot. The rows are samples, and the columns are k values. **(E)** PCA plot of three RBP-based subtypes. **(F)** Heatmap of expression values of RBPs with adjusted $p < 0.05$ across three subtypes. **(G)** The top 30 up-regulated marker genes in each subtype. **(H)** The top 30 down-regulated marker genes in each subtype. **(I,J)** Number of RBPs of up- and down-regulated marker genes in three subtypes.

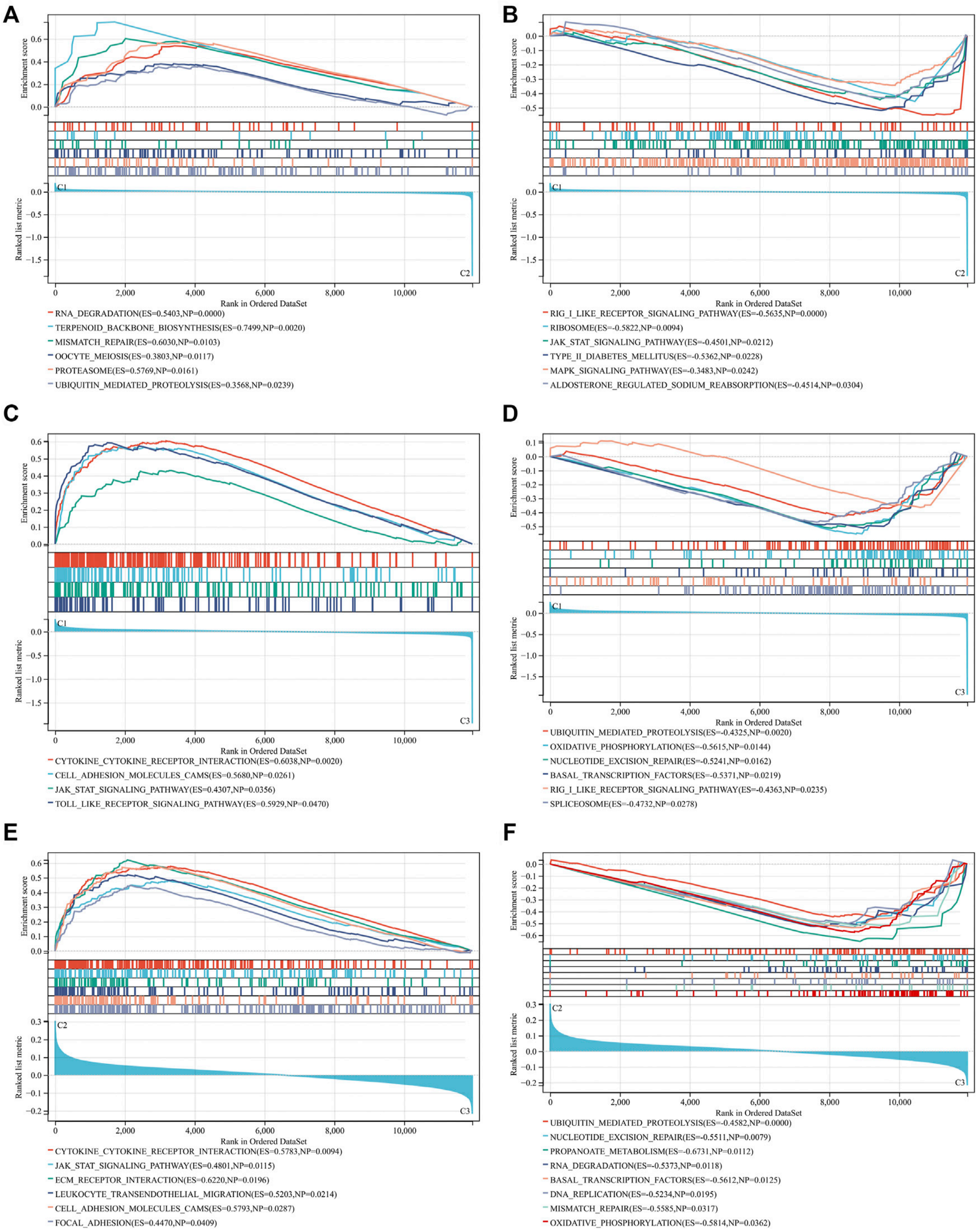


FIGURE 3
Different molecular mechanisms across RBP-based subtypes. **(A,B)** GSEA for comparing KEGG pathways between C1 and C2 subtypes. **(C,D)** GSEA for comparing KEGG pathways between C1 and C3 subtypes. **(E,F)** GSEA for comparing KEGG pathways between C2 and C3 subtypes.

identified the top 30 up- or down-regulated genes in each subtype compared with others, which were considered as specific marker genes of each subtype (Figures 2G,H). Potential RBPs of up- and down-regulated marker genes were predicted, respectively. As illustrated in Figures 2I,J, no notable differences in number of RBPs of up- and

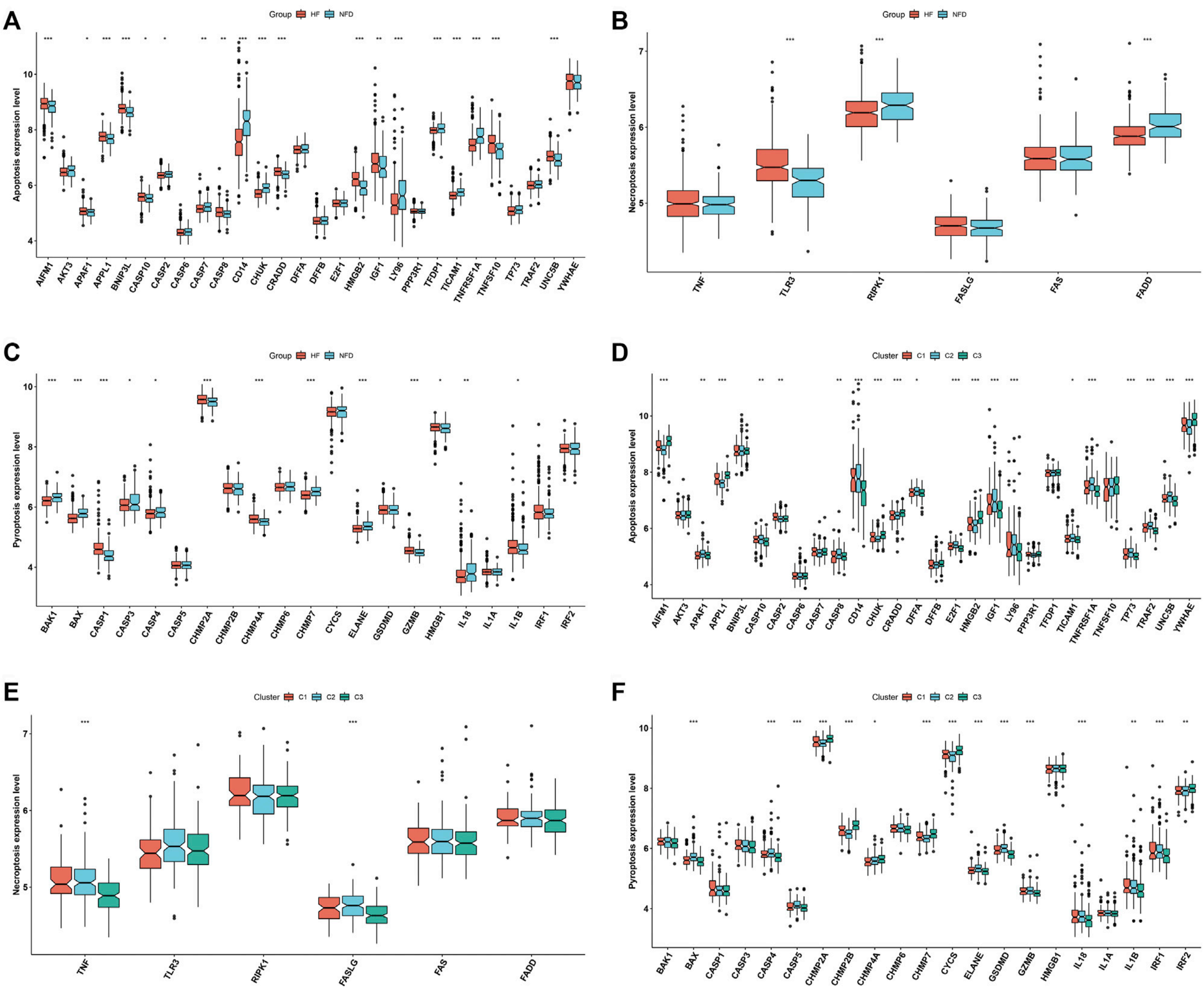


FIGURE 4
Distinct PANoptosis features across RBP-based subtypes. **(A–C)** Levels of apoptosis, necroptosis and pyroptosis genes in human failing and nonfailing control LV myocardium specimens. **(D–F)** Levels of apoptosis, necroptosis and pyroptosis genes across three RBP-based subtypes (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

down-regulated marker genes were found across three subtypes.

Different molecular mechanisms across RBP-based subtypes

Through GSEA, we dissected the differences in molecular mechanisms between distinct RBP-based subtypes. Compared to C2 subtype, RNA degradation, terpenoid backbone biosynthesis, mismatch repair, oocyte meiosis, proteasome, and ubiquitin mediated proteolysis were remarkably activated in C1 subtype (Figure 3A). Meanwhile, activation of RIG I like receptor signaling pathway, ribosome, JAK-STAT signaling pathway, type II diabetes mellitus, MAPK signaling pathway and aldosterone regulated sodium reabsorption was found in C2 subtype (Figure 3B). Molecular mechanisms between C1 and C3 subtypes were then compared. In Figure 3C, C1 subtype presented the remarkable activation of cytokine-cytokine receptor interaction, cell adhesion molecules CAMS, JAK-STAT signaling pathway and Toll-like receptor signaling pathway than C3 subtype. In comparison to C1 subtype, ubiquitin mediated proteolysis, oxidative phosphorylation, nucleotide excision repair, basal transcription factors, RIG I like receptor signaling pathway, and spliceosome were significantly activated in C3 subtype (Figure 3D). Additionally, we found the significant activation of cytokine-cytokine receptor interaction, JAK-STAT signaling pathway, ECM receptor

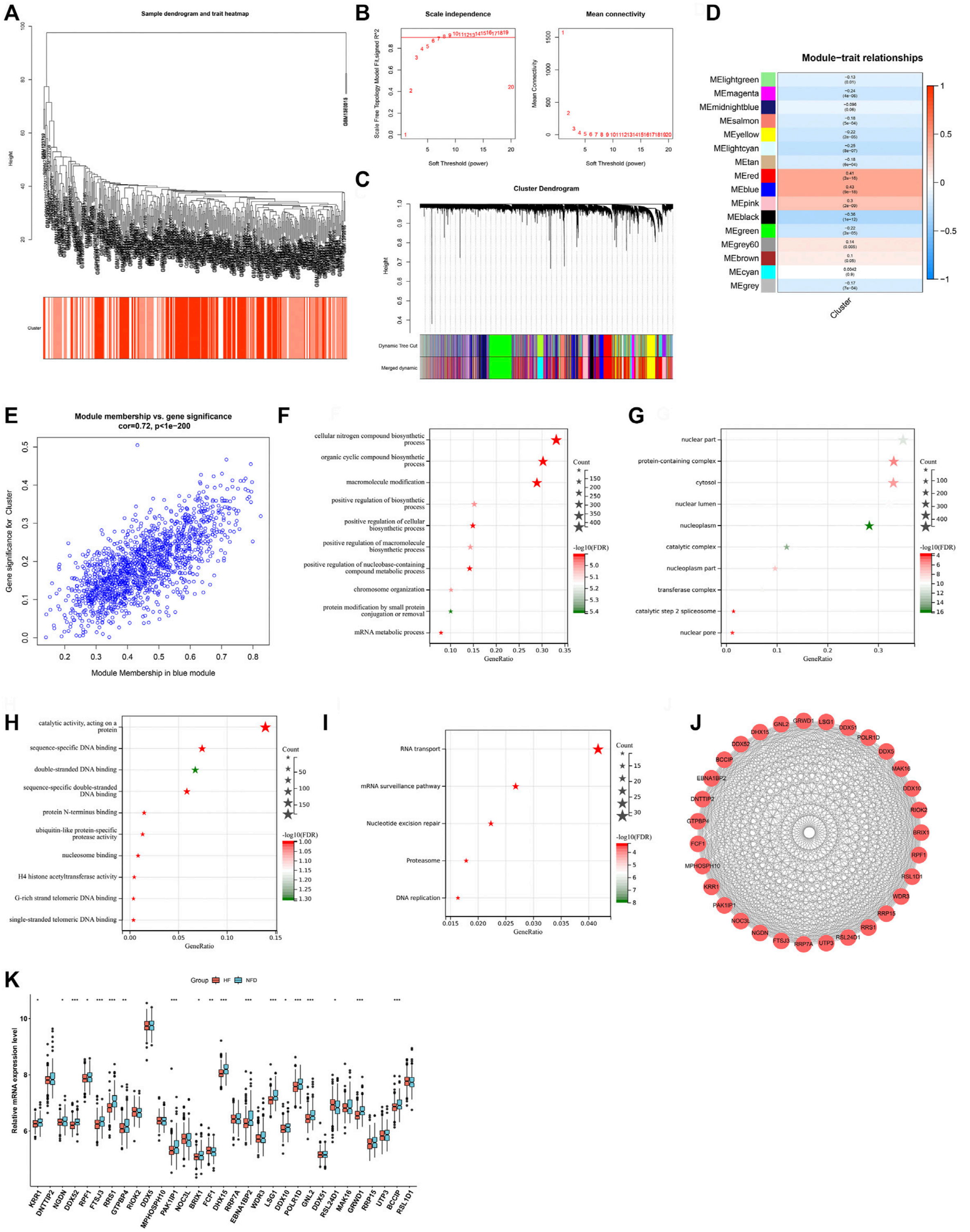


FIGURE 5
Identification of RBP-based subtype-related genes. **(A)** Clustering dendrogram of heart failure specimens on the basis of Euclidean distance. **(B)** Scale independence and mean connectivity at different power values. **(C)** Cluster dendrogram of distinct modules through dynamic tree cut approach and merged dynamic approach. The gray module represents unclassified genes. **(D)** Correlations between modules and RBP-based subtypes. The color indicates the strength of the correlation, and the number in parentheses indicates *p*-value. **(E)** Scatter plot of the relationship between module membership of blue module and gene significance for RBP-based subtypes. **(F–H)** The first ten biological processes, cellular components, and molecular functions of RBP-based subtype-related genes. **(I)** KEGG pathways significantly linked to RBP-based subtype-related genes. **(J)** The PPI subnetwork of key RBP-based subtype-related genes. **(K)** Box plot of the expressions of key RBP-based subtype-related genes in human failing and nonfailing control LV myocardium specimens. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

interactions, leukocyte trans-endothelial migration, cell adhesion molecules CAMS, and focal adhesion in C2 subtype in comparison to C3 subtype (Figure 3E). Meanwhile, ubiquitin mediated proteolysis, nucleotide excision repair, propanoate metabolism, RNA degradation, basal transcription factors, DNA replication, mismatch repair, and oxidative phosphorylation were markedly activated in C3 subtype (Figure 3F). Above data indicated the distinct molecular mechanisms across RBP-based subtypes.

Distinct PANoptosis features across RBP-based subtypes

Next, we focused on PANoptosis features in heart failure. Deregulation of PANoptosis (apoptosis, necroptosis and pyroptosis) genes was found in heart failure LV myocardium specimens *versus* nonfailing controls (Figures 4A–C), indicating that PANoptosis might be linked to heart failure. In addition, PANoptosis features were assessed across three RBP-based subtypes. The widespread heterogeneity in PANoptosis (apoptosis, necroptosis and pyroptosis) genes was observed across RBP-based subtypes (Figures 4D–F).

Identification of RBP-based subtype-related genes

WGCNA approach was utilized for determining RBP-based subtype-related genes. Hierarchical clustering analyses demonstrated no outlier specimens (Figure 5A). Soft-thresholding power is an import process of WGCNA. To establish a scale-free co-expression network, soft-thresholding power was set as 9 following scale independence and mean connectivity (Figure 5B). Genes were clustered through dynamic tree cut approach to obtain 16 modules (Figure 5C). Associations between modules and RBP-based subtypes were then evaluated. As a result, blue module presented the strongest correlation to RBP-based subtypes (Figure 5D). Additionally, we found the remarkable correlation between module membership of blue module and gene significance for RBP-based subtypes (Figure 5E). Thus, 1,460 genes in blue module were regarded as RBP-based subtype-related genes. Their biological functions and pathways were then probed. In Figure 5F, RBP-based subtype-related genes were remarkably linked to biosynthetic process. Also, they presented the associations with cellular components such as nuclear part, protein-containing complex, cytosol (Figure 5G). In Figure 5H, they possessed the molecular functions of catalytic activity acting on a protein, sequence-specific DNA binding, and double-stranded DNA binding, etc. RNA transport, mRNA surveillance pathway, nucleotide excision repair, proteasome and DNA replication were remarkably enriched by RBP-based subtype-related genes (Figure 5I). Above data proved their key roles in pathophysiologic processes of heart failure. Through MCODE approach, 31 key RBP-based subtype-related genes were selected, comprising KRR1, DNTTIP2, NGDN, DDX52, RPF1, FTSJ3, RRS1, GTPBP4, RIOK2, DDX5, MPHOSPH10, PAK1IP1, NOC3L, BRIX1, FCF1, DHX15, RRP7A, EBNA1BP2, WDR3, LSG1, DDX10, POLR1D, GNL2, DDX51, RSL24D1, MAK16, GRWD1, RRP15, UTP3, BCCIP, RSL1D1 (Figure 5J). Most key RBP-based subtype-related genes presented the down-regulation in failing than nonfailing control LV myocardium specimens (Figure 5K).

Identification of characteristic genes for heart failure *via* machine learning analyses

Two machine learning approaches LASSO and SVM-RPE were employed for selecting characteristic genes among RBP-based subtype-related genes. 17 and 10 characteristic genes were separately selected by LASSO (Figures 6A,B) and SVM-RPE (Figure 6C) methods. After intersection, ten characteristic genes were finally determined, including DDX52, DHX15, EBNA1BP2, FCF1, GNL2, GRWD1, LSG1, POLR1D, RRS1, and RSL24D1 (Figure 6D). C3 subtype presented the highest expressions of above characteristic genes, followed by C1 and C2 (Figure 6E). To assess the predictive efficacy of characteristic genes, ROC curves were plotted. As illustrated in Figures 6F–O, the AUC values (95%CI) of DDX52, DHX15, EBNA1BP2, FCF1, GNL2, GRWD1, LSG1, POLR1D, RRS1, and RSL24D1 were 0.68 (0.73–0.63), 0.67 (0.72–0.62), 0.64 (0.69–0.58), 0.57 (0.63–0.52), 0.64 (0.70–0.59), 0.70 (0.75–0.66), 0.68 (0.73–0.63), 0.60 (0.66–0.55), 0.73 (0.78–0.67), 0.56 (0.62–0.51), proving the excellent performance in diagnosing heart failure.

Establishment of a characteristic gene-based nomogram for heart failure

To facilitate the clinical performance of characteristic genes, a nomogram was established for heart failure (Figure 7A). As illustrated in calibration curve, the nomogram-predicted risk probabilities were close to the actual probabilities of heart failure (Figure 7B). Decision curve analyses demonstrated that the nomogram possessed the preferred prediction efficacy, with the higher net benefit (Figure 7C). Clinical impact curves were drawn for evaluating clinical applicability of the risk predictive nomogram. As illustrated in Figure 7D, the nomogram showed the superior overall net benefit within the wide and practical ranges of threshold probabilities and influenced patients' outcome, indicating that the nomogram possessed excellent predictive performance. Above data proved that the nomogram was clinically useful. Moreover, the AUC value (95%CI) of the nomogram was 0.84 (0.88–0.80) (Figure 7E),

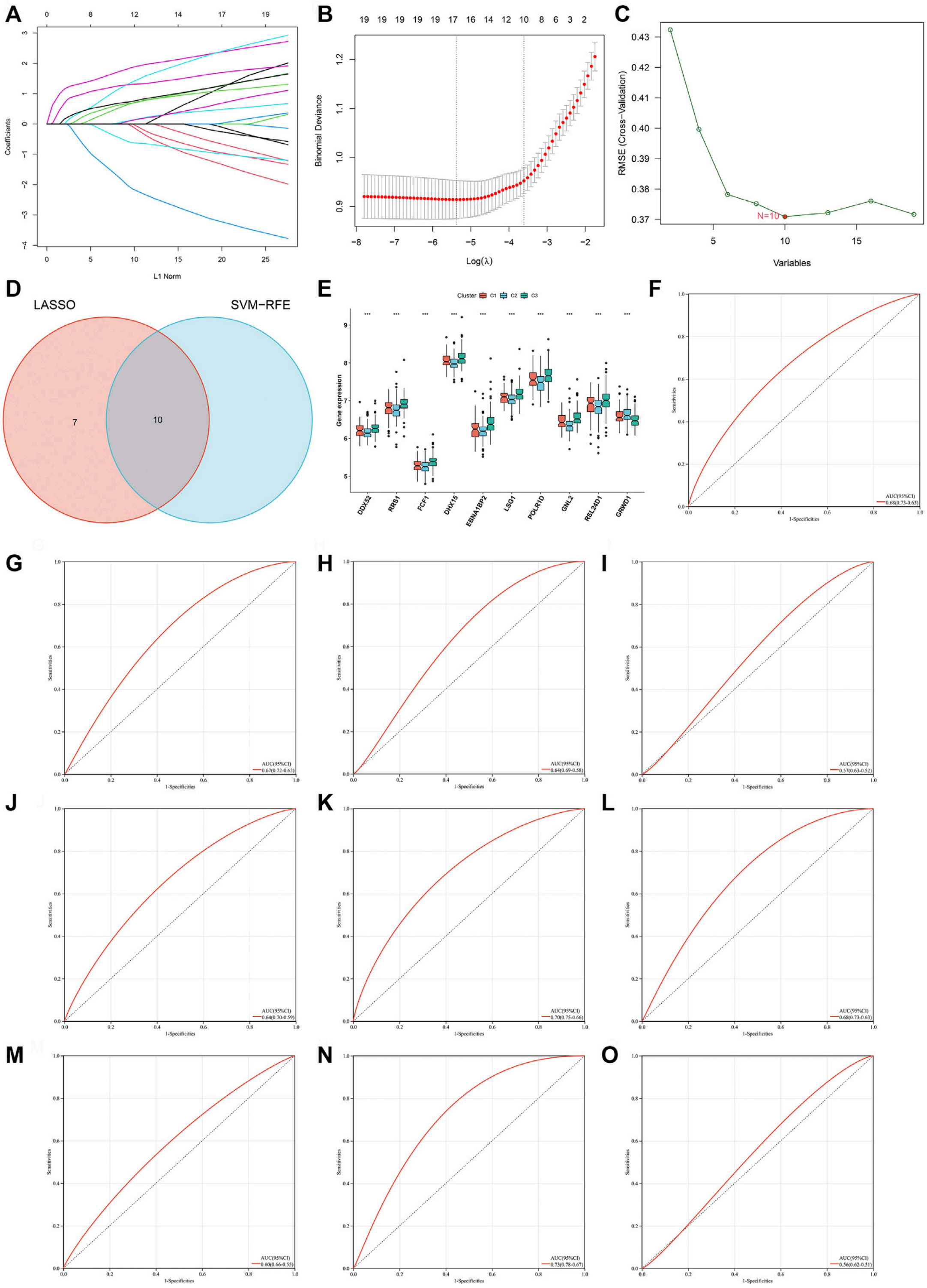


FIGURE 6
Identification of characteristic genes for heart failure *via* machine learning analyses. **(A)** Relationships between log-transformed lambda and regression coefficients. Each line indicates a variable. **(B)** LASSO regression profiling. The line represents 95% CI, and the dotted line represents the optimal number of variables. **(C)** Selection of characteristic genes through SVM-RPE approach. **(D)** Venn plot of characteristic genes shared by LASSO and SVM-RPE approaches. **(E)** Box plot of the expression of characteristic genes across three RBP-based subtypes (***$p < 0.001$). **(F–O)** ROC curves for assessing the predictive efficacy of characteristic genes: **(F)** DDX52, **(G)** DHX15, **(H)** EBNA1BP2, **(I)** FCF1, **(J)** GNL2, **(K)** GRWD1, **(L)** LSG1, **(M)** POLR1D, **(N)** RRS1, and **(O)** RSL24D1.

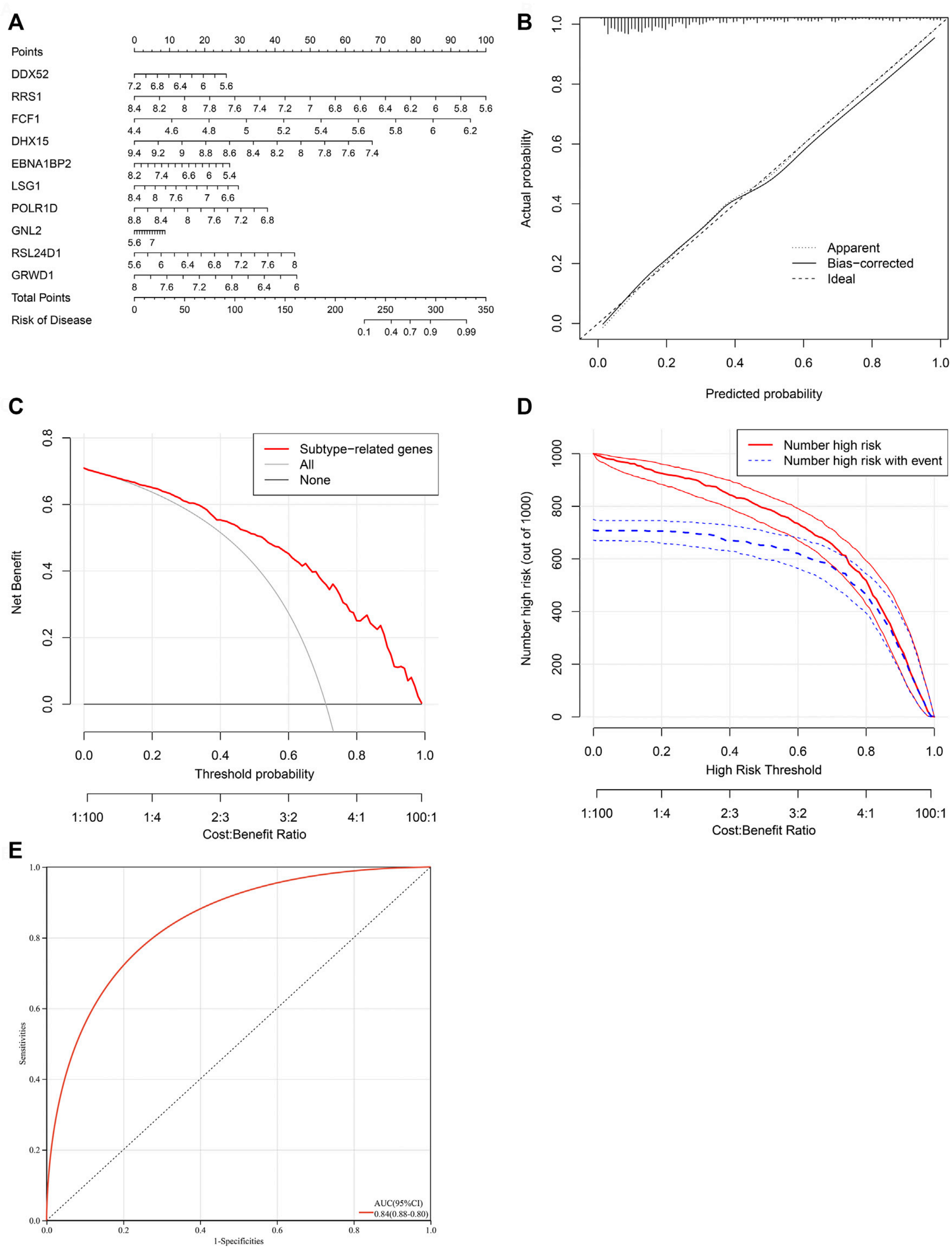


FIGURE 7
Establishment of a characteristic gene-based nomogram for heart failure. **(A)** A nomogram comprising characteristic genes for predicting risk of heart failure. **(B)** Calibration curve for actual and nomogram-predicted probability of heart failure. **(C)** Decision curve analyses for the net benefit curves of the nomogram. X-axis shows the threshold probability for heart failure and Y-axis represents the net benefit. **(D)** Clinical impact curves of the nomogram. Red curves indicate the number of patients classified as positive (high risk) by the nomogram at different threshold probabilities. Blue curves show the number of true positives at different threshold probabilities. **(E)** ROC curve for estimating the predictive efficacy of the nomogram.

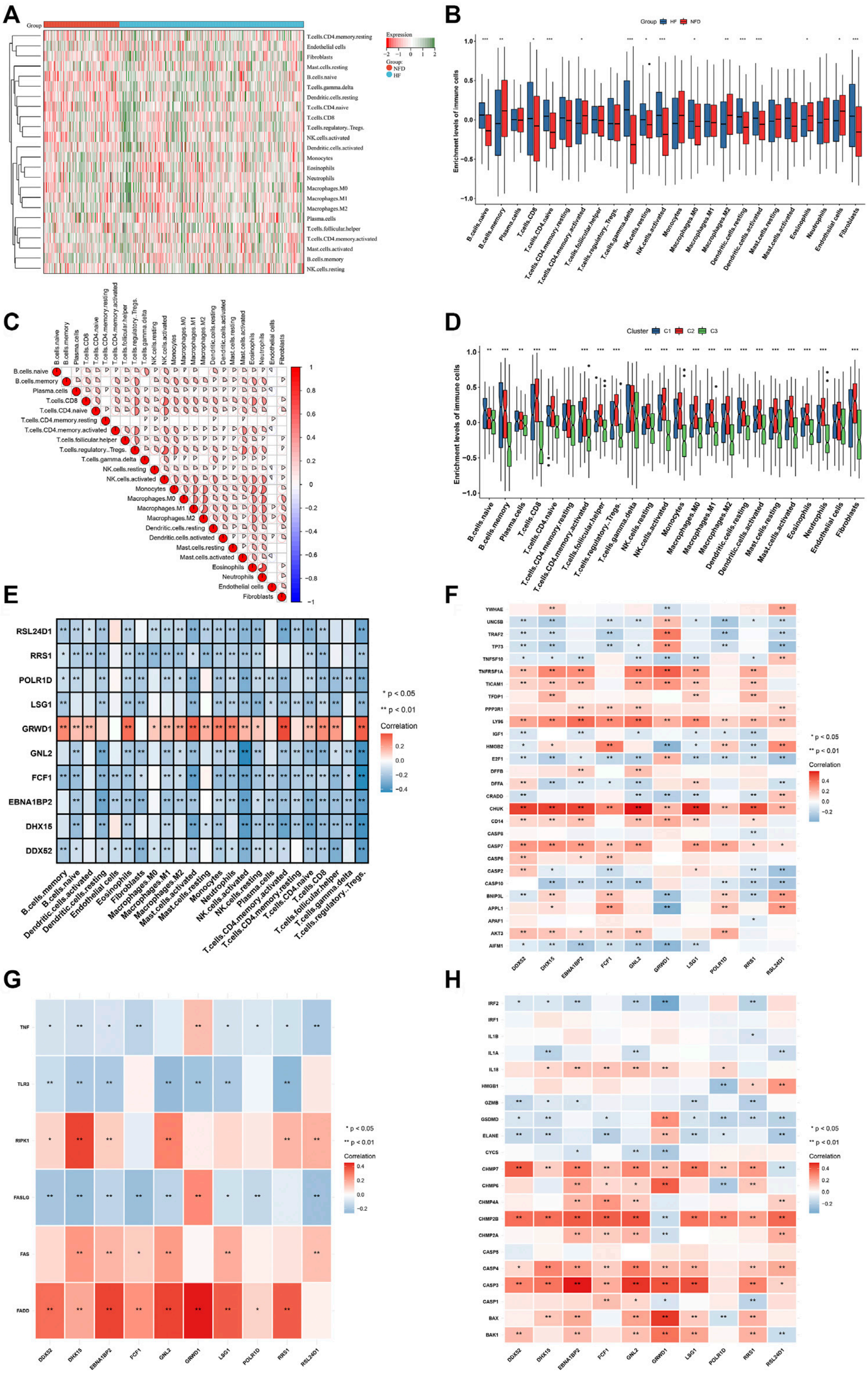


FIGURE 8
Landscape of immune cells and PANoptosis features in heart failure. **(A,B)** The abundance of 24 immune cell types in human failing (blue) and nonfailing control (red) LV myocardium specimens. **(C)** Associations between distinct immune cell types. **(D)** The abundance of 24 immune cell types across three RBP-based subtypes. **(E)** Associations between 24 immune cell types and characteristic genes. **(F–H)** Associations of characteristic genes with apoptosis, necroptosis and pyroptosis genes. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

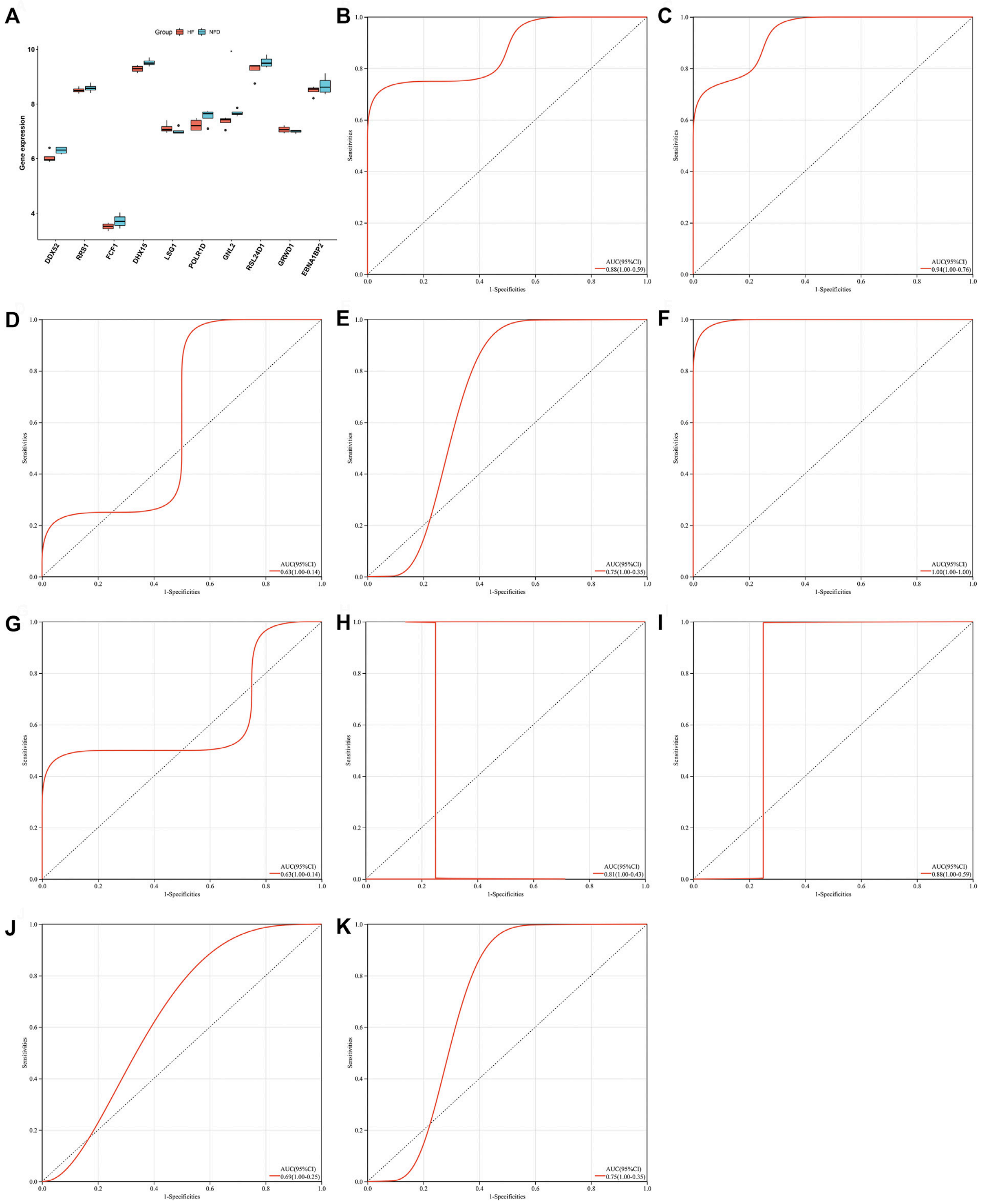


FIGURE 9
External verification of characteristic genes in heart failure in the GSE76701 dataset. **(A)** Box plot of the expressions of characteristic genes in human failing (red) and nonfailing control (blue) LV myocardium specimens (*$p < 0.05$). **(B–K)** Evaluation of diagnostic performance of **(B)** DDX52, **(C)** DHX15, **(D)** EBNA1BP2, **(E)** FCF1, **(F)** GNL2, **(G)** GRWD1, **(H)** LSG1, **(I)** POLR1D, **(J)** RRS1, and **(K)** RSL24D1 in heart failure through ROC curves.

which was higher than any one of characteristic genes, demonstrating that the predictive efficacy of the nomogram was better compared with a single characteristic gene.

Landscape of immune cells and PANoptosis features in heart failure

CIBERSORT was employed for estimating the abundance of 24 immune cell types. Firstly, the abundance of immune cell types was compared between human failing and nonfailing control LV myocardium specimens. As illustrated in Figures 8A,B, failing myocardium tissues presented the enhanced abundance of B cell naïve, T cells CD8, T cells CD4 naïve, T cells gamma delta, NK cells resting/activated, macrophages M0, dendritic cells resting/activated, and fibroblasts. Meanwhile, the reduced abundance of B cells memory, T cells CD4 memory activated, macrophages M2, eosinophils, endothelial cells was found in failing myocardium. Additionally, the positive interactions across immune cell types were found, as illustrated in Figure 8C. We also assessed the differences in immune cell types across three RBP-based subtypes. Generally, C2 had the highest abundance of most immune cells, followed by C1 and C3 (Figure 8D). Figure 8E depicted the associations between characteristic genes and immune cell infiltration. Most characteristic genes were negatively linked to the abundance of immune cells, but GRWD1 presented the positive interactions with most immune cell types. In addition, characteristic genes exhibited notable associations with PANoptosis (apoptosis, necroptosis and pyroptosis) genes (Figures 8F–H).

External verification of characteristic genes in heart failure

Characteristic genes in heart failure were externally verified in independent datasets. In Figure 9A, DDX52, RRS1, FCF1, DHX15, POLR1D, GNL2, RSL24D1, and EBNA1BP2 presented the low expressions in failing than nonfailing control heart. Inversely, LSG1, and GRWD1 expressions were up-regulated in failing compared with nonfailing control heart in the GSE76701 dataset. The abnormal expression of characteristic genes between human failing and nonfailing control LV myocardium tissues was confirmed in the GSE55296 dataset (Supplementary Figure S1). ROC curves were conducted for evaluating the diagnostic efficacy of above characteristic genes in heart failure in the GSE76701 dataset. The AUC values (95% CI) of DDX52, DHX15, EBNA1BP2, FCF1, GNL2, GRWD1, LSG1, POLR1D, RRS1, and RSL24D1 were 0.88 (1.00–0.59), 0.94 (1.00–0.76), 0.63 (1.00–0.14), 0.75 (1.00–0.35), 1.00 (1.00–1.00), 0.63 (1.00–0.14), 0.81 (1.00–0.43), 0.88 (1.00–0.59), 0.69 (1.00–0.25), and 0.75 (1.00–0.35), as illustrated in Figures 9B–K. However, no significant differences in characteristic genes were observed between HFpEF and HFrEF heart tissues in the GSE86569 dataset (Supplementary Figure S2). In addition, the GSE3585 dataset was adopted to validate the diagnostic efficacy of the characteristic genes. The excellent diagnostic performance of each characteristic gene was proven, as shown in Supplementary Figures S3A–H. Above evidence confirmed that characteristic genes enabled to excellently diagnose heart failure.

Prediction of potential therapeutic compounds of heart failure

Based on heart failure-specific RBPs, CMap analysis was adopted to screen potential compounds for the treatment of heart failure with $p < 0.05$. In accordance with MoA analysis, mebendazole, NPI-2358, vindesine, vincristine, flubendazole, vinorelbine, nocodazole, and ABT-751 shared tubulin inhibitor (Figure 10). GSK-3-inhibitor-IX, SB-415286, and SB-216763 shared glycogen synthase kinase inhibitor. Roscovitine and kenpaullone shared CDK inhibitor. GSK-3-inhibitor-II and PKCbeta-inhibitor shared PKC inhibitor.

Discussion

RBPs have been described to be expressed and modulated in a variety of organs especially human heart (Gupta et al., 2018). Despite this, little is known concerning the roles of RBPs in heart failure. Thus, our study implemented comprehensive analyses of RBPs in heart failure, and determined RBP-based subtypes, and RBP-based subtype-related characteristic genes, unveiling the crucial functions of RBPs in heart failure.

RBPs are crucial effectors of gene expression, and as such their abnormal expressions underlie the origin of heart failure (Gebauer et al., 2021). On the basis of the merged GSE5406, and GSE57338 expression profiling datasets, we determined five RBPs (EIF1AY, RPS4Y1, DDX3Y, RNASE2, and CSDC2) with adjusted $p < 0.05$ and |fold-change|>1.5 in heart failure LV myocardium specimens than nonfailing controls, which were regarded as hear failure-specific RBPs. RBPs with adjusted $p < 0.05$ were closely linked to RNA metabolism processes (RNA/mRNA/peptide/ncRNA metabolic processes, RNA splicing, and translation) as well as pathways (RNA transport, mRNA surveillance pathway, ribosome biogenesis in eukaryotes, aminoacyl-tRNA biosynthesis, RNA degradation, etc.), highlighting the crucial functions of RBPs in controlling gene expression. Evidence has demonstrated that deregulation of RNA metabolism leads to heart failure progression (Kim et al., 2018). On the basis of RBPs with adjusted $p < 0.05$, three RBP-based subtypes were established, characterized by distinct signaling pathway activities. Additionally, RBP-based subtype-related

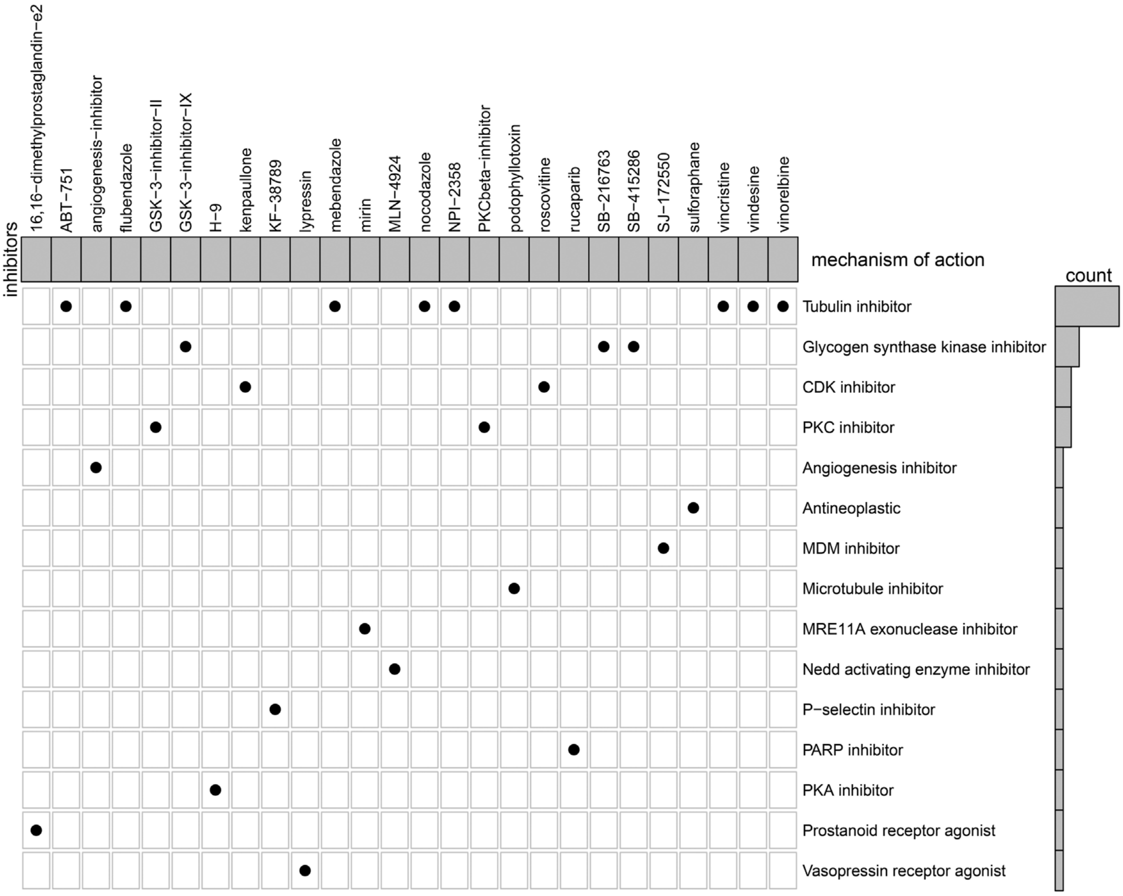


FIGURE 10
MoA analysis for the shared mechanisms of potential therapeutic compounds of heart failure.

genes were further determined, which might be modulated by RBPs.

Previous studies have determined heart-specific RBPs (RBM20, RBM24, HuR, etc.) that were not included in our heart failure-specific RBPs. For instance, suppressing RBM20 activity may improve diastolic dysfunction and cardiac atrophy (Hinze et al., 2016). RBM24 loss destroys global alternative splicing and contributes to dilated cardiomyopathy (Liu et al., 2019). HuR-induced SCN5A mRNA stability decreases arrhythmic risk in heart failure (Zhou et al., 2018). Thus, our study offered novel heart-specific RBPs. More experiments are awaited to validate the biological functions of heart failure.

LASSO is a regression analysis approach that utilizes regularization for improving the predictive accuracy (Xing et al., 2022). SVM-RFE is a reliable feature selection approach that determines the optimal variables through removing the feature vectors produced by SVM (Zhao et al., 2020). Through integrating two machine learning approaches, ten characteristic genes were eventually determined, comprising DDX52, DHX15, EBNA1BP2, FCF1, GNL2, GRWD1, LSG1, POLR1D, RRS1, and RSL24D1. All of them accurately predicted the risk of heart failure. Further, a characteristic gene-based nomogram was established, which was capable of accurately predicting heart failure risk, with the excellent clinical usability. DDX52 is a type of DEAD/H box RNA helicase, and its suppression exerts an anti-tumor effect (Yu et al., 2021). The DEAH-box RNA helicase DHX15 has been identified as a potential gene for pathological cardiac hypertrophy triggered by excessive exercise (Zhou et al., 2020) and pulmonary arterial hypertension (Wang et al., 2021). EBNA1BP2 functions as a dynamic scaffold for ribosome biogenesis (Hirano et al., 2009). FCF1 is a potential marker of circulating breast cancer cells for detecting metastasis (Fina et al., 2022). The nucleolar GTP-binding protein GNL2 is essential for retinal neurogenesis in developing zebrafish (Paridaen et al., 2011). Cdt1-binding protein GRWD1 acts as a histone-binding protein, which triggers MCM loading *via* influencing chromatin architecture (Sugimoto et al., 2015). LSG1 is a family member of essential GTPases, in relation to the evolution of compartmentalization

(Reynaud et al., 2005). POLR1D is a component of RNA polymerase I and RNA polymerase III complexes, mediating the synthesis of ribosomal RNA precursor and small RNA (Sanchez et al., 2020). RRS1 is a key factor of 5 S rRNA binding activity (Kharde et al., 2015). RSL24D1 participates in the biogenesis of the 60 S ribosomal subunit (Ni et al., 2022).

Heart failure is typically linked to cardiac remodeling, and inflammatory response plays a crucial role. During cardiac inflammation, immune cells invade the cardiac tissue as well as modulate tissue-damaging response (Bacmeister et al., 2019). In the present study, failing myocardium tissues exhibited the enhanced abundance of B cell naïve, T cells CD8, T cells CD4 naïve, T cells gamma delta, NK cells resting/activated, macrophages M0, dendritic cells resting/activated, and fibroblasts. Additionally, the decreased abundance of B cells memory, T cells CD4 memory activated, macrophages M2, eosinophils, endothelial cells was observed in failing myocardium. For example, CD8^{+} effector T cells may prevent cardioprotective macrophage differentiation in early heart failure (Komai et al., 2021). Posttranscriptional control of mRNA modulates inflammatory and immune responses. Several RBPs have been extensively explored, and bind target mRNAs to enhance or dampen above activities (Akira and Maeda, 2021). RBP-based C2 subtype presented the highest abundance of most immune cells, followed by C1 and C3. In addition, there was the extensive heterogeneity in PANoptosis traits across three RBP-based subtypes. Most characteristic genes presented negative correlations to the abundance of immune cells in heart failure, but GRWD1 was positively linked to most immune cell types, indicating their functions in mediating cardiac inflammation. Among them, evidence demonstrates that DHX15 may sense double-stranded RNA in myeloid dendritic cells to activate the immune response to RNA (Lu et al., 2014). Co-expression network analyses have determined DHX15 RNA helicase as a regulator of B cells (Detanico et al., 2019). DHX15 is a crucial regulator of natural killer-cell homeostasis and function (Wang et al., 2022). Additionally, characteristic genes exhibited remarkable interactions to PANoptosis features across heart failure. More experiments are required for verifying their regulatory functions in inflammatory and immune responses as well as PANoptosis in heart failure.

Conclusion

Collectively, our findings provided an overview of RBPs involved in heart failure. Three RBP-based subtypes as well as ten relevant characteristic genes were determined for heart failure, elucidating the critical roles of RBPs in pathophysiological process (especially immunity and PANoptosis) of heart failure as well as offering reliable targets for diagnosing heart failure risk. Despite this, in-depth research is required for verifying the effectiveness of RBP-based subtypes and characteristic genes in diagnostic utility of heart failure.

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

PC, ZW and WZ conceived and designed the study. JL, XZ and PR conducted most of the experiments and data analysis, and wrote the manuscript. YW and YW participated in collecting data and helped to draft the manuscript. All authors reviewed and approved the manuscript.

Funding

This research was supported by Natural Science Foundation of Xinjiang Uygur Autonomous Region (2020D01C088), Fund Project the People's Hospital of Xinjiang Uygur Autonomous Region (20200109).

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

Supplemenetary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1004163/full#supplementary-material

References

Akira, S., and Maeda, K. (2021). Control of RNA stability in immunity. *Annu. Rev. Immunol.* 39, 481–509. doi:10.1146/annurev-immunol-101819-075147

Bacmeister, L., Schwarzl, M., Warnke, S., Stoffers, B., Blankenberg, S., Westermann, D., et al. (2019). Inflammation and fibrosis in murine models of heart failure. *Basic Res. Cardiol.* 114 (3), 19. doi:10.1007/s00395-019-0722-5

Berg, D. D., Jhund, P. S., Docherty, K. F., Murphy, S. A., Verma, S., Inzucchi, S. E., et al. (2021). Time to clinical benefit of dapagliflozin and significance of prior heart failure hospitalization in patients with heart failure with reduced ejection fraction. *JAMA Cardiol.* 6 (5), 499–507. doi:10.1001/jamacardio.2020.7585

Chothani, S., Schäfer, S., Adami, E., Viswanathan, S., Widjaja, A. A., Langley, S. R., et al. (2019). Widespread translational control of fibrosis in the human heart by RNA-binding proteins. *Circulation* 140 (11), 937–951. doi:10.1161/circulationaha.119.039596

Detanico, T., Virgen-Slane, R., Steen-Fuentes, S., Lin, W. W., Rhode-Kurnow, A., Chappell, E., et al. (2019). Co-Expression networks identify DHX15 RNA helicase as a B cell regulatory factor. *Front. Immunol.* 10, 2903. doi:10.3389/fimmu.2019.02903

Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2019). Cytoscape StringApp: Network analysis and visualization of proteomics data. *J. Proteome Res.* 18 (2), 623–632. doi:10.1021/acs.jproteome.8b00702

Engebretsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. *Clin. Epigenetics* 11 (1), 123. doi:10.1186/s13148-019-0730-1

Fina, E., Cleris, L., Dugo, M., Lecchi, M., Ciniselli, C. M., Lecis, D., et al. (2022). Gene signatures of circulating breast cancer cell models are a source of novel molecular determinants of metastasis and improve circulating tumor cell detection in patients. *J. Exp. Clin. Cancer Res.* 41 (1), 78. doi:10.1186/s13046-022-02259-8

Gao, G., Chen, W., Yan, M., Liu, J., Luo, H., Wang, C., et al. (2020). Rapamycin regulates the balance between cardiomyocyte apoptosis and autophagy in chronic heart failure by inhibiting mTOR signaling. *Int. J. Mol. Med.* 45 (1), 195–209. doi:10.3892/ijmm.2019.4407

Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* 22 (3), 185–198. doi:10.1038/s41576-020-00302-y

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15 (12), 829–845. doi:10.1038/nrg3813

Gupta, S. K., Garg, A., Bär, C., Chatterjee, S., Foinquinos, A., Milting, H., et al. (2018). Quaking inhibits doxorubicin-mediated cardiotoxicity through regulation of cardiac circular RNA expression. *Circ. Res.* 122 (2), 246–254. doi:10.1161/circresaha.117.311335

Hannenhalli, S., Putt, M. E., Gilmore, J. M., Wang, J., Parmacek, M. S., Epstein, J. A., et al. (2006). Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation* 114 (12), 1269–1276. doi:10.1161/circulationaha.106.632430

Heidenreich, P. A., Bozkurt, B., Aguilar, D., Allen, L. A., Byun, J. J., Colvin, M. M., et al. (2022a). 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American college of cardiology/American heart association joint committee on clinical practice guidelines. *Circulation* 145 (18), e895–e1032. doi:10.1161/cir.0000000000001063

Heidenreich, P. A., Bozkurt, B., Aguilar, D., Allen, L. A., Byun, J. J., Colvin, M. M., et al. (2022b). 2022 AHA/ACC/HFSA guideline for the management of heart failure: A report of the American college of cardiology/American heart association joint committee on clinical practice guidelines. *J. Am. Coll. Cardiol.* 79 (17), e263–e421. doi:10.1016/j.jacc.2021.12.012

Hinze, F., Dieterich, C., Radke, M. H., Granzier, H., and Gotthardt, M. (2016). Reducing RBM20 activity improves diastolic dysfunction and cardiac atrophy. *J. Mol. Med.* 94 (12), 1349–1358. doi:10.1007/s00109-016-1483-3

Hirano, Y., Ishii, K., Kumeta, M., Furukawa, K., Takeyasu, K., and Horigome, T. (2009). Proteomic and targeted analytical identification of BXDC1 and EBNA1BP2 as dynamic scaffold proteins in the nucleolus. *Genes cells.* 14 (2), 155–166. doi:10.1111/j.1365-2443.2008.01262.x

Kharde, S., Calviño, F. R., Gumiero, A., Wild, K., and Sinning, I. (2015). The structure of Rpf2-Rrs1 explains its role in ribosome biogenesis. *Nucleic Acids Res.* 43 (14), 7083–7095. doi:10.1093/nar/gkv640

Kim, E. H., Galchev, V. I., Kim, J. Y., Misek, S. A., Stevenson, T. K., Campbell, M. D., et al. (2016). Differential protein expression and basal lamina remodeling in human heart failure. *Proteomics. Clin. Appl* 10 (5), 585–596. doi:10.1002/prca.201500099

Kim, G. H., Uriel, N., and Burkhoff, D. (2018). Reverse remodelling and myocardial recovery in heart failure. *Nat. Rev. Cardiol.* 15 (2), 83–96. doi:10.1038/nrcardio.2017.139

Komai, K., Ito, M., Nomura, S., Shichino, S., Katoh, M., Yamada, S., et al. (2021). Single-cell analysis revealed the role of CD8(+) effector T cells in preventing cardioprotective macrophage differentiation in the early phase of heart failure. *Front. Immunol.* 12, 763647. doi:10.3389/fimmu.2021.763647

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

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

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, J., Kong, X., Zhang, M., Yang, X., and Xu, X. (2019). RNA binding protein 24 deletion disrupts global alternative splicing and causes dilated cardiomyopathy. *Protein Cell* 10 (6), 405–416. doi:10.1007/s13238-018-0578-8

Liu, R., Yang, J., Yao, J., Zhao, Z., He, W., Su, N., et al. (2022). Optogenetic control of RNA function and metabolism using engineered light-switchable RNA-binding proteins. *Nat. Biotechnol.* 40 (5), 779–786. doi:10.1038/s41587-021-01112-1

Liu, Y., Morley, M., Brandimarto, J., Hannenhalli, S., Hu, Y., Ashley, E. A., et al. (2015). RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* 105 (2), 83–89. doi:10.1016/j.ygeno.2014.12.002

Lu, H., Lu, N., Weng, L., Yuan, B., Liu, Y. J., and Zhang, Z. (2014). DHX15 senses double-stranded RNA in myeloid dendritic cells. *J. Immunol.* 193 (3), 1364–1372. doi:10.4049/jimmunol.1303322

Maatz, H., Jens, M., Liss, M., Schafer, S., Heinig, M., Kirchner, M., et al. (2014). RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *J. Clin. Invest.* 124 (8), 3419–3430. doi:10.1172/jci74523

Mascolo, A., di Mauro, G., Cappetta, D., De Angelis, A., Torella, D., Urbanek, K., et al. (2022). Current and future therapeutic perspective in chronic heart failure. *Pharmacol. Res.* 175, 106035. doi:10.1016/j.phrs.2021.106035

Methawasin, M., Strom, J. G., Slater, R. E., Fernandez, V., Saripalli, C., and Granzier, H. (2016). Experimentally increasing the compliance of titin through RNA binding motif-20 (RBM20) inhibition improves diastolic function in a mouse model of heart failure with preserved ejection fraction. *Circulation* 134 (15), 1085–1099. doi:10.1161/circulationaha.116.023003

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

Ni, C., Schmitz, D. A., Lee, J., Pawłowski, K., Wu, J., and Buszczak, M. (2022). Labeling of heterochronic ribosomes reveals C1ORF109 and SPATA5 control a late step in human ribosome assembly. *Cell Rep.* 38 (13), 110597. doi:10.1016/j.celrep.2022.110597

Packer, M., Anker, S. D., Butler, J., Filippatos, G., Ferreira, J. P., Pocock, S. J., et al. (2021a). Influence of neprilysin inhibition on the efficacy and safety of empagliflozin in patients with chronic heart failure and a reduced ejection fraction: The EMPEROR-reduced trial. *Eur. Heart J.* 42 (6), 671–680. doi:10.1093/eurheartj/ehaa968

Packer, M., Butler, J., Zannad, F., Filippatos, G., Ferreira, J. P., Pocock, S. J., et al. (2021b). Effect of empagliflozin on worsening heart failure events in patients with heart failure and preserved ejection fraction: EMPEROR-preserved trial. *Circulation* 144 (16), 1284–1294. doi:10.1161/circulationaha.121.056824

Paridaen, J. T., Janson, E., Utami, K. H., Pereboom, T. C., Essers, P. B., van Rooijen, C., et al. (2011). The nucleolar GTP-binding proteins Gnl2 and nucleostemin are required for retinal neurogenesis in developing zebrafish. *Dev. Biol.* 355 (2), 286–301. doi:10.1016/j.ydbio.2011.04.028

Piepoli, M. F., Adamo, M., Barison, A., Bestetti, R. B., Biegus, J., Böhm, M., et al. (2022). Preventing heart failure: A position paper of the heart failure association in collaboration with the European association of preventive cardiology. *Eur. J. Heart Fail.* 24 (1), 143–168. doi:10.1002/ejhf.2351

Reynaud, E. G., Andrade, M. A., Bonneau, F., Ly, T. B., Knop, M., Scheffzek, K., et al. (2005). Human Lsg1 defines a family of essential GTPases that correlates with the evolution of compartmentalization. *BMC Biol.* 3, 21. doi:10.1186/1741-7007-3-21

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

Sanchez, E., Laplace-Builhé, B., Mau-Them, F. T., Richard, E., Goldenberg, A., Toler, T. L., et al. (2020). POLR1B and neural crest cell anomalies in Treacher Collins syndrome type 4. *Genet. Med.* 22 (3), 547–556. doi:10.1038/s41436-019-0669-9

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

Sugimoto, N., Maehara, K., Yoshida, K., Yasukouchi, S., Osano, S., Watanabe, S., et al. (2015). Cdt1-binding protein GRWD1 is a novel histone-binding protein that facilitates MCM loading through its influence on chromatin architecture. *Nucleic Acids Res.* 43 (12), 5898–5911. doi:10.1093/nar/gkv509

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49 (D1), D605–d612. doi:10.1093/nar/gkaa1074

Tarazón, E., Roselló-Lletí, E., Rivera, M., Ortega, A., Molina-Navarro, M. M., Triviño, J. C., et al. (2014). RNA sequencing analysis and atrial natriuretic peptide production in patients with dilated and ischemic cardiomyopathy. *PLoS One* 9 (3), e90157. doi:10.1371/journal.pone.0090157

Wang, G., Xiao, X., Wang, Y., Chu, X., Dou, Y., Minze, L. J., et al. (2022). The RNA helicase DHX15 is a critical regulator of natural killer-cell homeostasis and functions. *Cell. Mol. Immunol.* 19, 687–701. doi:10.1038/s41423-022-00852-7

Wang, W., Jiang, Z., Zhang, D., Fu, L., Wan, R., and Hong, K. (2021). Comparative transcriptional analysis of pulmonary arterial hypertension associated with three different diseases. *Front. Cell Dev. Biol.* 9, 672159. doi:10.3389/fcell.2021.672159

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

Xing, Q., Jiaochen, L., Shouyong, L., Limin, M., and Wang, Y. (2022). Six RNA binding proteins (RBPs) related prognostic model predicts overall survival for clear cell renal cell carcinoma and is associated with immune infiltration. *Bosn. J. Basic Med. Sci.* 22 (3), 435–452. doi:10.17305/bjbms.2021.6097

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, W., Ma, H., Li, J., Ge, J., Wang, P., Zhou, Y., et al. (2021). DDX52 knockdown inhibits the growth of prostate cancer cells by regulating c-Myc signaling. *Cancer Cell Int.* 21 (1), 430. doi:10.1186/s12935-021-02128-y

Zeng, C., Wang, R., and Tan, H. (2019). Role of pyroptosis in cardiovascular diseases and its therapeutic implications. *Int. J. Biol. Sci.* 15 (7), 1345–1357. doi:10.7150/ijbs.33568

Zhang, T., Zhang, Y., Cui, M., Jin, L., Wang, Y., Lv, F., et al. (2016). CaMKII is a RIP3 substrate mediating ischemia- and oxidative stress-induced myocardial necroptosis. *Nat. Med.* 22 (2), 175–182. doi:10.1038/nm.4017

Zhao, E., Xie, H., and Zhang, Y. (2020). Predicting diagnostic gene biomarkers associated with immune infiltration in patients with acute myocardial infarction. *Front. Cardiovasc. Med.* 7, 586871. doi:10.3389/fcvm.2020.586871

Zhou, A., Xie, A., Kim, T. Y., Liu, H., Shi, G., Kang, G. J., et al. (2018). HuR-mediated SCN5A messenger RNA stability reduces arrhythmic risk in heart failure. *Heart rhythm.* 15 (7), 1072–1080. doi:10.1016/j.hrthm.2018.02.018

Zhou, Z., Zheng, L., Tang, C., Chen, Z., Zhu, R., Peng, X., et al. (2020). Identification of potentially relevant genes for excessive exercise-induced pathological cardiac hypertrophy in zebrafish. *Front. Physiol.* 11, 565307. doi:10.3389/fphys.2020.565307

Frontiers in Genetics

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fgene.2022.1020757

Exploration the global single-cell ecological landscape of adenomyosis-related cell clusters by single-cell RNA sequencing

Jiajing Lin[1†], Li Liu[1†], Fengque Zheng[1], Saiqiong Chen[1], Weiwei Yang[1], Jingjing Li[2], Steven Mo[3*] and Ding-Yuan Zeng[2*]

[1]Department of Obstetrics and Gynecology, No. 4th Hospital Affiliated to Guangxi Medical University, Liuzhou, China, [2]Department of Gynecology, Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, China, [3]Department of Basic Science, YuanDong International Academy of Life Sciences, Hong Kong, China

Background: Adenomyosis (AM) is a common benign uterine disease that threatens the normal life of patients. Cells associated with microenvironmental immune ecology are crucial in AM, although they are not as well understood at the cellular level.

Methods: Single-cell sequencing (scRNA-seq) data were used to construct an AM global single-cell map, to further identify relevant cell clusters and infer chromosomal copy number variation (CNV) in AM samples. The biological functions of cell clusters were explored and cellular evolutionary processes were inferred by enrichment analysis and pseudotime analysis. In addition, a gene regulatory network (GRN) analysis was constructed to explore the regulatory role of transcription factors in AM progression.

Results: We obtained the expression profiles of 42260 cells and identified 10 cell clusters. By comparing the differences in cell components between AM patients and controls, we found that significant abundance of endometrial cells (EC), epithelial cells (Ep), endothelial cells (En), and smooth muscle cells (SMC) in AM patients. Cell clusters with high CNV levels possessing tumour-like features existed in the ectopic endometrium samples. Moreover, the Ep clusters were significantly involved in leukocyte transendothelial cell migration and apoptosis, suggesting an association with cell apoptosis and migration. En clusters were mainly involved in pathways in cancer and apoptosis, indicating that En has certain malignant features.

Conclusion: This study identified cell clusters with immune-related features, investigated the changes in the immune ecology of the microenvironment of these cells during AM, and provided a new strategy for the treatment of AM.

KEYWORDS

adenomyosis, single-cell RNA sequencing, malignant cells, immune microenvironment, biological function

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Saitian Zeng,
Cangzhou Central Hospital, China
Gang Deng,
Zhejiang University School of Medicine, China

*CORRESPONDENCE
Steven Mo,
Steven_Mo@ydlife.org
Ding-Yuan Zeng,
zengdingyuan@lmchh.org

†These authors contributed equally to this work

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

RECEIVED 16 August 2022
ACCEPTED 03 October 2022
PUBLISHED 17 October 2022

CITATION
Lin J, Liu L, Zheng F, Chen S, Yang W, Li J, Mo S and Zeng D-Y (2022), Exploration the global single-cell ecological landscape of adenomyosis-related cell clusters by single-cell RNA sequencing.
Front. Genet. 13:1020757.
doi: 10.3389/fgene.2022.1020757

Introduction

Adenomyosis (AM) is a common benign gynecological syndrome, characterized by infiltration of endometrial glands and stroma in the myometrium (Bird et al., 1972). The prevalence of AM ranges from 5% to 70% (Graziano et al., 2015), with an average of 20%–35% of women worldwide suffering from AM (Abbott, 2017). The most common manifestations of AM are dysmenorrhea, infertility, and abnormal uterine bleeding (AUB), but some women with AM are asymptomatic (Peric and Fraser, 2006). From the epidemiologic data, AM can increase the risk of cancers, including ovarian, endometrial, breast, colorectal, and other cancers of women (Borgfeldt and Andolf, 2004; Kok et al., 2015). Furthermore, Bergeron previously reported that the definitive diagnosis of AM was based on the presence of ectopic endometrial tissue in the myometrium (Bergeron et al., 2006), but is now diagnosed by non-invasive techniques such as pelvic imaging (Chapron et al., 2020). Studies have also reported that the standard method of managing the disease is hysterectomy, while most patients desire to preserve their fertility (Stratopoulou et al., 2021). Despite improvements in diagnostic tools, awareness of AM remains poor (Leyendecker et al., 2015).

At the cellular level, the microenvironmental immune cells of AM play an important role. The number of macrophages, natural killer cells, and T cells in the endometrial stroma of AM increased significantly compared with women with mild focal AM or without the disease (Yang et al., 2004; Tremellen and Russell, 2012). Several malignant features exist in the epithelial cells (Ep) of AM, such as high migration capacity, which contribute to disease progression (Liu et al., 2021). Studies have confirmed that the endothelial cells (En) are damaged, and the uterus occurs the symptoms of bleeding, which is also important in adhesion and migration (Kruger-Genge et al., 2019). Furthermore, smooth muscle cells (SMC) have the ability to shrink and diastole, lack of contraction can cause uterine bleeding, which may lead to the occurrence of inflammation (Owens et al., 2004). It may thus be possible to comprehend the emergence of AM by concentrating on the mechanisms of change in cells related to the immunological microenvironment.

Single-cell RNA sequencing (scRNA-seq), an indispensable technique to dissect cellular heterogeneity and analyze cell types, can assist us in thoroughly comprehending the biological roles (Hedlund and Deng, 2018). Numerous effective methods to examine molecular alterations at the cellular level are provided by the scRNA-seq (Tang et al., 2009). Moreover, research has demonstrated that rare clusters of AM were identified by scRNA-seq, confirming that the occurrence originates from endometrial migration (Liu et al., 2021). However, more studies are needed for further validation. In this study, we explored the states and transitions of the immune microenvironment cells of AM from a single-cell perspective. A comprehensive map of the AM single-cell ecosystem was depicted, relevant cell clusters were identified, and chromosomal copy number variation (CNV) was inferred for each AM sample. Furthermore, it was further confirmed the associated cluster of markers was involved in the signaling pathways and gene regulatory networks (GRN), which contributes to our understanding of the functions of the cluster markers in AM and at the cellular level.

Materials and methods

Data sources

The AM scRNA-seq data including SRR12791871, SRR12791872, and SRR12791873 (Liu et al., 2021) were obtained from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). A 50-year-old woman with uterine fibroids, excluding the AM, and the endometrium tissue from this patient were used as a control sample. Moreover, two endometrium tissue samples were obtained from a 46-year-old AM patient and the samples were taken from eutopic endometrium (AM_EM) and ectopic endometrial (AM_EC) tissues.

Data preprocessing and construction of the single-cell atlas

We used the IntegrateData function (Butler et al., 2018) in the Seurat package (Stuart et al., 2019) to merge the scRNA-seq data, and performed cell clustering analysis according to default parameters. Uniform Manifold Approximation and Projection (UMAP) algorithm (Becht et al., 2018) was adopted for dimensionality reduction and visualization and mapped into single cell profiling. Subsequently, the FindAllMarkers function in Seurat package identified the specific marker genes in each cell cluster. Furthermore, the cell types underwent an immune response based on annotation and re-clustering of known marker genes.

Differential gene expression analysis

Differential expression analysis was performed based on the FindMarkers function in the Seurat package (Butler et al., 2018). Differentially expressed genes (DEGs) of different clusters in the Control, AM_EM, or AM_EC groups were identified. DEGs between normal tissues and AM tissues were screened by an adjusted p value <0.05 and |log fold change (log FC)|>0.5 being considered significant.

Evaluation of CNV in single cells

CNVs are primarily used to identify subclones of diseased cells and to infer tumor evolution (Yates and Campbell, 2012). The CNVs of each cell were assessed from the AM patients by the inferCNV package (inferCNV of the Trinity CTAT Project; https://github.com/broadinstitute/inferCNV) (Patel et al., 2014). To calculate the CNVs of AM_EM and AM_EC cells, the average or normal expression of genes from immune cells was applied as a reference and then determine the expression.

Functional enrichment analysis

To explore the biological functions involved in each cell cluster. We performed the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) on the clustered markers using the clusterProfiler package (Yu et al., 2012). $p < 0.05$ was considered statistically significant.

Pseudotime analysis

Pseudotime analysis could determine the dynamics of gene expression within cell types and trajectories over time (Cao et al., 2019), and infer cell evolution during the AM. A "branch" trajectory was constructed based on the Monocle three package (Trapnell et al., 2014) to explore the dysregulated changes in immune cells of AM patients and project the cells into low-dimensional space by UMAP, the parameters of Monocle three package are set to the default.

Construction of GRN

In this study, we constructed a GRN with transcription factors as the core to infer co-expression modules, to further explore the dysregulate regulatory mechanism of immune cellogenesis. Single cell regulatory network inference and clustering (SCENIC) (Aibar et al., 2017; Van de Sande et al., 2020) was used to infer gene regulatory networks based on single-cell expression profiles and identify cell states, providing important biological insights into the mechanism driving cell heterogeneity. Among them, the binding motifs of the transcription factors in the co-expression module were obtained from the JASPAR database (https://jaspar.genereg.net).

Statistical analysis

Statistical analyses were performed using R (https://www.r-project.org/). Expression levels of genes were analyzed using unpaired t-tests. If the $p < 0.05$ that considered statistical significance. The analyses in this study were based on the Bioinforcloud platform (http://www.bioinforcloud.org.cn).

Results

Global single-cell atlas of the adenomyosis

To investigate the early cell population dynamics in AM patients, we analyzed the scRNA-seq data from endometrial tissue samples of AM patients and control donors. The analysis flow of this study is shown in Figure 1A, where we constructed a global single-cell landscape of AM. By cluster analysis, we divided 42,260 cells into 31 cell clusters, which were identified into 11 cell types based on known markers (Figure 1B) (Supplementary Table S1), including endometrial cells (EC), fibroblasts, epithelial cells (Ep), endothelial cells (En), $CD8^{+}$T, $CD4^{+}$T, Naive T, macrophages (Mac), plasmacytoid predendritic cells (pDC), smooth muscle cells (SMC), and innate lymphoid cells (ILC). Among them, each cell marker exhibited a specific expression for the cell cluster (Figure 1C). Chromosomal CNV analysis based on expression patterns at genomic intervals showed the presence of multicopy events in AM in ectopic endometrial samples (Figure 1D). Further comparing the differences in cell composition between control and AM patients, we found that the highest abundance of En and EC was found in the AM_EC and AM_EM groups. However, the fibroblasts in the control group had the highest abundance (Figure 1E). In summary, we delineated the single-cell profiles of AM patients to reveal the differences in microenvironmental cell components in AM patients.

Ecological landscape of adenomyosis-associated endometrial cells clusters

AM occurs mainly in endometrial tissue, which has a higher cell abundance in EC of AM patients, therefore, subsequent studies will focus on this cellular cluster. We obtained 10 EC clusters by cluster analysis (Figure 2A). As shown in Figure 2B, almost all of these cell clusters were present in different groups of AM patients. Further exploration of the abundance of the cellular cluster in AM patients revealed a significant increase in the proportion of EC_TIMP3 cell clusters and a significant decrease in the proportion of EC_ZFAND2A clusters (Figure 2C). Markers for the different clusters of the EC were mapped to the single-cell atlas, including ZFAND2A, KPT17, TIMP3, SPARCL1, PLAAT3, SPINT2, SCGB2A1, RGS5, CXCL2, and COL1A2 (Figure 2D). Furthermore, EC clusters may be associated with cell motility, which was closely associated with focal adhesion and leukocyte transendothelial migration and apoptosis (Figure 2E). Based on the pseudotime trajectory

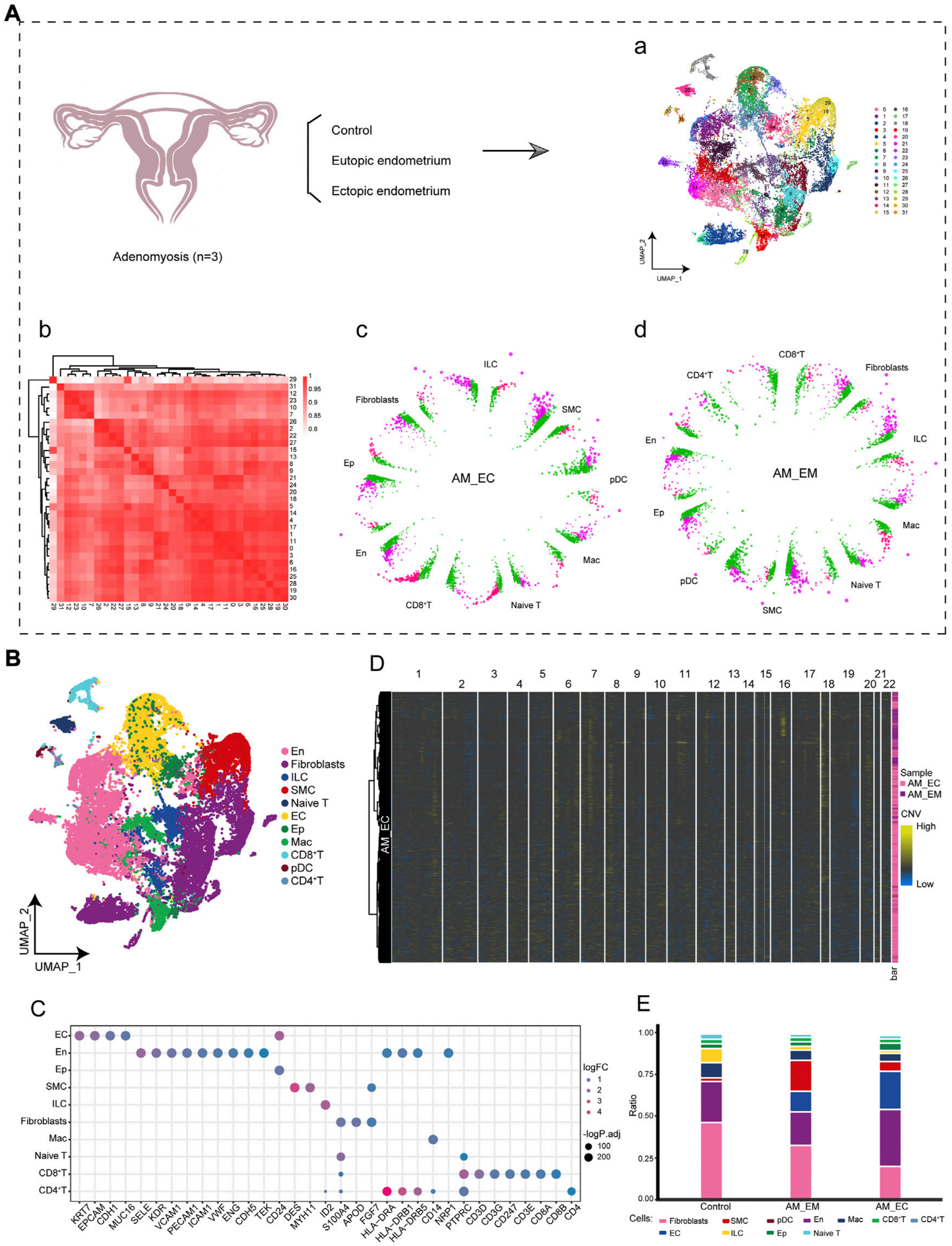


FIGURE 1
Single-cell landscape of adenomyosis (AM) in the immune microenvironment. **(A)**. The workflow of this study analyzed single-cell RNA sequencing (10X Genomics) data between control and two AM between eutopic endometrium (AM_EM) and ectopic endometrium (AM_EC) samples, identifying 31 cell clusters of 42,260 cells. **(B).** The cell types of single cells mapping AM includes endometrial cells (EC), fibroblasts (Fibroblasts), epithelial cells (Ep), endothelial cells (En), CD8 + T, CD4 + T, initial T cells (Naive T), macrophages (Mac), predendritic cells (pDC), smooth muscle cells (SMC), and innate lymphoid cells (ILC). **(C).** Bubble plots show specific marker genes in different cell types. **(D).** Heatmap shows the levels of copy number variation in chromosomes 1 to 22 during AM. **(E).** Differences in cell abundance components between control and AM patients. Different colors represent different cell types. AM_EM, eutopic endometrium; AM_EC, ectopic endometrium; CNV, copy number variations; EC, endometrial cells; Ep, epithelial cells; En, endothelial cells; Naive T, initial T cells; Mac, macrophages, pDC, predendritic cells; SMC, smooth muscle cells; ILC, innate lymphoid cells; UMAP, Uniform Manifold Approximation and Projection.

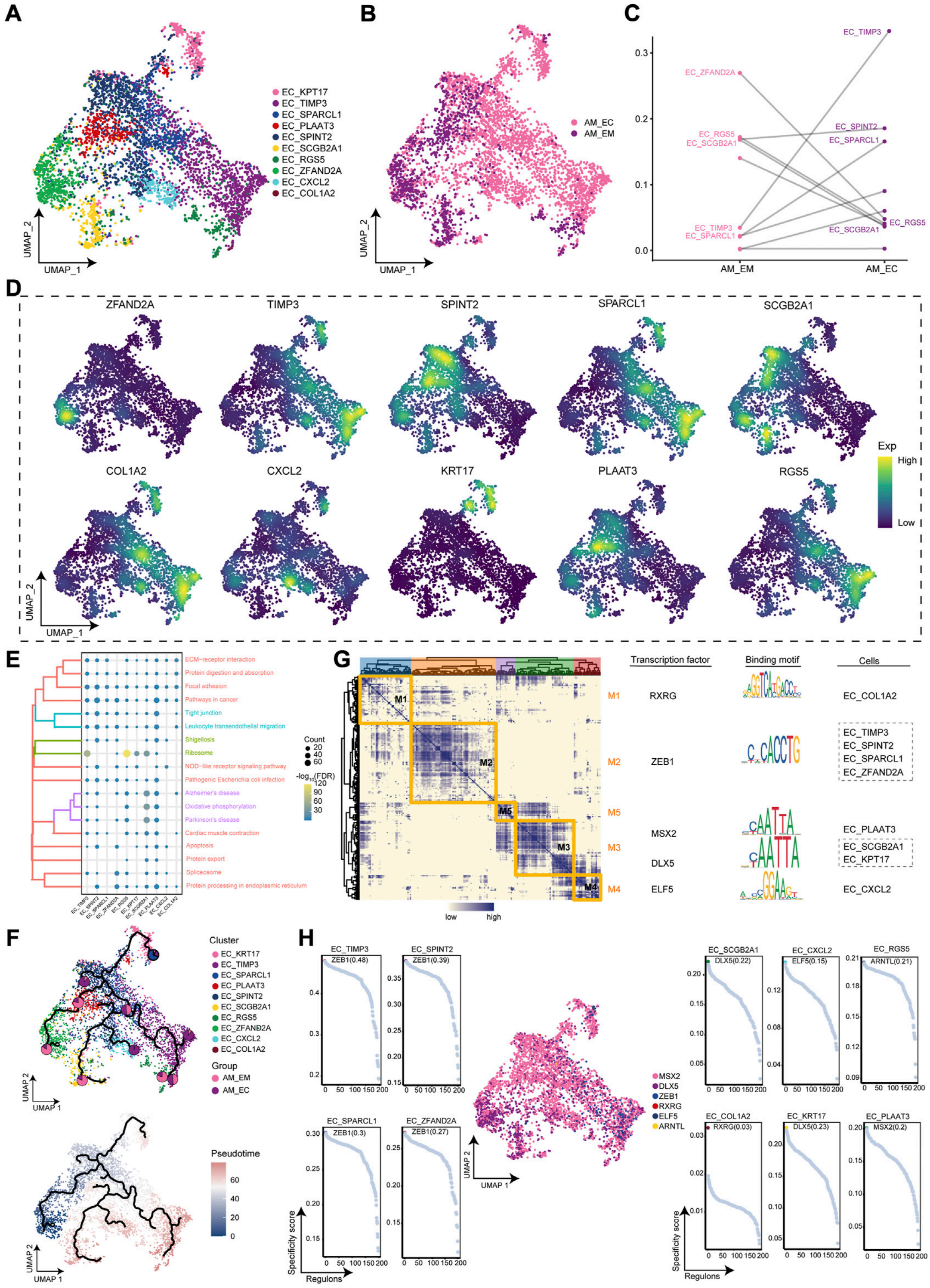


FIGURE 2
Identification of the endometrial cells (EC) clusters in adenomyosis (AM). **(A)**. Single-cell atlas shows the cellular clusters of EC. **(B).** Single-cell atlas shows clusters of EC cells in eutopic endometrium and ectopic endometrial samples from AM patients. **(C)**. Differences in the abundance of EC clusters in the eutopic and ectopic endometrium groups of AM patients. **(D).** Marker genes for the specific EC clusters. **(E).** Biological pathways in specific clusters of EC. **(F).** Single-cell atlas map the trajectory and pseudotime values of EC progression. Pie charts show the proportion of the different groups in the clusters. **(G).** Heatmap-TF-binding sequence-cell clusters demonstrate the gene regulatory network of EC clusters. **(H).** Transcription factors regulate the makers of EC clusters. AM_EM, eutopic endometrium; AM_EC, ectopic endometrium; EC, endometrial cells; UMAP, Uniform Manifold Approximation and Projection.

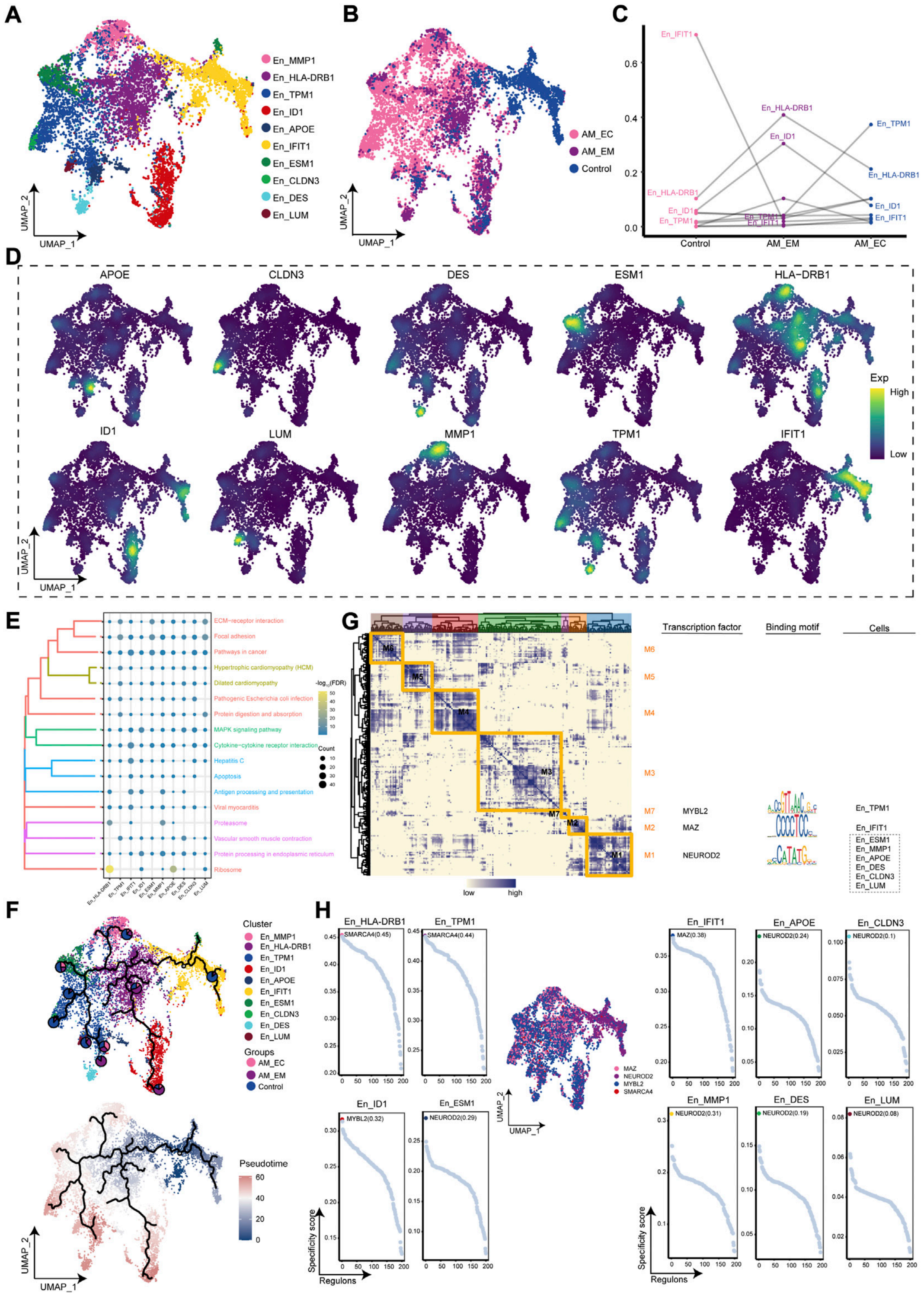


FIGURE 3
The adenomyosis (AM)-associated endothelial cell (En) clusters. **(A)**. Single-cell atlas shows the cell clusters of endothelial cells (En). **(B)**. Single-cell mapped the En clusters in the eutopic endometrium (AM_EM), ectopic endometrial (AM_EC), and control groups of AM patients. **(C)**. Differences in the abundance of En clusters among the AM_EM, AM_EC, and control groups. **(D)**. Marker genes for specific En clusters. **(E)**. Biological pathways in specific En clusters. **(F)**. Single-cell atlas map the trajectory and pseudotime values of development in En from AM. Pie charts show the proportion of the different groups in the clusters. **(G)**. Transcription factors of En clusters in a co-expression pattern. Left: Heat map identifies co-expression modules; middle: major transcription factors and binding sequences; right: cell clusters of transcription factors. **(H)**. Transcription factors regulated makers of EC clusters. AM_EM, eutopic endometrium; AM_EC, ectopic endometrium; UMAP, Uniform Manifold Approximation and Projection.

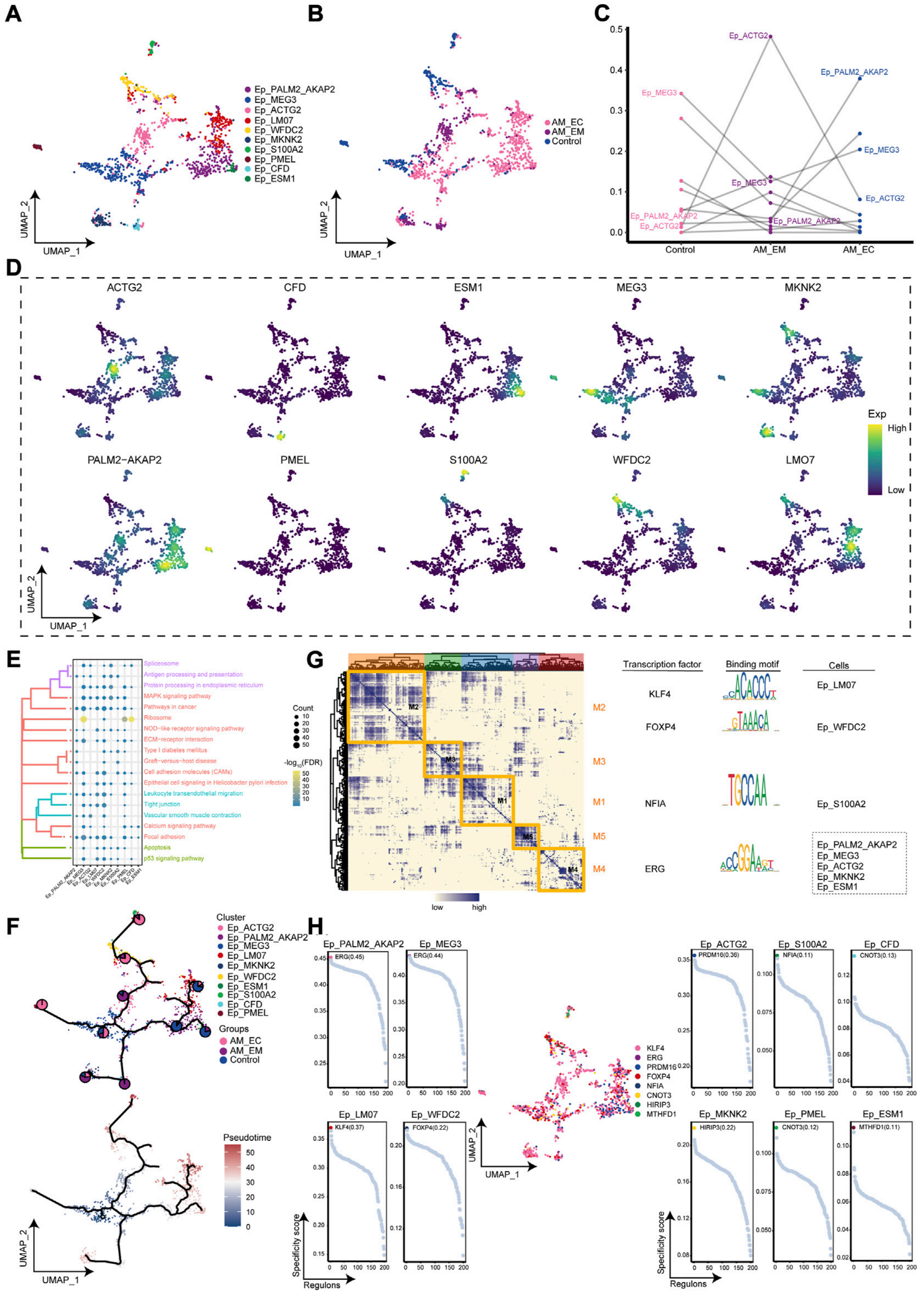


FIGURE 4
The adenomyosis (AM)-associated epithelial cell (Ep) clusters. **(A)**. Single-cell mapped of cell clusters of Ep. **(B).** Single-cell atlas shows the AM_EM, AM_EC, and control groups with clusters of Ep. **(C).** Differences in the cellular ecology of Ep clusters in different groups. **(D).** Marker genes for specific Ep clusters. **(E).** Biological pathways in Ep clusters. **(F).** Single-cell atlas map the developmental trajectory of Ep in AM. Pie charts show the proportion of the different groups in the cluster. **(G).** Transcription factors of Ep clusters in a co-expression pattern. Left: Heat map identified co-expression modules; Middle: major transcription factors and their binding sequences; Right: cell clusters of transcription factors. **(H).** Scatter plot of transcription factors in Ep clusters. AM_EM, eutopic endometrium; AM_EC, ectopic endometrium; UMAP, Uniform Manifold Approximation and Projection.

analysis, it was inferred that the EC_ZFAND2A cluster served as the developmental starting point, and then differentiated into other cell clusters (Figure 2F). Furthermore, we constructed the GRN and found that the GRN with TFs as pivots was organized into five modules (Figure 2G), such as RXRG, ZEB1, MSX2, DLX5, ELF5, ARNTL, to regulate the specific gene expression (Figure 2H). Above all, we identified EC clusters of AM patients, defined marker genes for their specific expression, and elucidated the GRN of evolved EC clusters.

Ecological landscape of adenomyosis-associated en clusters

The abundance of En was significant in the AM_EC group, from which we inferred that En may play a facilitating role in the disease course. Therefore, the cluster analysis of En yielded 10 cell clusters (Figure 3A). Further mapping of these cell clusters to the AM_EC, AM_EM, and control groups, and we found that mainly originated from AM patients (Figure 3B). En_HLA-DRB1 and En_ID1 were increased in AM_EM, En_TPM1 was significant in the AM_EC group, and En_IFIT1 was significantly decreased in AM patients (Figure 3C). Gene expression of markers for En clusters weas mapped to a single-cell atlas, including APOE, CLDN3, MMP1, HLA-DRB1, TPM1, ID1, IFIT1, ESM1, DES, and LUM (Figure 3D). En clusters were mainly related to cell proliferation, such as pathways in cancer, apoptosis and extracellular matrix receptor interactions (Figure 3E). Pseudotime trajectory analysis showed that the En_IFIT1 cluster evolved as a developmental starting point towards AM patients (Figure 3F). The results of GRN for En clusters indicated that marker genes were divided into seven modules, and regulated by TFs, such as MYBL2, MAZ, and NEUROD2 (Figure 3G). Figure 3H shows the transcription factors of En specific cell cluster. En cluster markers are regulated by transcription factors that promote the development of AM.

Exploring the ecological landscape in ep clusters of adenomyosis

Interestingly, Ep loses polarity and intercellular adhesion to gain migration capacity (Acloque et al., 2009), and the cell abundance of Ep clusters was significant in the AM_EC group. Therefore, cluster analysis of Ep clusters was again performed to obtain 10 cell clusters (Figure 4A), which were mapped to AM_EC, AM_EM, and control groups according to their sample source (Figure 4B). Compared with the control group, the cell abundance of Ep_ACTG2 was significantly increased in the AM_EM group, cell abundance of Ep_PALM2_AKAP2 was significant in the AM_EC group (Figure 4C). Subsequently, we mapped the expression of cluster markers (PALM2_AKAP2, MEG3, ACTG2, LM07, WFDC2, MKNK2, S100A2, PMEL, CFD, and ESM1) to the single-cell atlas of Ep clusters (Figure 4D). To further explore the biological signatures for the involvement of the Ep clusters in the AM, we performed enrichment analysis of the marker genes in the Ep clusters, showing that extracellular matrix receptor interactions, MAPK signaling pathway and apoptosis were significantly involved in Ep clusters (Figure 4E). The developmental trajectory of Ep was explored by pseudotime trajectory analysis, and the results indicated Ep_PMEL cluster was in an early developmental stage and evolved into AM_MKNK2, AM_ESM1, AM_LM07, and AM_S100A2 (Figure 4F). We further performed a GRN analysis of Ep clusters, showing that the marker genes of the Ep clusters were divided into five modules regulated by the transcription factors, such as KLF4, FOXP4, NFIA, and ERG (Figure 4G). Furthermore, we explored the expression of these TFs in specific Ep clusters and found that KLF4 was the most highly expressed in the cluster (Figure 4H), suggesting that high expression of KLF4 may be associated with the development of AM.

SMC-associated cell clusters of adenomyosis

During AM, ecological components of SMC clusters were significantly observed, so we will investigate their microenvironmental immune properties. The SMC clusters were re-clustered to obtain 10 cell clusters (Figure 5A) and mapped to different samples (Figure 5B). The SMC_TP53BP2 cluster had significant cellular abundance in the AM_EC and AM_EM groups compared to the control group, while SMC_IFI6 had significant cellular abundance in the control group (Figure 5C). Next, we showed the expression of marker genes was significantly changed in SMC clusters for clusters (Figure 5D). The SMC clusters were significantly involved in pathways, such as cytokine receptor interaction, vascular smooth muscle contraction and apoptosis (Figure 5E). We also further explored the developmental trajectory of the SMC, clarifying the evolution trajectory from the SMC_TP53BP2 cluster to the SMC_IFI6, SMC_VCAN, and the SMC_CXCL8 cluster (Figure 5F). In addition, the GRN analysis with TFs as the fulcrum yielded five modules (Figure F5G) with gene expression of specific SMC regulated by REL, MAZ, and ETV7 (Figure 5H). Taken together, these results suggest that certain specific clusters are closely associated with vascular smooth muscle contraction, can lead to smooth muscle ischemia in AM patients, and may promote the development of AM.

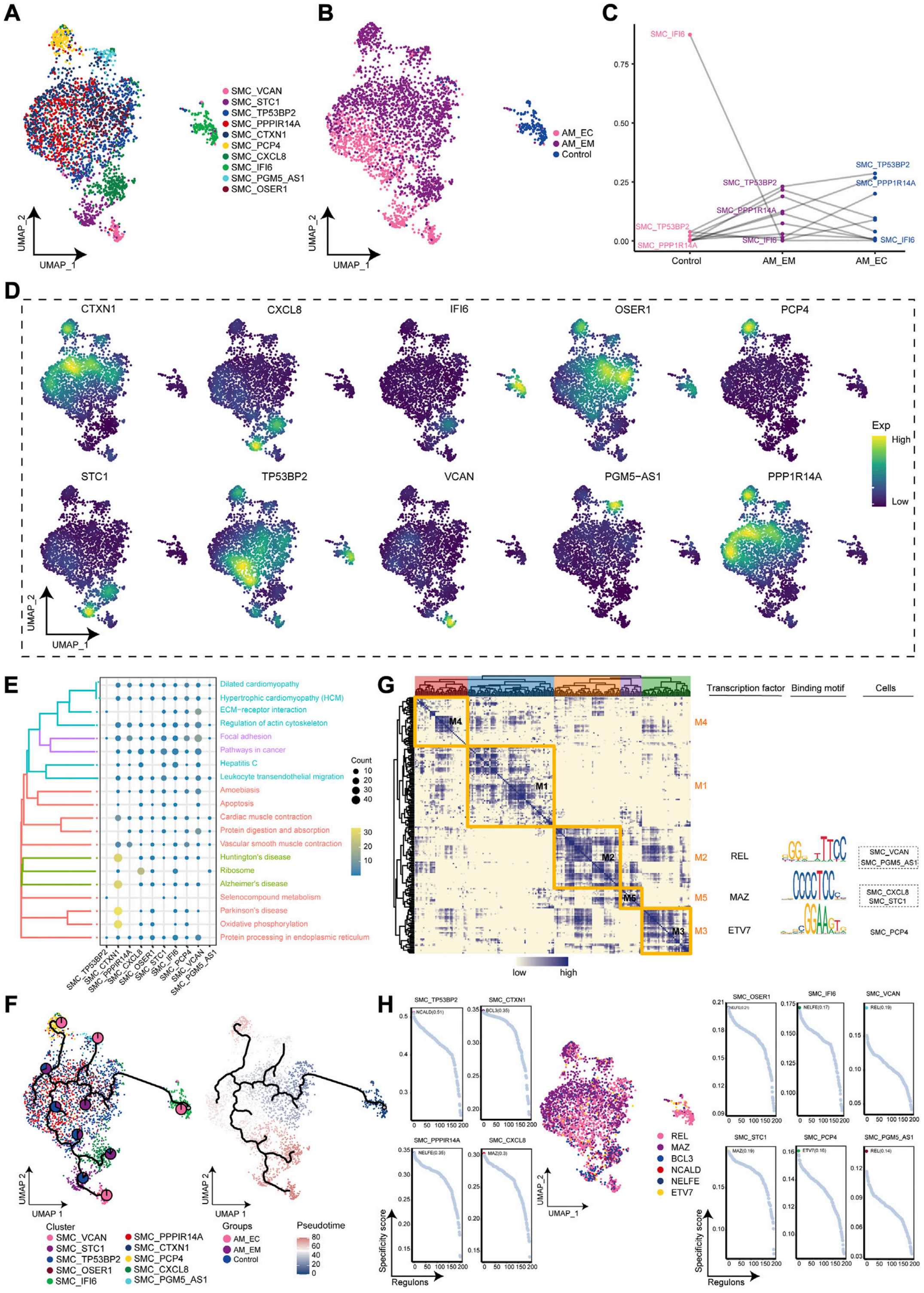


FIGURE 5

AM-related smooth muscle cell (SMC) clusters **(A)**. Single-cell atlas maps the cell clusters of SMC. **(B).** Single-cell atlas shows different groups of SMC clusters. **(C).** Differences in the cellular ecology of SMC clusters in different groups. **(D).** Marker genes for specific SMC clusters. **(E).** Biological pathways are significantly involved in the SMC clusters. **(F).** Single-cell atlas maps the evolutionary trajectories of SMC subgroups in AM. Pie charts show the proportion of the different groups in the subgroup. **(G)**. Transcription factors for clusters of SMC in a co-expression pattern. Left: Heat map identifies co-expression modules; middle: major transcription factors and their binding sequences; right: cell clusters of transcription factors. **(H).** Scatter plots showed transcription factors in clusters of SMC. AM_EM, eutopic endometrium; AM_EC, ectopic endometrium; SMC, smooth muscle cell; UMAP, Uniform Manifold Approximation and Projection.

Discussion

To date, most studies of orthoendometrium in women with AM have focused on the expression of a single gene or a limited number of genes (Benagiano and Brosens, 2012; Xiaoyu et al., 2014). However, few studies have revealed the uniqueness of each cell in the AM process at the individual cellular level. In this study, we analyzed the scRNA-seq data from the endometrial tissue of one patient with AM and one fibroid patient as controls, and explored changes in the cellular state and immune microenvironment of AM. Ectopic endometrium samples present high CNV levels, and are considered a potential factor in the development of the AM. The presence of certain specific cell clusters was associated with the progression of AM. In conclusion, it is possible to explore the global status of AM patients at the cellular level, providing new insights into the in-depth study of AM.

The ecological compositions of EC in the AM_EC and AM_EM groups were significant compared to the controls, indicating that EC plays an important role in the AM microenvironment. EC clusters were significantly involved in cell motility-related pathways such as focal adhesion and leukocyte transendothelial migration by enrichment analysis. AM patients suffer from uterine bleeding, pelvic pain, or infertility in women due to endometrial adhesion and destruction of the endometrium (Lacheta, 2019; Bulun et al., 2021). In the GRN, RXRG, ZEB1, MSX2, DLX5, and ELF5 regulated the EC clusters. Knockdown of SKP2 was found to reduce ZEB1 expression in endometrial stromal cells, thereby inhibiting their invasion and proliferation (Guo et al., 2022). However, few other TFs have been reported in AM. although many relevant mechanisms have been studied based on endometrial tissue, few studies have determined the effect of EC cells on AM at the cellular level. In conclusion, these results suggested that EC cells mediate and disrupt the endometrium, moreover, a cluster of markers regulated by TFs may promote the development of AM.

En has a dual role in immunology and pathology. On the one hand, the dysfunction will mediate the development of certain diseases, and on the other hand, they will actively mediate the immune response at the site of injury or infection (Pober and Sessa, 2007). Liu found co-localization of Ep and En markers in cluster one and promoted cell growth in AM (Liu et al., 2021), identical to the tumor-like characteristics reported by AM (Liu et al., 2018). Enrichment analysis showed that the En clusters were significantly involved in cancer-related pathways and extracellular matrix receptor interactions, indicating that the En was associated with cell proliferation and had certain malignant characteristics in the AM_EC group. Moreover, KLF4, FOXP4, NFIA, and ERG can regulate the markers of En clusters in the GRN. By inhibiting the biological functions of autophagy and metaphase during AM onset, KLF4 is abnormally reduced (Mei et al., 2022). However, other regulated TFs were almost rarely reported in AM, and whether En migrate needs further investigation.

Ep recognizes perturbations in their microenvironment, sends reinforcement signals, and transmits the signals to the immune system (Larsen et al., 2020). Studies have shown that epithelial immune cells of endometria can enhance cell survival and epithelial protective barrier function (Ho et al., 2006). Notably, the highest percentage of EP was found in the AM_EC and AM_EM groups; however, the number of EP was lower in the control group, which may be due to the significant postoperative endometrial thinning in the control group. Moreover, the Ep clusters were also significantly involved in the biological pathways such as leukocyte transendothelial migration, the MAPK pathway, and focal adhesions. Migration across the En after leukocyte adhesion, indicated that Ep was associated with high motility and migration. The MAPK pathway was required for the cell migration process, and the cytokines can also mediate cell migration (Liu et al., 2021). Furthermore, development of AM was improved by inhibiting the activated MAPK/ERK signaling pathway (Ying et al., 2021). In addition, study has shown that Ep loses their polarity and intercellular adhesion during adhesive epithelial interstitial transformation (EMT), to gain the ability to migrate to a mesenchymal phenotype (Acloque et al., 2009) and that EMT may play a key role in the pathogenesis of AM (Wang et al., 2021). In particular, during the conversion of Ep into En, a significant accumulation of angiogenic mimicry formation in AM_EC was found (Liu et al., 2021). Our results suggested that EP subsets significantly involved in pro-migratory pathways may play an important role in AM progression.

Abnormal proliferation of SMC in the endometrium-myometrial junction area is an important cause of AM (Huang et al., 2021), and the emergence of AM causes hyperplasia and hypertrophy of the surrounding SMC (Zhai et al., 2020). Compared to the SMC in the normal uterus, the uterine SMC has hypertrophy and ultrastructural changes, which may have contractility effects on the myometrium (Mehasseb et al., 2010). In the AM global single-cell ecosystem, the components of the SMC are more prominent in the AM than in the controls due to AM-induced SMC abnormalities. SMC clusters were significantly involved in cytokine receptor interactions and vascular smooth muscle contraction. SMCs are capable of significant phenotypic changes in response to changes in local environmental cues, including cell-cell and cell-matrix interactions, as well as various inflammatory mediators (Owens et al., 2004). In addition, vascular smooth muscle contraction occurs, leading to smooth muscle ischemia and dysmenorrhea in AM patients (Zhai et al., 2020). In GRN, we obtained five co-expression modules and three TFs, including REL, MAZ, and ETV7. Studies have demonstrated that REL was expressed and localized in the epithelial or stromal cells after castrated prostate patients (Rosa-Ribeiro et al., 2014). MAZ, as a transcriptional activator, may participate in the development of atherosclerosis (Ponnusamy et al., 2015). ETV7 promotes the resistance of breast cancer cells to chemotherapy and radiotherapy (Pezze et al., 2021). However, these transcription factors have hardly been reported in AM. Generally speaking, maintaining the microenvironment homeostasis of SMC is critical for inhibition of development in AM.

In the present study, we found expression profiles of 42,260 cells and identified 10 cell clusters, including EC, fibroblasts, Ep, En, CD8$^+$T, CD4$^+$T, Naive T, Mac, pDC, SMC, and ILC. Among there, significant abundance of EC, Ep, En, and SMC in AM patients comparing the controls. Furthermore, the Ep clusters were mainly involved in leukocyte transendothelial cell migration and apoptosis; En clusters were mainly involved in pathways in cancer and apoptosis. Especially, some cell clusters were involved in cell migration and apoptosis may be promote the development of AM patients. Moreover, we obtained co-expression modules and TFs associated with the significant cell clusters by GRN comparing with the previous studies.

In conclusion, our study established a single-cell ecological landscape of the endometrium between control and AM patients, and explored the dynamic changes of immune cells during AM. However, this study has several limitations. First, the samples were too small and larger sample size is needed for a large-scale study. Moreover, this study was mainly based on bioinformatics analysis and therefore requires relevant molecular and cellular experimental validation.

Data availability statement

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

Author contributions

JL and LL designed the study. FZ, WY and SM contributed to the literature research and analyzed and interpreted the data. SC, JL wrote the initial draft of the manuscript. JL, LL and D-YZ reviewed and edited the manuscript. All authors read and approved the manuscript.

Funding

This research was supported by Guangxi traditional Chinese medicine appropriate technology development and promotion project (GZSY22-83), Natural Science Foundation of China (82,060,470), the Natural Science Foundation of Guangxi Province of China (2020GXNSFAA297110) and the Science and Technology Plan Project of Liuzhou (2020NBAB0823).

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.2022.1020757/full#supplementary-material

References

Abbott, J. A. (2017). Adenomyosis and abnormal uterine bleeding (AUB-A)-Pathogenesis, diagnosis, and management. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 40, 68–81. doi:10.1016/j.bpobgyn.2016.09.006

Acloque, H., Adams, M. S., Fishwick, K., Bronner-Fraser, M., and Nieto, M. A. (2009). Epithelial-mesenchymal transitions: The importance of changing cell state in development and disease. *J. Clin. Invest.* 119 (6), 1438–1449. doi:10.1172/JCI38019

Aibar, S., Gonzalez-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., et al. (2017). Scenic: Single-cell regulatory network inference and clustering. *Nat. Methods* 14 (11), 1083–1086. doi:10.1038/nmeth.4463

Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44. doi:10.1038/nbt.4314

Benagiano, G., and Brosens, I. (2012). The endometrium in adenomyosis. *Womens Health (Lond)* 8 (3), 301–312. doi:10.2217/whe.12.8

Bergeron, C., Amant, F., and Ferenczy, A. (2006). Pathology and physiopathology of adenomyosis. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 20 (4), 511–521. doi:10.1016/j.bpobgyn.2006.01.016

Bird, C. C., McElin, T. W., and Manalo-Estrella, P. (1972). The elusive adenomyosis of the uterus--revisited. *Am. J. Obstet. Gynecol.* 112 (5), 583–593. doi:10.1016/0002-9378(72)90781-8

Borgfeldt, C., and Andolf, E. (2004). Cancer risk after hospital discharge diagnosis of benign ovarian cysts and endometriosis. *Acta Obstet. Gynecol. Scand.* 83 (4), 395–400. doi:10.1111/j.0001-6349.2004.00305.x

Bulun, S. E., Yildiz, S., Adli, M., and Wei, J. J. (2021). Adenomyosis pathogenesis: Insights from next-generation sequencing. *Hum. Reprod. Update* 27 (6), 1086–1097. doi:10.1093/humupd/dmab017

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36 (5), 411–420. doi:10.1038/nbt.4096

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566 (7745), 496–502. doi:10.1038/s41586-019-0969-x

Chapron, C., Vannuccini, S., Santulli, P., Abrao, M. S., Carmona, F., Fraser, I. S., et al. (2020). Diagnosing adenomyosis: An integrated clinical and imaging approach. *Hum. Reprod. Update* 26 (3), 392–411. doi:10.1093/humupd/dmz049

Graziano, A., Lo Monte, G., Piva, I., Caserta, D., Karner, M., Engl, B., et al. (2015). Diagnostic findings in adenomyosis: A pictorial review on the major concerns. *Eur. Rev. Med. Pharmacol. Sci.* 19 (7), 1146–1154.

Guo, Y., Wang, J., Jia, C., and Liao, Y. (2022). SKP2 regulates ZEB1 expression and stimulates eutopic endometrial stromal cell invasion and proliferation of adenomyosis. *Reprod. Biol.* 22 (1), 100578. doi:10.1016/j.repbio.2021.100578

Hedlund, E., and Deng, Q. (2018). Single-cell RNA sequencing: Technical advancements and biological applications. *Mol. Asp. Med.* 59, 36–46. doi:10.1016/j.mam.2017.07.003

Ho, L. S., Tsang, L. L., Chung, Y. W., and Chan, H. C. (2006). Establishment of a mouse primary co-culture of endometrial epithelial cells and peripheral blood leukocytes: Effect on epithelial barrier function and leukocyte survival. *Cell. Biol. Int.* 30 (12), 977–982. doi:10.1016/j.cellbi.2006.07.004

Huang, J. H., Duan, H., Wang, S., Wang, Y. Y., and Lv, C. X. (2021). Upregulated microRNA let-7a accelerates apoptosis and inhibits proliferation in uterine junctional zone smooth muscle cells in adenomyosis under conditions of a normal activated hippo-YAP1 axis. *Reprod. Biol. Endocrinol.* 19 (1), 81. doi:10.1186/s12958-021-00753-w

Kok, V. C., Tsai, H. J., Su, C. F., and Lee, C. K. (2015). The risks for ovarian, endometrial, breast, colorectal, and other cancers in women with newly diagnosed endometriosis or adenomyosis: A population-based study. *Int. J. Gynecol. Cancer* 25 (6), 968–976. doi:10.1097/IGC.0000000000000454

Kruger-Genge, A., Blocki, A., Franke, R. P., and Jung, F. (2019). Vascular endothelial cell biology: An update. *Int. J. Mol. Sci.* 20 (18), E4411. doi:10.3390/ijms20184411

Lacheta, J. (2019). Uterine adenomyosis: Pathogenesis, diagnostics, symptomatology and treatment. *Ceska Gynekol.* 84 (3), 240–246.

Larsen, S. B., Cowley, C. J., and Fuchs, E. (2020). Epithelial cells: Liaisons of immunity. *Curr. Opin. Immunol.* 62, 45–53. doi:10.1016/j.coi.2019.11.004

Leyendecker, G., Bilgicyildirim, A., Inacker, M., Stalf, T., Huppert, P., Mall, G., et al. (2015). Adenomyosis and endometriosis. Re-visiting their association and further insights into the mechanisms of auto-traumatisation. An MRI study. *Arch. Gynecol. Obstet.* 291 (4), 917–932. doi:10.1007/s00404-014-3437-8

Liu, L., Luo, N., Guo, J., Xie, Y., Chen, L., and Cheng, Z. (2018). Berberine inhibits growth and inflammatory invasive phenotypes of ectopic stromal cells: Imply the possible treatment of adenomyosis. *J. Pharmacol. Sci.* 137 (1), 5–11. doi:10.1016/j.jphs.2017.12.001

Liu, Z., Sun, Z., Liu, H., Niu, W., Wang, X., Liang, N., et al. (2021). Single-cell transcriptomic analysis of eutopic endometrium and ectopic lesions of adenomyosis. *Cell. Biosci.* 11 (1), 51. doi:10.1186/s13578-021-00562-z

Mehasseb, M. K., Bell, S. C., Pringle, J. H., and Habiba, M. A. (2010). Uterine adenomyosis is associated with ultrastructural features of altered contractility in the inner myometrium. *Fertil. Steril.* 93 (7), 2130–2136. doi:10.1016/j.fertnstert.2009.01.097

Mei, J., Sheng, X., Yan, Y., Cai, X., Zhang, C., Tian, J., et al. (2022). Decreased Kruppel-like factor 4 in adenomyosis impairs decidualization by repressing autophagy in human endometrial stromal cells. *BMC Mol. Cell. Biol.* 23 (1), 24. doi:10.1186/s12860-022-00425-6

Owens, G. K., Kumar, M. S., and Wamhoff, B. R. (2004). Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol. Rev.* 84 (3), 767–801. doi:10.1152/physrev.00041.2003

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344 (6190), 1396–1401. doi:10.1126/science.1254257

Peric, H., and Fraser, I. S. (2006). The symptomatology of adenomyosis. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 20 (4), 547–555. doi:10.1016/j.bpobgyn.2006.01.006

Pezze, L., Meskyte, E. M., Forcato, M., Pontalti, S., Badowska, K. A., Rizzotto, D., et al. (2021). ETV7 regulates breast cancer stem-like cell features by repressing IFN-response genes. *Cell. Death Dis.* 12 (8), 742. Supplementary Table 1 The cell clusters and cell type of the adenomyosis. doi:10.1038/s41419-021-04005-y

Pober, J. S., and Sessa, W. C. (2007). Evolving functions of endothelial cells in inflammation. *Nat. Rev. Immunol.* 7 (10), 803–815. doi:10.1038/nri2171

Ponnusamy, M., Zhuang, M. A., Zhou, X., Tolbert, E., Bayliss, G., Zhao, T. C., et al. (2015). Activation of sirtuin-1 promotes renal fibroblast activation and aggravates renal fibrogenesis. *J. Pharmacol. Exp. Ther.* 354 (2), 142–151. doi:10.1124/jpet.115.224386

Rosa-Ribeiro, R., Nishan, U., Vidal, R. O., Barbosa, G. O., Reis, L. O., Cesar, C. L., et al. (2014). Transcription factors involved in prostate gland adaptation to androgen deprivation. *PLoS One* 9 (6), e97080. doi:10.1371/journal.pone.0097080

Stratopoulou, C. A., Donnez, J., and Dolmans, M. M. (2021). Conservative management of uterine adenomyosis: Medical vs. Surgical approach. *J. Clin. Med.* 10 (21), 4878. doi:10.3390/jcm10214878

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, et al. (2019). Comprehensive integration of single-cell data. *Cell.* 177 (7), 1888–1902. doi:10.1016/j.cell.2019.05.031

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6 (5), 377–382. doi:10.1038/nmeth.1315

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32 (4), 381–386. doi:10.1038/nbt.2859

Tremellen, K. P., and Russell, P. (2012). The distribution of immune cells and macrophages in the endometrium of women with recurrent reproductive failure. II: Adenomyosis and macrophages. *J. Reprod. Immunol.* 93 (1), 58–63. doi:10.1016/j.jri.2011.12.001

Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* 15 (7), 2247–2276. doi:10.1038/s41596-020-0336-2

Wang, Y. Y., Duan, H., Wang, S., Quan, Y. J., Huang, J. H., and Guo, Z. C. (2021). Talin1 induces epithelial-mesenchymal transition to facilitate endometrial cell migration and invasion in adenomyosis under the regulation of microRNA-145-5p. *Reprod. Sci.* 28 (5), 1523–1539. doi:10.1007/s43032-020-00444-8

Xiaoyu, L., Weiyuan, Z., Ping, J., Anxia, W., and Liane, Z. (2014). Serum differential proteomic analysis of endometriosis and adenomyosis by iTRAQ technique. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 182, 62–65. doi:10.1016/j.ejogrb.2014.08.034

Yang, J. H., Chen, M. J., Chen, H. F., Lee, T. H., Ho, H. N., and Yang, Y. S. (2004). Decreased expression of killer cell inhibitory receptors on natural killer cells in eutopic endometrium in women with adenomyosis. *Hum. Reprod.* 19 (9), 1974–1978. doi:10.1093/humrep/deh372

Yates, L. R., and Campbell, P. J. (2012). Evolution of the cancer genome. *Nat. Rev. Genet.* 13 (11), 795–806. doi:10.1038/nrg3317

Ying, P., Li, H., Jiang, Y., Yao, Z., Lu, S., Yang, H., et al. (2021). Qiu's neiyi recipe regulates the inflammatory action of adenomyosis in mice via the MAPK signaling pathway. *Evid. Based. Complement. Altern. Med.* 2021, 9791498. doi:10.1155/2021/9791498

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

Zhai, J., Vannuccini, S., Petraglia, F., and Giudice, L. C. (2020). Adenomyosis: Mechanisms and pathogenesis. *Semin. Reprod. Med.* 38 (2-03), 129–143. doi:10.1055/s-0040-1716687

Frontiers in Genetics

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fgene.2022.1022236

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Li Tao,
Chinese Academy of Sciences (CAS),
China
Haitao Xu,
Anqing Hospital affiliated to Anhui
Medical University, China

*CORRESPONDENCE
Xiaoliang Wang,
wangxiaoliang99@126.com

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

RECEIVED 18 August 2022
ACCEPTED 12 October 2022
PUBLISHED 28 October 2022

CITATION
Zhu B, Wang S, Wang R and Wang X
(2022), Identification of molecular
subtypes and a six-gene risk model
related to cuproptosis for triple negative
breast cancer.
Front. Genet. 13:1022236.
doi: 10.3389/fgene.2022.1022236

Identification of molecular subtypes and a six-gene risk model related to cuproptosis for triple negative breast cancer

Baoxi Zhu, Songping Wang, Rui Wang and Xiaoliang Wang*

Department of Thyroid and Breast Surgery, Anhui No.2 Provincial People's Hospital,, Hefei, China

Background: Breast cancer is the mostly diagnosed cancer worldwide, and triple negative breast cancer (TNBC) has the worst prognosis. Cuproptosis is a newly identified form of cell death, whose mechanism has not been fully explored in TNBC. This study thought to unveil the potential association between cuproptosis and TNBC.

Materials and Methods: Gene expression files with clinical data of TNBC downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases were included in this study. Consensus clustering was utilized to perform molecular subtyping based on cuproptosis-associated genes. Limma package was applied to distinguish differentially expressed genes. Univariate Cox regression was used to identify prognostic genes. Least absolute shrinkage and selection operator and stepwise Akaike information criterion optimized and established a risk model.

Results: We constructed three molecular subtypes based on cuproptosis-associated genes, and the cuproptosis-based subtyping showed a robustness in different datasets. Clust2 showed the worst prognosis and immune-related pathways such as chemokine signaling pathway were significantly activated in clust2. Clust2 also exhibited a high possibility of immune escape to immune checkpoint blockade. In addition, a six-gene risk model was established manifesting a high AUC score over 0.85 in TCGA dataset. High- and low-risk groups had distinct prognosis and immune infiltration. Finally, a nomogram was constructed with strong performance in predicting TNBC prognosis than the staging system.

Conclusion: The molecular subtyping system related to cuproptosis had a potential in guiding immunotherapy for TNBC patients. Importantly, the six-gene risk model was effective and reliable to predict TNBC prognosis.

KEYWORDS

triple negative breast cancer, cuproptosis, molecular subtypes, risk model, SIX-gene

Introduction

Breast cancer is one of the leading cause of cancer death in women, which is the top one diagnosed cancer type with 2,261,419 new cases (11.7% of total cases) in 2020 according to the global cancer statistics (Sung et al., 2021). The overall survival of breast cancer is markedly different in developed and developing countries, with an estimated 5-year survival of 80% and below 40%, respectively (Coleman et al., 2008). The incidence of breast cancer elevates with age but seldomly found before the age of 20 years and breast cancer most commonly occurs in 40–50 aged women (Akram et al., 2017). Although many versions of guidelines for the diagnosis and treatment of breast cancer have been established, such as European Breast Guidelines (Schünemann et al., 2020) and the American Joint Committee on Cancer's (AJCC) guideline (Plichta et al., 2020), the treatment for triple negative breast cancer (TNBC) still remains a challenge. TNBC is a clinically aggressive type of breast cancer with poor survival, compared with other breast cancer types, including HER2-positive, oestrogen receptor (ER)-positive and progesterone receptor (PR)-positive. Chemotherapy resistance and immune escape common occur in TNBC, which makes an obstacle in TNBC treatment (Wein and Loi, 2017). Therefore, accurate molecular biomarkers or subtypes are strongly needed to guide personalized therapy for TNBC.

Programmed cell death is recognized as a promising therapeutic target in cancer therapy, where necroptosis, pyroptosis, and apoptosis are the most studied types (Bertheloot et al., 2021). Cuproptosis is considered as a new form of programmed cell death involved in the proliferation of lung cancer cells (Tang et al., 2022). Copper chelators such as RPTDH/R848 nanoparticles are demonstrated to be able to suppress cancer cell growth and metastasis in breast cancer (Zhou et al., 2019), inspiring a possibility that cuproptosis is a potential target for cancer treatment. Up to now, studies have discovered a series of prognostic signatures related to cuproptosis for different cancer types such as kidney renal clear cell carcinoma (Ji et al., 2022), melanoma (Lv et al., 2022), and hepatocellular carcinoma (Zhang et al., 2022). However, the relation between cuproptosis and TNBC has not been revealed.

Therefore, in this study, we aimed to analyze the role of cuproptosis in TNBC, and construct molecular subtypes based on cuproptosis-associated genes by using gene expression data of TNBC obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. By comparing the molecular features of different subtypes, we unveiled the relation between cuproptosis and immune infiltration. Moreover, a risk model related to cuproptosis was established for predicting TNBC prognosis. The risk model was effective to distinguish TNBC patients into different risk types. Notably, the model outperformed the AJCC staging system, which had a potential to be used as a prognostic signature in TNBC.

Materials and methods

Data collection and preprocessing

The RNA-seq data of TNBC was downloaded from Genomic Data Commons (GDC) Data Portal by TCGA GDC API (https://portal.gdc.cancer.gov/projects/TCGA-BRCA, named as TCGA dataset). GSE103091 dataset was downloaded from GEO database (https://www.ncbi.nlm.nih.gov/geo/). TNBC samples without progression-free survival (PFS) or survival status were eliminated. TNBC samples with PFS shorter than 30 days or more than 10 years were excluded. In GSE103091 dataset, Ensembl ID was converted to gene symbol and we used the averaged expression level when a gene had multiple Ensembl IDs. Finally, 105 TNBC samples and 113 paracancerous samples were remained in TCGA dataset, and 91 TNBC samples were remained in GSE103091 dataset.

The source of cuproptosis genes

Cuproptosis genes were obtained from a previous study (Tsvetkov et al., 2022), and a total of 13 cuproptosis genes were used in the study including FDX1, LIPT1, LIAS, DLD, DBT, GCSH, DLST, DLAT, PDHA1, PDHB, SLC31A1, ATP7A, and ATP7B.

Identification of prognostic cuproptosis-associated genes

Firstly, single sample gene set enrichment anlaysis (ssGSEA) was used to calculate the enrichment score of 13 cuproptosis genes for each sample in TCGA dataset. Limma R package (Ritchie et al., 2015) was applied to screen differentially expressed genes (DEGs) between paracancerous and tumor samples (false discovery rate (FDR) < 0.05 and $|\log2FC| > 1$). Then Pearson correlation analysis was performed to evaluate the correlation between the DEG expression and the ssGSEA score of cuproptosis. DEGs with $|$correlation coefficient (R)$| > 0.4$ and $p < 0.05$ were selected. Next, univariate Cox regression analysis in the survival R package was conducted on the DEGs and DEGs with $p < 0.05$ as the input for unsupervised consensus clustering.

Constructing molecular subtypes based on prognostic cuproptosis-associated genes

ConsensusClusterPlus R package (Wilkerson and Hayes, 2010) was used for conducting unsupervised consensus clustering to identify molecular subtypes. The expression of prognostic cuproptosis-associated genes were used as a basis for clustering samples. KM algorithm and Euclidean distance were set to carry out 500 bootstraps with each bootstrap consisting of 80% of samples in TCGA dataset. Cluster number k was chosen from 2 to 10. The optimal cluster number was determined according to cumulative distribution function (CDF) and area under CDF curve.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was utilized to calculate the enrichment score of functional pathways for molecular subtypes. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were obtained from KEGG database (https://www.genome.jp/kegg/).

Establishing a risk model

Firstly, DEGs between different molecular subtypes were identified with limma R package (FDR <0.05 and |log2FC| > 1.5). Least absolute shrinkage and selection operator (LASSO) regression analysis (Friedman et al., 2010) decreased the number of DEGs in glmnet R package. Stepwise Akaike information criterion (stepAIC) was applied for further optimizing the risk model through MASS R package (Zhang, 2016). We determined the risk model according to the formula: risk score = $\Sigma\beta i \times \text{Expi}$, where β indicates the coefficient of prognostic genes and Expi indicates the expression level of prognostic genes. Each sample obtained a risk score, which was subsequently transferred to z-score. Samples were stratified into high-risk and low-risk groups according to the z-score = 0. Kaplan-Meier survival analysis was conducted to evaluate the prognosis of the two risk groups.

Assessment of immune infiltration

Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) tool was implemented to evaluate stromal and immune infiltration (Yoshihara et al., 2013). Microenvironment Cell Populations (MCP)-counter methodology was applied to assess the enrichment of 10 immune cells (Becht et al., 2016). SsGSEA algorithm in GSVA R package was performed to predict estimated proportion of 28 immune cells (Hänzelmann et al., 2013).

Statistical analysis

The bioinformatics analysis in this study was supported by Sangerbox platform (http://vip.sangerbox.com/) (Shen et al., 2022). R software (v4.1) was used as a platform to conduct all statistical analysis. Log-rank test was performed in Kaplan-Meier survival analysis, univariate and multivariate Cox regression analysis. Student t test was performed to examine the difference between two groups. ANOVA was conducted to test the difference among three groups. $p < 0.05$ was considered as statistically significant.

Results

Identification of prognostic genes associated with cuproptosis

Firstly, we calculated the ssGSEA score of cuproptosis pathway based on 13 cuproptosis genes for each TNBC sample in TCGA dataset (Supplementary Table S1). Paracancerous samples had obviously higher cuproptosis score than tumor samples (Figure 1A). Then differential analysis was performed to identify DEGs between TNBC and paracancerous samples. A total of 3125 DEGs were screened under FDR <0.05 and |log2FC| > 1 (Figure 1B). Next, we analyzed the relation between the expression of DEGs and ssGSEA of cuproptosis by Pearson correlation analysis. 1,275 DEGs with $|R| > 0.4$ and $p < 0.05$ were selected for further univariate Cox regression analysis (Supplementary Table S2). 39 prognostic DEGs were found to be significantly associated with TNBC prognosis in TCGA dataset ($p < 0.05$, Supplementary Table S3), whose expression levels were significantly different between paracancerous and tumor samples ($p < 0.0001$, Figure 1C).

Construction of molecular subtypes based on cuproptosis-associated genes

Based on the expression profiles of the 39 cuproptosis-associated genes, we then constructed molecular subtypes through consensus clustering. According to the CDF curve, cluster number k = 3 was determined as the optimal (Figures 2A–C). Three molecular subtypes (clust1, clust2, and clust3) were distinguished based on the 39 cuproptosis-associated genes, and they showed distinct PFS in both TCGA and GSE103091 datasets (Figures 2D,E; Supplementary Figure S1, log-rank $p = 0.0038$ and 0.036, respectively). Clust2 had the shortest PFS and the most number of dead samples, while clust1 had the favorable prognosis (Figure 2F), indicating that cuproptosis-associated genes may be involved in the TNBC progression.

Differential pathways and immune infiltration of three molecular subtypes

Next we analyzed the enriched pathways of the three subtypes by GSEA. By comparing clust2 to non-clust2 (clust1 and clust3), we observed that immune-related pathways and tumor-related pathways were obviously activated in clust2, such as cytokine-

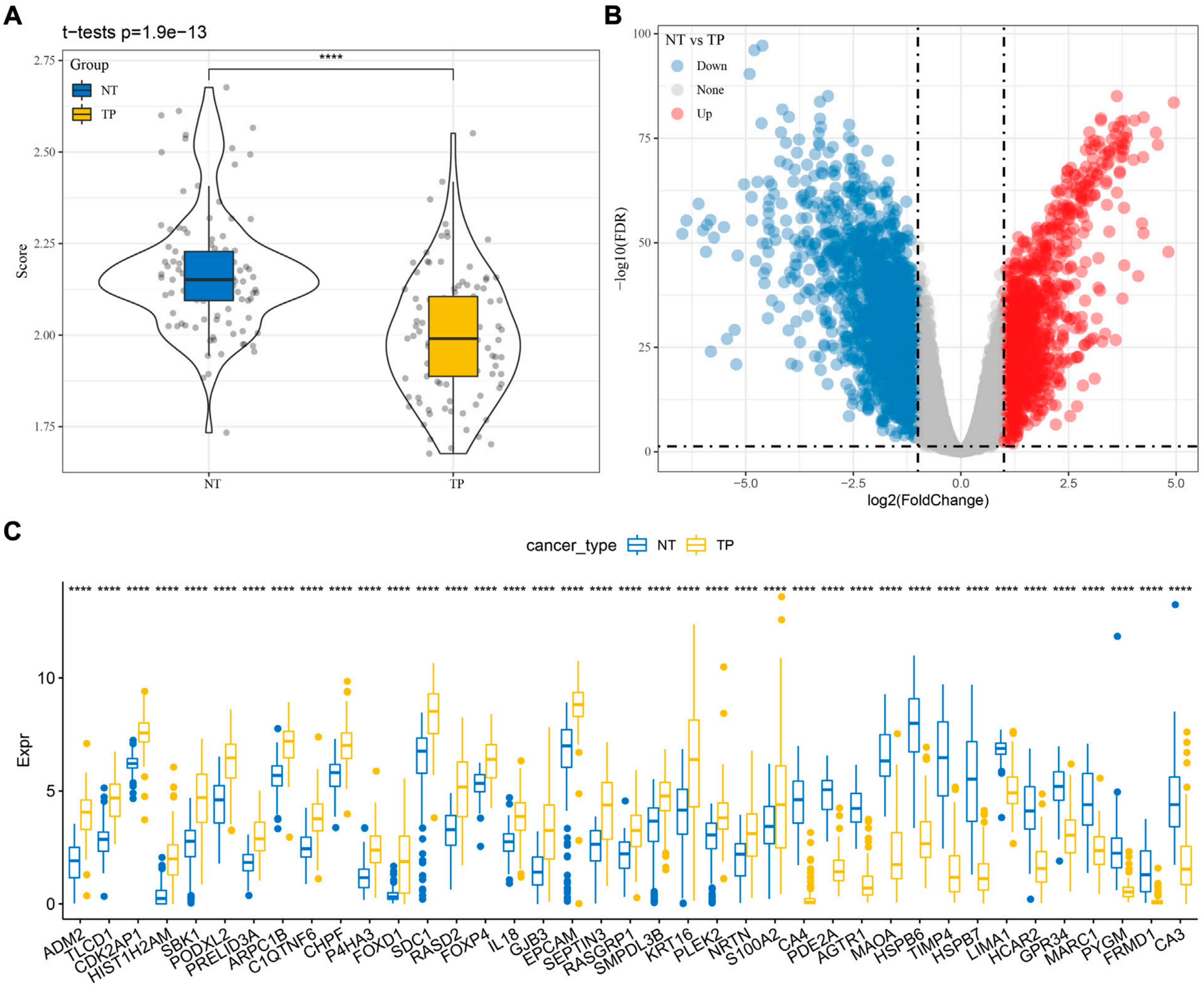


FIGURE 1
Identification of cuproptosis-associated genes related to TNBC prognosis in TCGA dataset. **(A)** The ssGSEA score of cuproptosis in paracancerous (NT) and tumor samples (TP). **(B)** Volcano plot of DEGs between NT and TP samples. **(C)** The expression of seven prognostic cuproptosis-associated genes in NT and TP samples. Student t test was performed. ****$p < 0.0001$.

cytokine receptor interaction, chemokine signaling pathway, MAPK signaling pathway, toll-like receptor signaling pathway, TGF-β signaling pathway, and pathways in cancer (Figure 3A). In clust1 vs non-clust1, the above pathways were significantly suppressed (Supplementary Figure S2), suggesting that cuproptosis-associated genes were involved in the immune regulation. Pathways related to cell proliferation and cell death were evaluated in the three subtypes. Among the six pathways, P53 signaling pathway was the most enriched in clust2 and clust1 had the lowest enrichment of cell death-related pathways including necroptosis, ferroptosis, and apoptosis (Figure 3B, ANOVA $p < 0.05$). This indicated an interaction of cuproptosis with other cell death pathways.

Given that immune-related pathways were differentially enriched in three subtypes, we then assessed the immune infiltration. Not surprisingly, clust1 had the lowest stromal score and immune score, compared with other two subtypes (Figure 2C, ANOVA $p < 0.0001$). Estimation of 10 immune cell types by MCP-counter also showed a lowest enrichment of them in clust1 such as T cells, monocytic lineage, and myeloid dendritic cells ($p < 0.05$, Figure 3D). Notably, clust2 had the highest enrichment of fibroblasts ($p < 0.01$, Figure 3D). Similar results were outputted through ssGSEA that majority of immune cells had a low estimated proportion in clust1 (Supplementary Figure S2B). Furthermore, we also determined the expression of immune checkpoint genes in the three subtypes. The result showed that 22 of 47 immune checkpoints were differentially expressed in the three subtypes (Supplementary Figure S2C). We suspected that cuproptosis-associated genes had an influence in tumor microenvironment and therefore affected the efficiency of

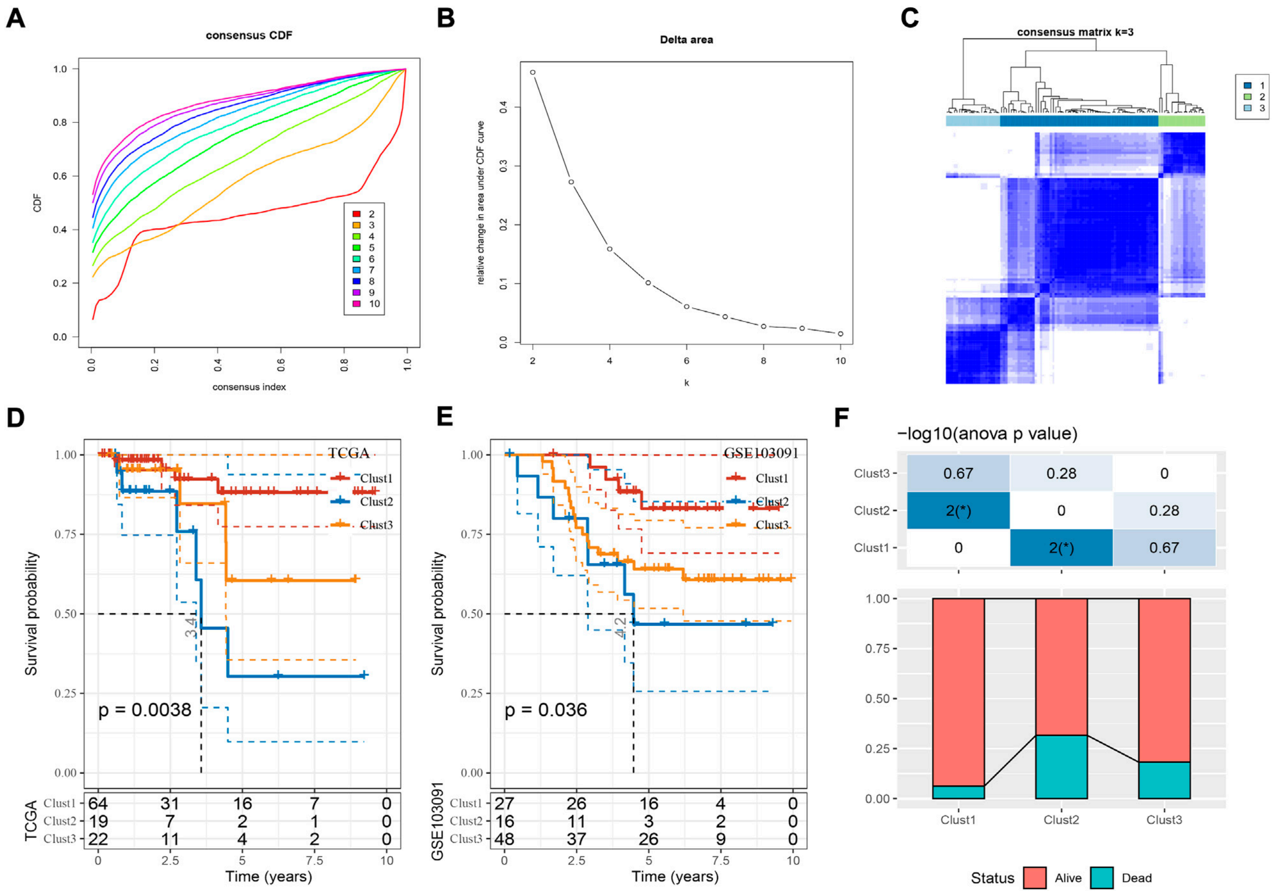


FIGURE 2
Construction of molecular subtypes based on cuproptosis-associated genes. **(A,B)** Consensus CDF curve and delta area under CDF curve when cluster number k = 2 to 10. **(C)** Consensus matrix when k = 3. **(D,E)** Kaplan-Meier survival curve of three molecular subtypes in TCGA **(D)** and GSE103091 **(E)** datasets. Log-rank test was performed. **(F)** The distribution of alive and dead samples in three subtypes. ANOVA was conducted. *$p < 0.05$.

immunotherapy in TNBC. TIDE analysis revealed the predicted sensitivity of three subtypes to immune checkpoint blockade therapy (Figure 3E). Clust2 had the highest TIDE score, suggesting a high possibility of immune escape to immunotherapy, which may be resulted from a high enrichment of myeloid-derived suppressor cells (MDSCs), T cell exclusion and T cell dysfunction (Figure 3E). The proportion of responders in clust2 was also the lowest compared with other two subtypes (Figure 3F).

Construction of a cuproptosis-related risk model for predicting TNBC prognosis

As three subtypes performed different molecular signatures, we then identified the DEGs between clust1 vs non-clust1, clust2 vs non-clust2, clust3 vs non-clust3. As a result, 2,723 DEGs were screened (FDR <0.05 and |log2FC| > 1.5). Then univariate Cox regression was used to further filter 1,213 DEGs, and finally 89 DEGs (prognostic genes) with 77 risk genes and 12 protective genes remained (Supplementary Figure S3A). Moreover, LASSO regression was performed on 89 genes to generate an optimal risk model. The model reached the optimal when lambda = 0.057, where 14 prognostic genes remained (Supplementary Figure S3B, C). StepAIC was further performed to optimize the prognostic model, and finally six prognostic genes were remained including *PTPRN2*, *SCARB1*, *SLC37A2*, *YES1*, *LY6D*, and *NOTCH3* (Supplementary Figure S3D). The risk model was determined according to the following formula:

$$\begin{aligned}\text{risk score} &= 0.384*\text{PTPRN2} + (-0.754*\text{SCARB1})\\ &\quad + 0.703*\text{SLC37A2} + (-0.586*\text{YES1}) + 0.264*\text{LY6D}\\ &\quad + 0.622*\text{NOTCH3}\end{aligned}$$

For each sample, a risk score was calculated according to the formula. The risk model showed a favorable performance

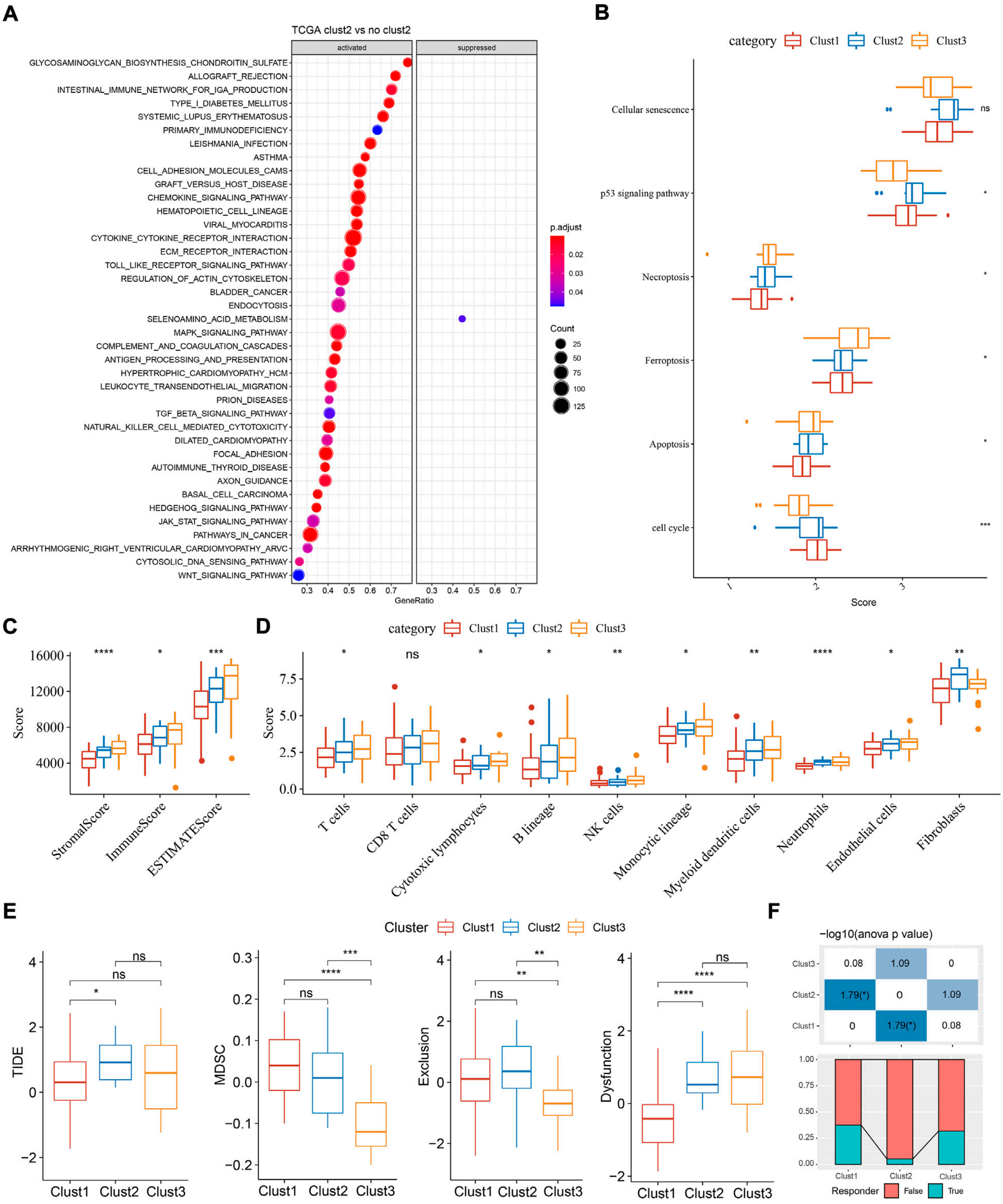


FIGURE 3

Functional analysis and immune analysis of three molecular subtypes in TCGA dataset. **(A)** GSEA result of clust2 vs non-clust2. **(B)** The ssGSEA score of six pathways related to cell death. **(C)** Stromal score and immune score calculated by ESTIMATE. **(D)** The estimated enrichment score of 10 immune cells. **(E)** TIDE analysis for predicting the sensitivity to immunotherapy. **(F)** The proportion of responders and non-responders in three subtypes. ANOVA was conducted. ns, not significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

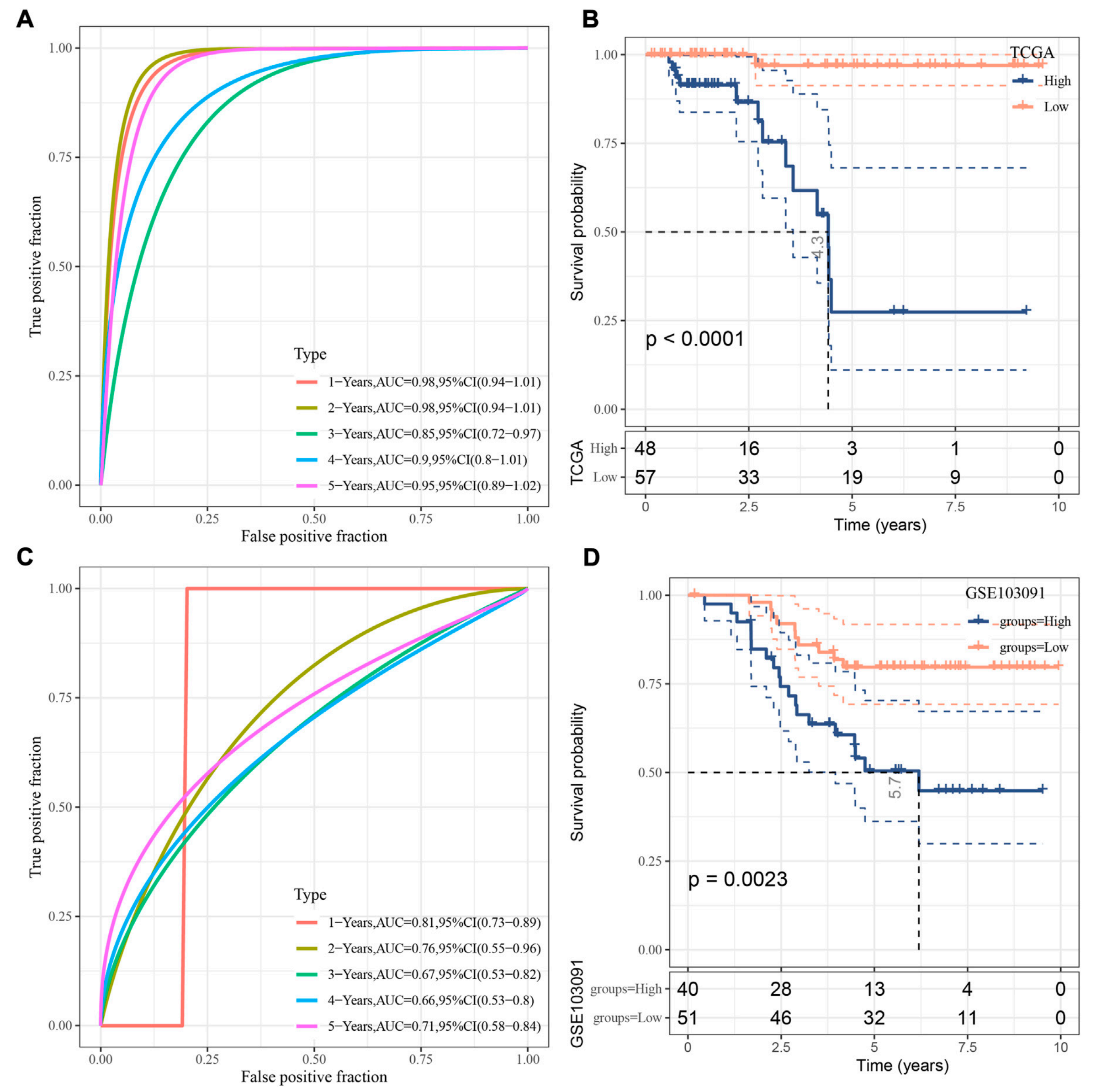


FIGURE 4
Verification of the risk model. **(A)** ROC analysis of the risk model in TCGA dataset. **(B)** Kaplan-Meier survival curve of two risk groups in TCGA dataset. **(C)** ROC analysis of the risk model in GSE103091 dataset. **(D)** Kaplan-Meier survival curve of two risk groups in GSE103091 dataset. Log-rank test was performed.

in predicting one- to 5-year PFS with AUC all over than 0.85 in TCGA dataset (Figure 4A). Determined by the optimal cut-off value of risk score, the samples were classified to different risk types (high-risk and low-risk). Kaplan-Meier survival plot showed that high- and low-risk groups had markedly different PFS (Figure 4B, $p < 0.0001$). In GSE103091 dataset, a favorable AUC of the risk model and differential prognosis between two risk groups was also observed (Figures 4C,D).

The association of risk score with clinical stages and immune infiltration

In the relation between risk score and clinical features, we found that a difference of risk score was shown between stage I+II and stage III+IV (Figure 5A). In addition, alive samples had a lower risk score than the deceased samples. Kaplan-Meier survival analysis revealed that the risk model could effectively divide samples into high- and low-risk groups

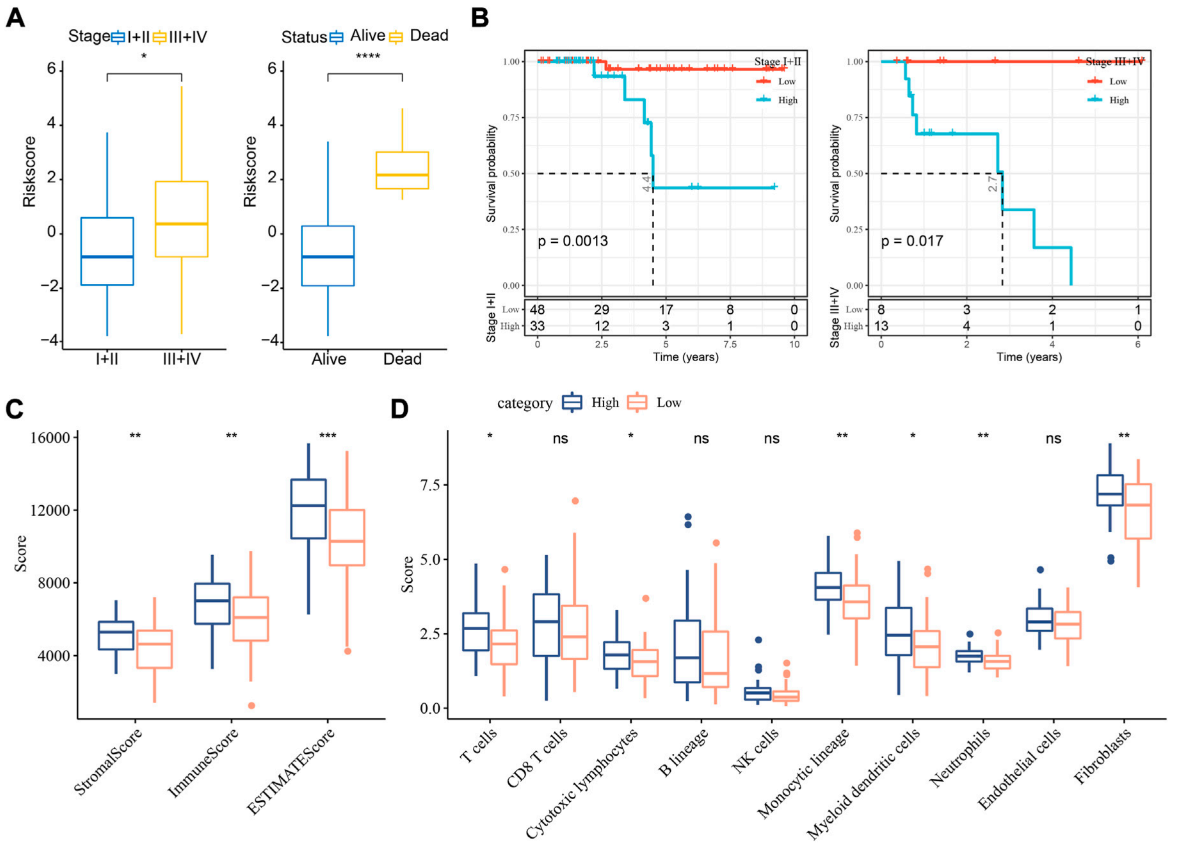


FIGURE 5
The relation of risk score to clinical features and immune infiltration in TCGA dataset. **(A)** The risk score in different stages and survival status. Student t test was conducted. **(B)** Kaplan-Meier survival analysis of high- and low-risk groups with different stages. Log-rank test was conducted. **(C)** ESTIMATE analysis for calculating stromal score and immune score of two groups. **(D)** MCP-counter analysis for calculating the enrichment score of 10 immune cells. Student t test was performed. ns, not significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

grouping by different clinical features (Figure 5B; Supplementary Figure S4A).

To understand whether a difference on tumor microenvironment was shown between two risk groups, we applied different tools, including ESTIMATE, MCP-counter, and ssGSEA, to evaluate the immune infiltration. The three tools showed consistent result that high immune infiltration was displayed in samples with high risk (Figures 5C,D, Supplementary Figure S4B). The above findings further demonstrated that cuproptosis-associated genes were possibly involved in the modulation of tumor microenvironment.

Establishing a nomogram for clinical application based on risk score and clinical characteristics

Univariate and multivariate Cox regression analysis revealed that stage and risk score were independent risk factors (Figures 6A,B). Consequently, we established a nomogram based on stage and risk score, of which risk score contributed the most to the nomogram (Figure 6C). Calibration curve showed that the predicted PFS was similar to the observed PFS (Figure 6D). Decision curve analysis (DCA) demonstrated the reliability of the nomogram and risk model (Figure 6E). Compared with other clinical characteristics, the nomogram and risk model exhibited a better performance in predicting PFS, especially long-term PFS (Figure 6F).

Discussion

An increased level of copper can result in cell death and the disruption of cupper homeostasis can lead to life-threatening diseases such as Wilson's disease and neurodegenerative disorders (Gaggelli et al., 2006; Bandmann et al., 2015). Tsvetkov et al. have revealed that

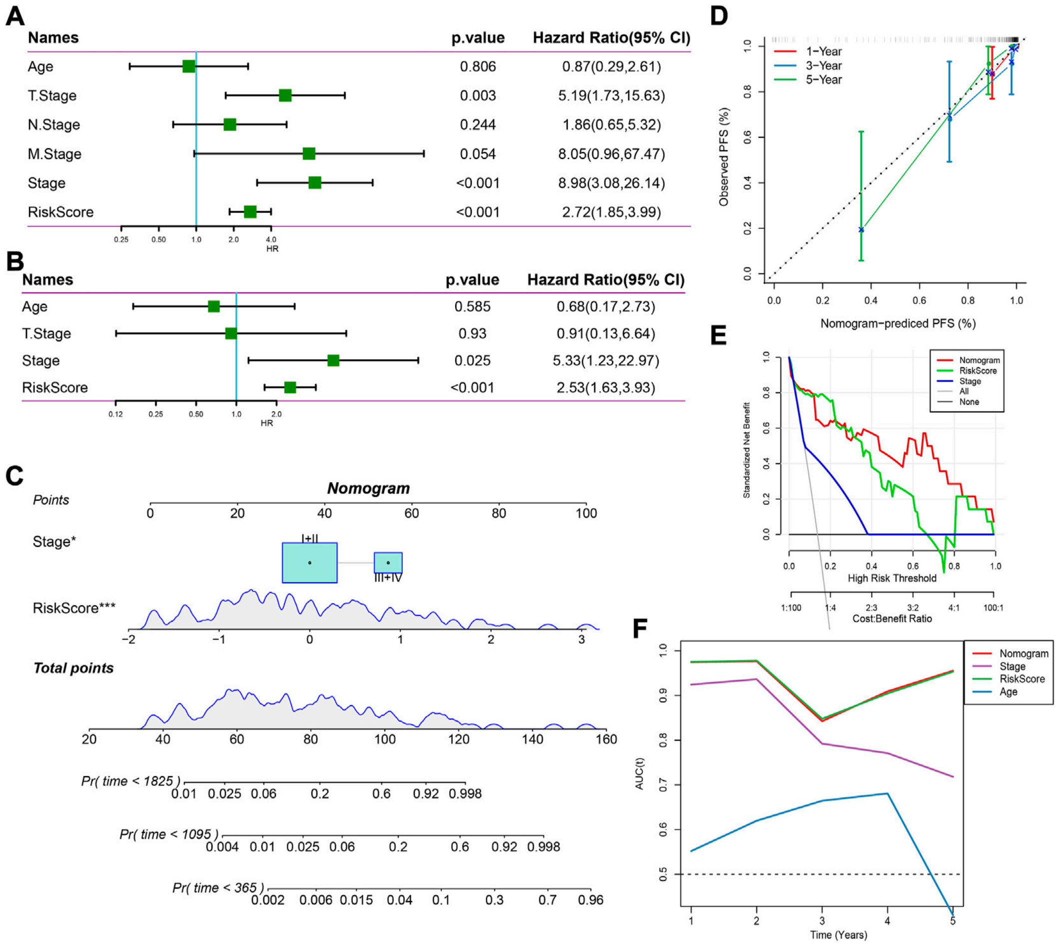


FIGURE 6
Establishing a nomogram based on the risk score. **(A,B)** Univariate **(A)** and multivariate **(B)** Cox regression analysis on age, stage, and risk score. **(C)** A nomogram for predicting death rate based on risk score and stage. **(D)** Calibration curve of 1-, 3-, and 5-year OS. **(E)** Decision curve of stage, risk score, and nomogram. **(F)** AUC of age, stage, risk score, and nomogram.

copper-induced cell death, which defined as cuproptosis, is mediated by protein lipoylation involved in tricarboxylic acid (TCA) cycle (Tsvetkov et al., 2022). Unlike other cell death forms including apoptosis, ferroptosis, necroptosis, and pyroptosis, cuproptosis functions in a new mechanism through which mitochondrial ferredoxin 1-mediated protein lipoylation leads to proteotoxic stress and ultimately cell death (Tsvetkov et al., 2022). Cupper ionophores and cupper chelators have been explored as potential anti-cancer molecules (O'Day et al., 2013; Cui et al., 2021), which inspires the research on the potential of cuproptosis in cancer treatment.

We observed a significant difference of cuproptosis score between normal and TNBC samples, suggesting the instability of cuproptosis homeostasis in cancer cells. Normal samples have a higher cuproptosis score than TNBC samples, indicating a higher activity of cupper dwindling. Therefore, we further explored the association of cuproptosis with TNBC prognosis, functional pathways, and tumor immune microenvironment through constructing molecular subtypes based on cuproptosis-associated genes. The current study have shown that the three molecular subtypes had distinct prognosis and enrichment of activated pathways. Clust2 had the worst prognosis and the highest proportion of dead samples. Notably, immune related pathways were significantly activated in clust2, such as cytokine-cytokine signaling pathway, chemokine signaling pathway, and Toll-like receptor signaling pathway, which drove a possibility that cuproptosis may participate in the modulation of immune microenvironment. Not surprisingly, three molecular subtypes demonstrated different immune infiltration and response to immune checkpoint blockade. Clust2 was predicted to have a great possibility of immune escape in immunotherapy, compared to other two subtypes, which may be resulted from T cell exclusion and T cell dysfunction. Differential analysis on three molecular subtypes illustrated that cuproptosis was involved in cancer progression and immune microenvironment.

Furthermore, we established a risk model based on cuproptosis-related genes, where six prognostic biomarkers were included (PTPRN2, SCARB1, SLC37A2, YES1, LY6D, and NOTCH3). Most of these biomarkers have been reported to promote cancer progression. PTPRN2 is a protein tyrosine phosphatase receptor, which was found to be upregulated in metastatic breast cancer and could promote cancer metastasis through PI(4,5)P2-dependent actin remodeling (Sengelaub et al., 2016). Immature isoform of PTPRN2 (proPTPRN2) expression was closely associated with lymph node-positive breast cancer and poor clinical outcome (Sorokin et al., 2015). Scavenger receptor class B member 1 (SCARB1) is a cell-surface glycoprotein mediating low density lipoprotein-cholesteryl ester (LDL-CE), which is involved in lipid internalization (Swarnakar et al., 1999). David de Gonzalo-Calvo et al. suggested that SCARB1 potentially promote CE accumulation and aggressive feature in breast cancer (de Gonzalo-Calvo et al., 2015). Proto-oncogene tyrosine-protein kinase (YES1) has been widely reported to stimulate cancer cell growth and migration in various cancer types such as lung cancer (Garmendia et al., 2019), gastric cancer (Mao et al., 2021), and breast cancer (Takeda et al., 2017), which is therefore considered as a novel therapeutic target for cancer therapy (Garmendia et al., 2022). Targeting YES1 was effective to restore the sensitivity to chemotherapeutic drugs (trastuzumab and lapatinib) in drug-resistance breast cancer cell lines (Takeda et al., 2017). Moreover, downregulation of YES1 *via* miR-133 was demonstrated to inhibit cancer cell proliferation triple-negative breast cancer cell lines (Zhang et al., 2020). Lymphocyte antigen six superfamily member D (LY6D) has been identified as a biomarker for bladder cancer and a chemoresistance marker laryngeal squamous cell carcinoma (Andersson et al., 2020; Wang et al., 2020). NOTCH3 signaling is a well-known pathway contributing to cancer development (Aburjania et al., 2018). SLC37A2 has not been reported to be involved in cancerigenesis or cancer progression.

The risk model manifested a favorable performance in predicting TNBC prognosis in the two independent datasets. Two risk groups also showed different immune infiltration, which was consistent with the result on molecular subtypes. To increase the accuracy of the risk model in predicting TNBC prognosis, we further established a nomogram that exhibited a better performance than the staging system.

Conclusion

In conclusion, this study revealed the important role of cuproptosis in TNBC development and its crosstalk with tumor immune microenvironment. We distinguished three molecular subtypes related to cuproprotiss, which had a potential to guide the personalized immunotherapy. In addition, we established a six-gene risk model with robust performance to predict TNBC prognosis.

Data availability statement

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

Author contributions

All authors contributed to this present work: [BZ] designed the study, [SW] acquired the data. [RW] drafted the manuscript, [XW] revised the manuscript. All authors read and approved the 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1022236/full#supplementary-material

References

Aburjania, Z., Jang, S., Whitt, J., Jaskula-Stzul, R., Chen, H., and Rose, J. B. (2018). The role of Notch3 in cancer. *Oncologist* 23, 900–911.

Akram, M., Iqbal, M., Daniyal, M., and Khan, A. U. (2017). Awareness and current knowledge of breast cancer. *Biol. Res.* 50, 33.

Andersson, N., Ohlsson, J., Wahlin, S., Nodin, B., Boman, K., Lundgren, S., et al. (2020). Lymphocyte antigen 6 superfamily member D is A marker of urothelial and squamous differentiation: Implications for risk stratification of bladder cancer. *Biomark. Res.* 8, 51. doi:10.1186/s40364-020-00232-1

Bandmann, O., Weiss, K. H., and Kaler, S. G. (2015). Wilson's disease and other neurological copper disorders. *Lancet. Neurol.* 14, 103–113. doi:10.1016/S1474-4422(14)70190-5

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

Bertheloot, D., Latz, E., and Franklin, B. S. (2021). Necroptosis, pyroptosis and apoptosis: An intricate game of cell death. *Cell. Mol. Immunol.* 18, 1106–1121. doi:10.1038/s41423-020-00630-3

Coleman, M. P., Quaresma, M., Berrino, F., Lutz, J. M., De Angelis, R., Capocaccia, R., et al. (2008). Cancer survival in five continents: A worldwide population-based study (concord). *Lancet. Oncol.* 9, 730–756. doi:10.1016/S1470-2045(08)70179-7

Cui, L., Gouw, A. M., Lagory, E. L., Guo, S., Attarwala, N., Tang, Y., et al. (2021). Mitochondrial copper depletion suppresses triple-negative breast cancer in mice. *Nat. Biotechnol.* 39, 357–367. doi:10.1038/s41587-020-0707-9

De Gonzalo-Calvo, D., López-Vilaró, L., Nasarre, L., Perez-Olabarria, M., Vázquez, T., Escuin, D., et al. (2015). Intratumor cholesteryl ester accumulation is associated with human breast cancer proliferation and aggressive potential: A molecular and clinicopathological study. *Bmc Cancer* 15, 460. doi:10.1186/s12885-015-1469-5

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01

Gaggelli, E., Kozlowski, H., Valensin, D., and Valensin, G. (2006). Copper homeostasis and neurodegenerative disorders (Alzheimer's, Prion, And Parkinson's Diseases And Amyotrophic Lateral Sclerosis). *Chem. Rev.* 106, 1995–2044. doi:10.1021/cr040410w

Garmendia, I., Pajares, M. J., Hermida-Prado, F., Ajona, D., Bértolo, C., Sainz, C., et al. (2019). Yes1 Drives Lung Cancer Growth And Progression And Predicts Sensitivity To Dasatinib. *Am. J. Respir. Crit. Care Med.* 200, 888–899. doi:10.1164/rccm.201807-1292OC

Garmendia, I., Redin, E., Montuenga, L. M., and Calvo, A. (2022). Yes1: A Novel Therapeutic Target And Biomarker In Cancer. *Mol. Cancer Ther.* 21, 1371–1380. doi:10.1158/1535-7163.MCT-21-0958

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene Set Variation Analysis For Microarray And Rna-Seq Data. *Bmc Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Ji, Z. H., Ren, W. Z., Wang, H. Q., Gao, W., and Yuan, B. (2022). Molecular Subtyping Based On Cuproptosis-Related Genes And Characterization Of Tumor Microenvironment Infiltration In Kidney Renal Clear Cell Carcinoma. *Front. Oncol.* 12, 919083. doi:10.3389/fonc.2022.919083

Lv, H., Liu, X., Zeng, X., Liu, Y., Zhang, C., Zhang, Q., et al. (2022). Comprehensive Analysis Of Cuproptosis-Related Genes In Immune Infiltration And Prognosis In Melanoma. *Front. Pharmacol.* 13, 930041. doi:10.3389/fphar.2022.930041

Mao, L., Yuan, W., Cai, K., Lai, C., Huang, C., Xu, Y., et al. (2021). Epha2-Yes1-Anxa2 Pathway Promotes Gastric Cancer Progression And Metastasis. *Oncogene* 40, 3610–3623. doi:10.1038/s41388-021-01786-6

O'day, S. J., Eggermont, A. M., Chiarion-Sileni, V., Kefford, R., Grob, J. J., Mortier, L., et al. (2013). Final Results Of Phase Iii Symmetry Study: Randomized, Double-Blind Trial Of Elesclomol Plus Paclitaxel Versus Paclitaxel Alone As Treatment For Chemotherapy-Naive Patients With Advanced Melanoma. *J. Clin. Oncol.* 31, 1211–1218. doi:10.1200/JCO.2012.44.5585

Plichta, J. K., Ren, Y., Thomas, S. M., Greenup, R. A., Fayanju, O. M., Rosenberger, L. H., et al. (2020). Implications For Breast Cancer Restaging Based On The 8th Edition Ajcc Staging Manual. *Ann. Surg.* 271, 169–176. doi:10.1097/SLA.0000000000003071

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

Schünemann, H. J., Lerda, D., Quinn, C., Follmann, M., Alonso-Coello, P., Rossi, P. G., et al. (2020). Breast Cancer Screening And Diagnosis: A Synopsis Of The European Breast Guidelines. *Ann. Intern. Med.* 172, 46–56. doi:10.7326/M19-2125

Sengelaub, C. A., Navrazhina, K., Ross, J. B., Halberg, N., and Tavazoie, S. F. (2016). Ptprn2 And Plcβ1 Promote Metastatic Breast Cancer Cell Migration Through Pi(4, 5)P2-Dependent Actin Remodeling. *Embo J.* 35, 62–76. doi:10.15252/embj.201591973

Shen, W., Song, Z., Xiao, Z., Huang, M., Shen, D., Gao, P., et al. (2022). Sangerbox: A Comprehensive, Interaction-Friendly Clinical Bioinformatics Analysis Platform. *Imeta* 1. doi:10.1002/Imt2.36

Sorokin, A. V., Nair, B. C., Wei, Y., Aziz, K. E., Evdokimova, V., Hung, M. C., et al. (2015). Aberrant Expression Of Proptprn2 In Cancer Cells Confers Resistance To Apoptosis. *Cancer Res.* 75, 1846–1858. doi:10.1158/0008-5472.CAN-14-2718

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

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: Globocan Estimates Of Incidence And Mortality Worldwide For 36 Cancers In 185 Countries. *Ca. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Swarnakar, S., Temel, R. E., Connelly, M. A., Azhar, S., and Williams, D. L. (1999). Scavenger Receptor Class B, Type I, Mediates Selective Uptake Of Low Density Lipoprotein Cholesteryl Ester. *J. Biol. Chem.* 274, 29733–29739. doi:10.1074/jbc.274.42.29733

Takeda, T., Yamamoto, H., Kanzaki, H., Suzawa, K., Yoshioka, T., Tomida, S., et al. (2017). Yes1 Signaling Mediates The Resistance To Trastuzumab/Lap Atinib In Breast Cancer. *Plos One* 12, E0171356. doi:10.1371/journal.pone.0171356

Tang, D., Chen, X., and Kroemer, G. (2022). Cuproptosis: A Copper-Triggered Modality Of Mitochondrial Cell Death. *Cell Res.* 32, 417–418. doi:10.1038/s41422-022-00653-7

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

Wang, J., Fan, J., Gao, W., Wu, Y., Zhao, Q., Chen, B., et al. (2020). Ly6d As A Chemoresistance Marker Gene And Therapeutic Target For Laryngeal Squamous Cell Carcinoma. *Stem Cells Dev.* 29, 774–785. doi:10.1089/scd.2019.0210

Wein, L., and Loi, S. (2017). Mechanisms Of Resistance Of Chemotherapy In Early-Stage Triple Negative Breast Cancer (Tnbc). *Breast* 34, S27–S30. doi:10.1016/j.breast.2017.06.023

Wilkerson, M. D., and Hayes, D. N. (2010). Consensusclusterplus: A Class Discovery Tool With Confidence Assessments And Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity And Stromal And Immune Cell Admixture From Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, G., Sun, J., and Zhang, X. (2022). A Novel Cuproptosis-Related Lncrna Signature To Predict Prognosis In Hepatocellular Carcinoma. *Sci. Rep.* 12, 11325. doi:10.1038/s41598-022-15251-1

Zhang, G., Wang, J., Zheng, R., Song, B., Huang, L., Liu, Y., et al. (2020). Mir-133 Targets Yes1 And Inhibits The Growth Of Triple-Negative Breast Cancer Cells. *Technol. Cancer Res. Treat.* 19, 1533033820927011. doi:10.1177/1533033820927011

Zhang, Z. (2016). Variable Selection With Stepwise And Best Subset Approaches. *Ann. Transl. Med.* 4, 136. doi:10.21037/atm.2016.03.35

Zhou, P., Qin, J., Zhou, C., Wan, G., Liu, Y., Zhang, M., et al. (2019). Multifunctional Nanoparticles Based On A Polymeric Copper Chelator For Combination Treatment Of Metastatic Breast Cancer. *Biomaterials* 195, 86–99. doi:10.1016/j.biomaterials.2019.01.007

Frontiers in Genetics

TYPE Original Research
PUBLISHED 14 November 2022
DOI 10.3389/fgene.2022.1037011

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Dan Yu,
Shanghai Changzheng Hospital, China
Zhang Zhenyu,
First Affiliated Hospital of Jinzhou
Medical University, China

*CORRESPONDENCE
Guodong Wu,
wjbzmz@sina.cn

[†]These authors have contributed equally to this work

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

RECEIVED 05 September 2022
ACCEPTED 26 October 2022
PUBLISHED 14 November 2022

CITATION
Wu G, Feng D, Zhang Z, Zhang G and Zhang W (2022), Establishment of lung adenocarcinoma classification and risk model based on necroptosis-related genes.
Front. Genet. 13:1037011.
doi: 10.3389/fgene.2022.1037011

Establishment of lung adenocarcinoma classification and risk model based on necroptosis-related genes

Guodong Wu[1*†], Dingwei Feng[2†], Ziyu Zhang[1], Gao Zhang[1] and Wei Zhang[1]

[1]Department of Thoracic and Cardiovascular Surgery, The First Hospital of Fangshan District, Beijing, China, [2]Department of Thoracic Surgery, Beijing Yanhua Hospital, Beijing, China

Lung adenocarcinoma (LUAD) is the most widely known histological subtype of lung cancer. Its classification is significant for the characteristic evaluation of patients. The aim of this research is to assess the categorization of LUAD and its risk model based on necroptosis and to investigate its potential regulatory mechanisms for diagnosing and treating LUAD. According to the expression profile data along with the clinical information related to LUAD from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), we constructed a consistency matrix through consistency clustering, and used the ConsensusClusterPlus as the measurement distance to cluster and subtype the samples, and performed gene set enrichment analysis and immune infiltration analysis. Least absolute shrinkage and selection operator (Lasso) regression was utilized for obtaining prognostic significant necroptosis phenotype-related genes. Finally, we measured each patient's riskscore (RS) and build a risk model, and predicted the effect of immunotherapy for different groups of risk factors in the model. Three molecular subtypes of LUAD were obtained by cluster analysis of necroptosis-related genes in LUAD samples. Compared with C1, C3 had a better prognosis and higher immune cell infiltration. The prognosis of the C1 subtype was poor and had a high clinical grade. The proportion of Stage II, Stage III, and Stage IV was much more in comparison with that of the other two subtypes. TP53 gene had a high mutation frequency in the C1 subtype. Gene Set Enrichment Analysis (GSEA) indicated that the aberrant pathways in the C1 and C3 subtypes mainly included some cell cycle-related pathways. In addition, seven genes were identified as related genes of necroptosis phenotype affecting prognosis. High RS had a poor prognosis, while low RS had a good prognosis. The RS was verified to have a strong ability to predict survival. LUAD can be classified by the genes linked with cell necrosis and apoptosis. The difference among various types is helpful to deepen the understanding of LUAD. In addition, a risk model was constructed based. In conclusion, this study provides potential detection targets and treatment methods for LUAD from a new perspective.

KEYWORDS

LUAD, necroptosis, immune microenvironment, risk model, immunotherapy

Introduction

Lung cancer has the highest death rate around the globe (Bray et al., 2018). Its most widely known histological subtype is the Lung adenocarcinoma (LUAD), making up about 50% of the total lung cancer cases. It has a high risk of distant metastasis at each stage (Shi et al., 2016) and is linked with increased malignancy and a worse prognosis (Gong et al., 2019; Zhou et al., 2019). LUAD treatment is based on grade and stage and is mainly determined by the evaluation of tumor histology and patient characteristics by pathologists (Wei et al., 2019). The prognosis of lung cancer is unsatisfactory even though there has been improvement in its present treatment approach (chemotherapy, surgical resection, radiotherapy, immunotherapy, and molecular targeted therapy). Even at present, the 5-year survival rate of lung cancer patients is only 4%–17%, while the 5-year survival rate of metastatic tumor patients is <5% (Hirsch et al., 2017; Arbour and Riely, 2019; Anusewicz et al., 2020). Consequently, it is very important to diagnose this disease on time along with a detailed and precise risk assessment. Most of the risk assessment and monitoring tools that are being used at present for lung cancer use the clinical features and pathological parameters, among these the most widely used approach, is TNM stratification. Though, the current tumor-node-metastasis (TNM) models are usually linked with limited confidence in lung cancer prognosis prediction, which is composed of great heterogeneity among individuals. Therefore, it is necessary to coordinate the clinicopathological features of the genome when evaluating the survival prognosis of individuals.

Necroptosis, a kind of programmed necrotic cell apoptosis, is the gatekeeper of the host against pathogen invasion. It is a recently found type of programmed cell death that unlike apoptosis is unrelated to caspase (Robinson et al., 2019). The morphological manifestations of necroptosis are cell rounding and swelling, explosive rupture of the cell membrane, cell membrane perforation, mitochondrial dysfunction, and loss of mitochondrial membrane potential (Nikoletopoulou et al., 2013). During the inhibition or low level of caspase-8, receptor-interacting protein 1 (RIP1) can use receptor-interacting protein 3 (RIP3) to develop the complex of RIP1-RIP3, therefore, stimulating the mixed spectrum of pseudokinases. Phosphorylation of mixed-lineage kinase domain-like protein (MLKL) occurs to synthesize necrotic bodies, leading to necroptosis (Vandenabeele et al., 2010). The necroptosis imbalance is also a key factor in many inflammatory diseases. Necroptosis is known to have both positive and negative effects, and it has a complicated link with cancer. Even though research shows that upon the blockage of apoptosis, necroptosis can inhibit tumor growth as well as metastasis, however, its key regulators will promote tumor growth and metastasis (Liu et al., 2021). Increasing evidence shows that necroptosis has the ability to inhibit the growth and metastasis of tumors, so it can be used as a potential method to treat cancer (Li et al., 2020a; Park et al., 2020; Tan et al., 2020). These reports have highlighted the significant involvement of necroptosis in tumorigenesis and metastasis, suggesting the potential of targeting necroptosis as a new tumor classification and treatment.

In this study, we identified stable molecular subtypes by consensus clustering using genes associated with cell necroptosis and compared the clinicopathological features, mutation features, immune features, and pathway features among subtypes. Finally, the genes linked with the prognosis score and necroptosis were found by expression difference analysis and Lasso. Then, the risk model and clinical prognosis model were constructed, which could assist in the personalized treatment of LUAD patients.

Materials and methods

Collection and processing of data

The mutation, as well as RNA-Seq data of LUAD, were taken from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/abouttcga) using TCGA GDC API. In the RNA-Seq data, we removed the samples with no information regarding clinical follow-up, survival time, and status. After selecting, a total of 500 samples of primary LUAD were obtained. Then, the Ensembl in the data was changed into a Gene symbol, and the expression of numerous gene symbols was considered the mid-value. The expression data of the GSE72094 and GSE31210 datasets were taken from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). 398 and 226 LUAD samples were included respectively after selection. For the above GEO data set, the annotation information of the corresponding chip platform was downloaded. According to the annotation information, the probe was mapped to the gene, and the single probe matching numerous genes was eliminated. When a gene was matched with multiple probes, we considered the median as the gene expression value. The current study utilized the TCGA as the training set, and GSE72094 and GSE31210 data sets were utilized as independent verification sets. In addition, our necroptosis-related genes came from previous study (Xin et al., 2022), with a total of 74 genes.

Molecular typing of necroptosis-related genes

Univariate Cox analysis by Cox function in the R package highlighted the genes substantially linked with LUAD prognosis ($p < 0.05$). A consistency matrix was constructed by ConsensusClusterPlus (Wilkerson and Hayes, 2010) to cluster and divide the samples according to these genes. The molecular subtypes of samples were provided by the expression data of genes linked with necroptosis. We carried out 500 bootstraps with the "PAM" algorithm and "1-Pearson correlation" as the distance measurement. Each bootstrap had 80% of the subjects in the training set. The cluster number was set from 2 to 10. The best division was done by measuring the consistency matrix and the cumulative distribution function (CDF), and we got the molecular subtypes of the samples.

Establishing of risk model

The differentially expressed necroptosis genes (false discovery rate (FDR) < 0.05 and |log2fold change (FC)| >1) were selected by limma package in molecular subtypes. Afterward, we chose genes that were expressed differentially and had a significant prognosis ($p < 0.05$). The proportion of genes was further reduced by Least absolute shrinkage and selection operator (Lasso) regression, and major prognostic genes related to the phenotype of necroptosis were obtained. By creating a penalty function, it can obtain a more precise model by compressing some coefficients and setting others to zero. To process data with complicated collinearity is a biased estimation that yet preserves the benefit of subset contraction. It makes variable selection during parameter estimation possible and improves the way multicollinearity in regression analysis is dealt with. The risk model was subsequently created. The prognosis risk score (RS) for individual patients was determined with: $RS = \Sigma\, \beta i \times Expi$.

Expi is referred to as the level of expression of genes linked with the prognosis of necroptosis phenotype, β is referred to as the Cox regression coefficient of the corresponding gene. The patients were sorted into RS-high and RS-low groups according to the threshold "classification." We drew the survival curve by the Kaplan-Meier method for prognosis analysis, and the significance of the difference was found with the help of the log-rank test.

Prediction of immunotherapy effect

The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm (Jiang et al., 2018) was employed for verification of the impact that immune microenvironment score (IMS) has on the prediction of clinical response of immune checkpoint inhibitors (ICIs). TIDE algorithm is a calculation method for predicting immune checkpoint blockade (ICB) reactivity by using a gene expression profile. It evaluates three types of cells that inhibit the infiltration of T cells in tumors, including myeloid-derived suppressor cells (MDSCs), tumor-associated fibroblasts (TAF), and the M2 subtypes of tumor-associated macrophages (TAMs), as well as two distinct subtypes of tumor immune escape mechanisms, including tumor-infiltrating cytotoxic T lymphocytes (CTL) dysfunction score and CTL immunosuppressive factor rejection score. The higher TIDE prediction score indicated an increased likelihood of immune escape, showing immunotherapy to be less beneficial for patients.

Gene set enrichment analysis

For understanding the pathways of various biological mechanisms in a variety of molecular subtypes, GSEA was employed for pathway analysis. We utilized all candidate gene sets present in the Hallmark (Liberzon et al., 2015) for GSEA. FDR <0.05 was taken as a significant enrichment.

Immune infiltration analysis

Cell type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) algorithm (Chen et al., 2018) (https://cibersort.stanford.edu/) was used for the quantification of 22 immune cells' relative abundance in LUAD. Simultaneously, the number of immune cells was measured with the help of the Estimation of Stromal and Immune cells in Malignant Tumor tissues utilizing Expression data (ESTIMATE) software (Yoshihara et al., 2013).

Statistical analysis

All R packages and statistical analysis were conducted in R software (4.1.1). Parameters with no specific indication were default. Statistical methods were indicated in the figure legends. $p < 0.05$ was considered as significant. ns, no significance. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

Results

Molecular typing on the basis of genes related to necroptosis

Firstly, the expression of necroptosis-related genes was taken from the expression matrix of TCGA, and 20 necroptosis genes with significant prognosis related to LUAD were selected (Figure 1A, $p < 0.05$). Patients were classified by consensus clustering in accordance with the expression profiles of these 20 genes. We finally determined that the optimal number of clusters was 3 as it gave us comparatively stable clustering outcomes (Figures 1B,C), i.e., k = 3 to get three separate molecular subtypes (Figure 1D). Further analysis revealed that there were major prognostic variations in the prognostic features of the three molecular subtypes (Figure 1E). In general, C3 showed a good prognosis, and the C1 subtype had a poor prognosis. In addition, for the GSE72094 data set, after molecular typing with the same method, it was found that there were major

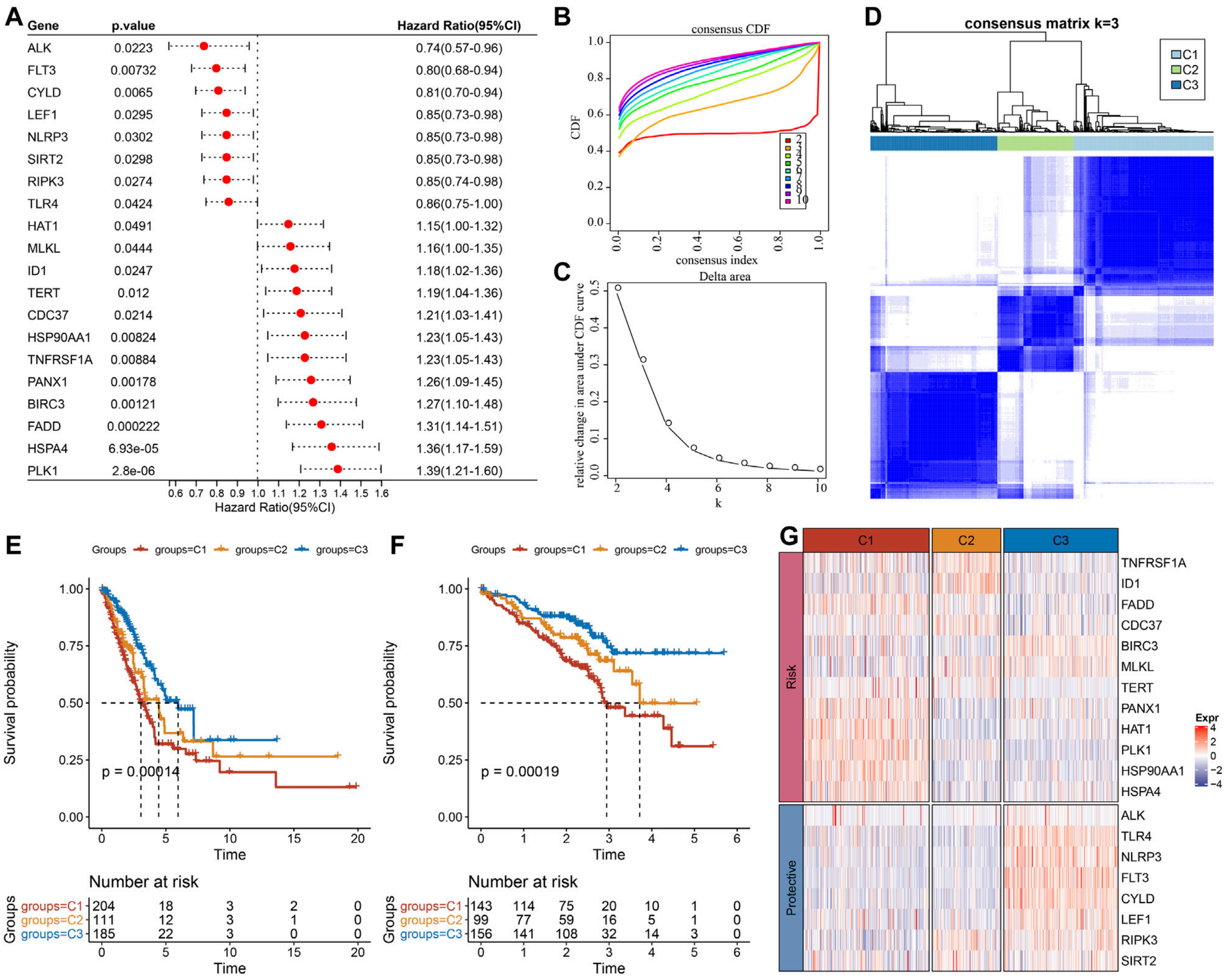


FIGURE 1

Molecular typing results according to the necroptosis-related genes. **(A)** Forest map of genes related to necroptosis with significant prognosis in TCGA cohort; **(B)** cumulative distribution function (CDF) curve of the samples in TCGA cohort; **(C)** The CDF Delta area curve of TCGA cohort sample highlights the relative change of the area under the CDF curve of each category number (k) in comparison with K-1. The horizontal axis is for k, and the vertical axis is for the relative change of the area under the CDF curve; **(D)** Sample clustering Heatmap when consensus k = 3 in TCGA queue; **(E)** Kaplan-Meier (KM) curve of overall survival (OS) prognosis of three subtypes in TCGA cohort; **(F)** Prognostic variations in the three molecular subtypes in the GSE72094 cohort; **(G)** The heat map of the expression of necroptosis genes with significant prognosis in different subtypes of TCGA.

variations in prognosis (Figure 1F), similar to the training set. At the same time, the expression differences of these 20 necroptosis genes that were substantially linked with the prognosis in separate molecular subtypes of TCGA were compared (Figure 1G).

Clinicopathological characteristics among molecular subtypes

We kept on exploring the differences in clinicopathological characteristics in separate molecular subtypes present in the TCGA cohort. In the TCGA data set, there were variations in the distribution of diverse clinical features among the three molecular subtypes. It could be observed that the C1 subtype had a high clinical grade, and male patients accounted for a large proportion of C1 and C2 subtypes (Figure 2A). Moreover, we also compared the clinicopathological characteristics of various molecular subtypes in the GSE72094 cohort and observed that the proportion of Stage II in the C1 subtype was substantially greater than that of the other two subtypes, and the proportion of Stage II was considerably reduced in comparison with that of the other two subtypes. KRAS, STK11, and TP53 gene mutations in patients with the C3 subtype were considerably reduced in comparison with those in patients with C1 and C2 subtypes,

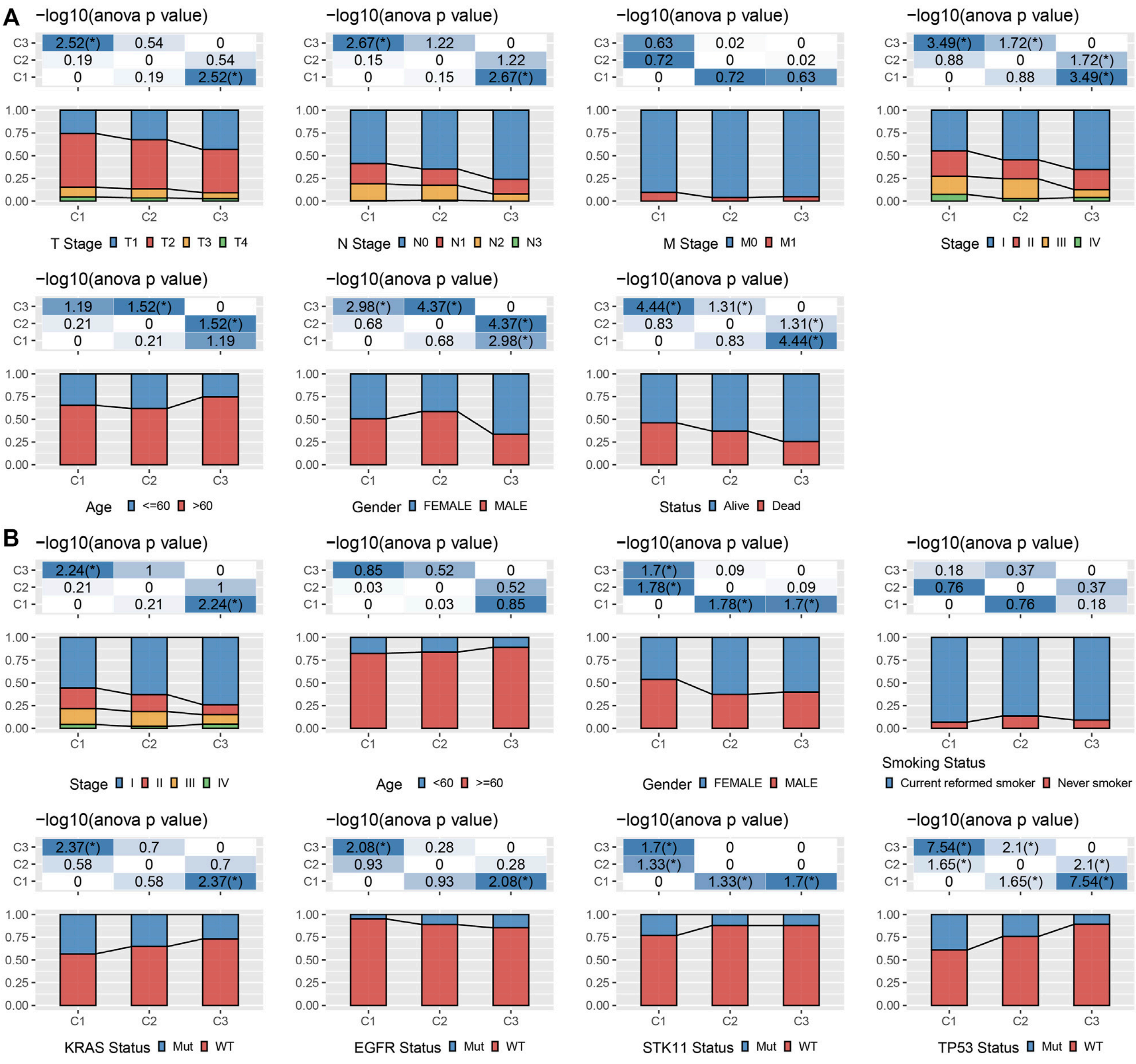


FIGURE 2
Clinicopathological properties of molecular subtypes. **(A)** The clinicopathological characteristics of molecular subtypes in TCGA cohort; **(B)** The clinicopathological characteristics of molecular subtypes of GSE72094 cohort; The lower and upper parts of the proportion are the statistical significance of the distribution difference in two pairs -log10 (*p*-value).

and EGFR gene mutations were significantly more than those in patients with C1 and C2 subtypes (Figure 2B).

Mutation characteristics among molecular subtypes

This report also explained the variations of genomic alterations in the three molecular subtypes in the TCGA cohort. Firstly, the molecular characteristic information of TCGA was obtained from the previous pan-cancer research (Thorsson et al., 2018). Among them, the C1 subtype showed a higher Homologous Recombination Defects, Aneuploidy Score, Number of Segments, Fraction Altered, and Tumor mutation burden (Figure 3A). In addition, according to 160 different immune signatures, LUAD was divided into five immune subtypes, of which the best prognosis was observed in the immune subtypes C3 and C4 and C6 had the poorest prognosis. It was discovered that, of the three types of molecular subtypes defined in this study, the C3 subtype

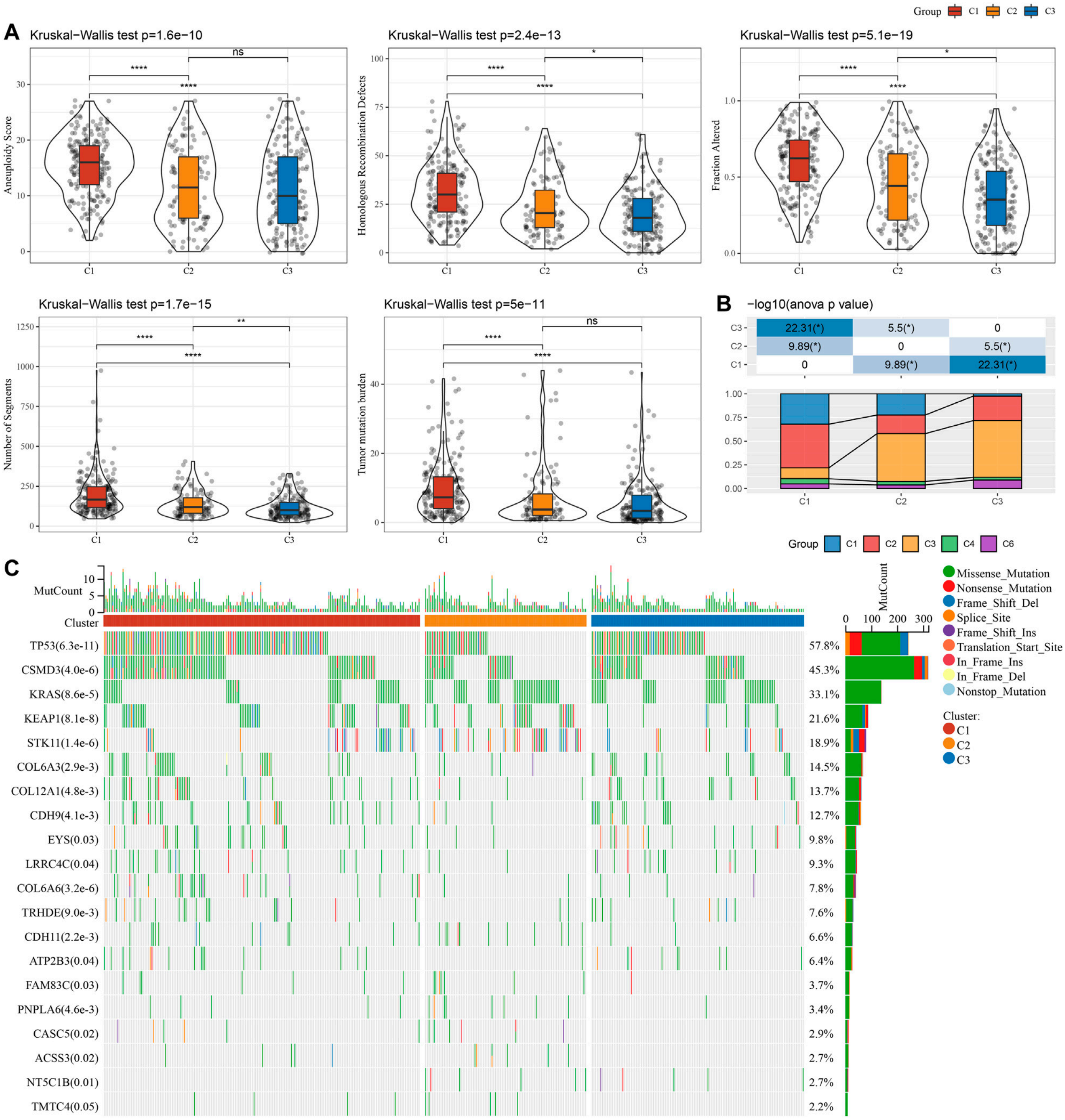


FIGURE 3
Genomic changes of molecular subtypes in TCGA cohort. **(A)** The differences in Homologous Recombination Defects, Aneuploidy Score, Fraction Altered, Number of Segments, and Tumor mutation burden among the molecular subtypes of TCGA cohort were compared; **(B)** Comparison of three molecular subtypes and immune molecular subtypes; **(C)** Somatic mutations in three molecular subtypes (chi-square test). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

described in the previous study accounted for more of the C3 subtypes described when the relationship between these five immune subtypes and the three types of molecular subtypes described by us was compared (Figure 3B). In addition, based on the correlation analysis between gene mutation and molecular subtype, we concluded that there was a major link between molecular subtype and gene mutation. TP53, CSMD3, and KRAS, and other genes had numerous

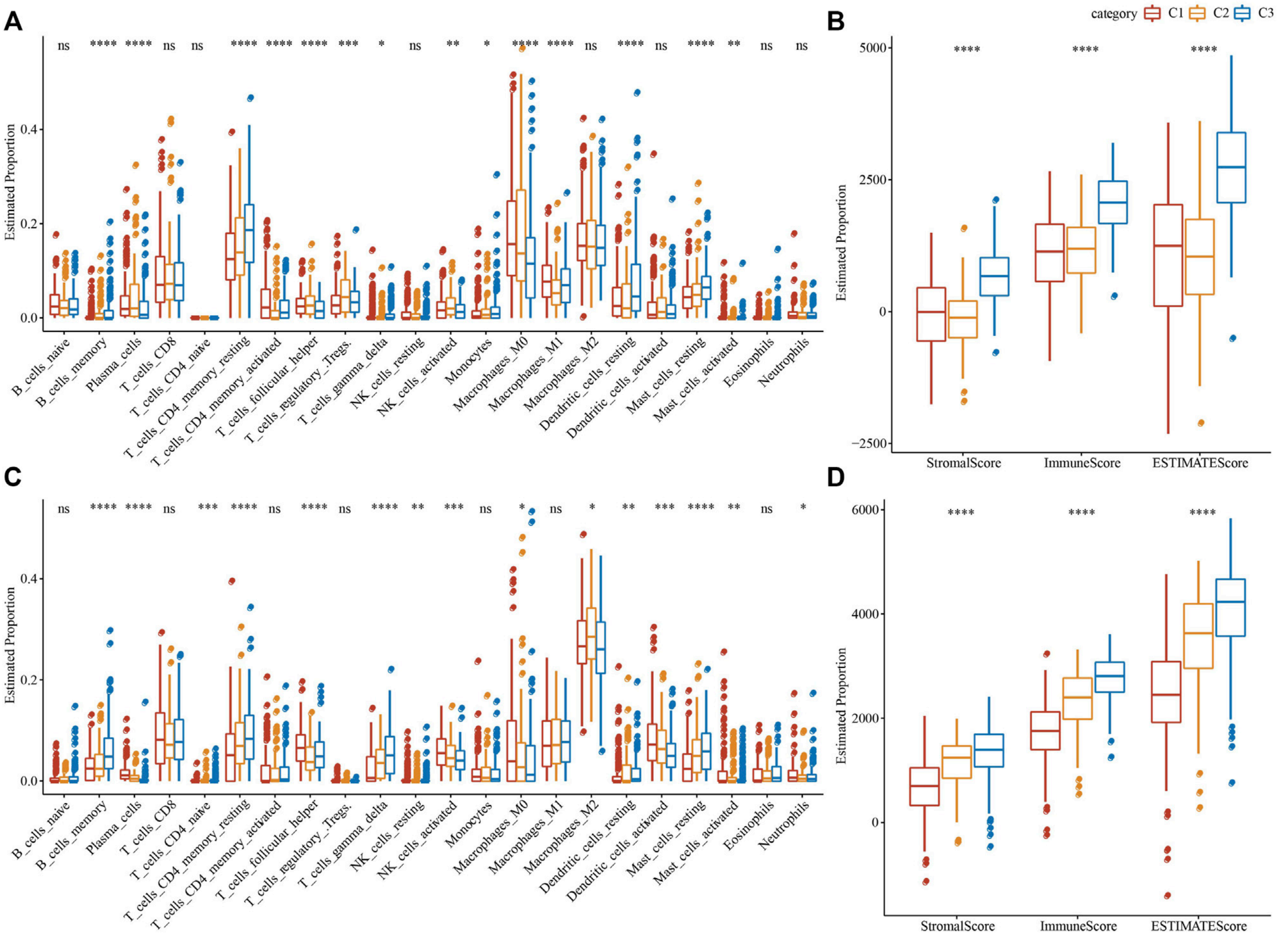


FIGURE 4
Proportion of immune cell components in two LUAD cohorts. **(A)** The variation of 22 immune cell scores among different molecular subtypes in the TCGA cohort; **(B)** The difference of ESTIMATE immune infiltration among different molecular subtypes in the TCGA cohort; **(C)** Differences in scores of 22 immune cells in various molecular subtypes in GSE72094 cohort; **(D)** Differences of ESTIMATE immune infiltration in various molecular subtypes in GSE72094 cohort. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$.

somatic mutations in LUAD, and the TP53 gene had the highest mutation frequency in the C1 subtype (Figure 3C).

Immune characteristics among molecular subtypes

To clearly understand the difference in immune microenvironment among subjects with various molecular subtypes, the level of immune cell infiltration in patients in the TCGA cohort was evaluated by the gene expression level in immune cells. Firstly, based on the relative abundance of 22 immune cells (Figure 4A), it was observed that most immune cell types had significant differences among subtypes. For example, macrophages of the M1 type were substantially more infiltrated in C1 and C3 subtypes than in C2, while regulatory T cells (Tregs) were substantially more infiltrated in C2 subtypes in comparison with the C1 and C3. Simultaneously, the "immune score" of the C3 subtype was increased in comparison with that of other subtypes, i.e., C1 and C2 subtypes, with higher immune cell infiltration (Figure 4B). In addition, by comparing the immune infiltration of the GSE72094 cohort (Figures 4C,D), a similar phenomenon to TCGA could be observed.

Pathway analysis between molecular subtypes

GSEA analysis was done to identify the differentially activated pathways in various molecular subtypes. The outcomes revealed that in comparison with the C3 subtype,

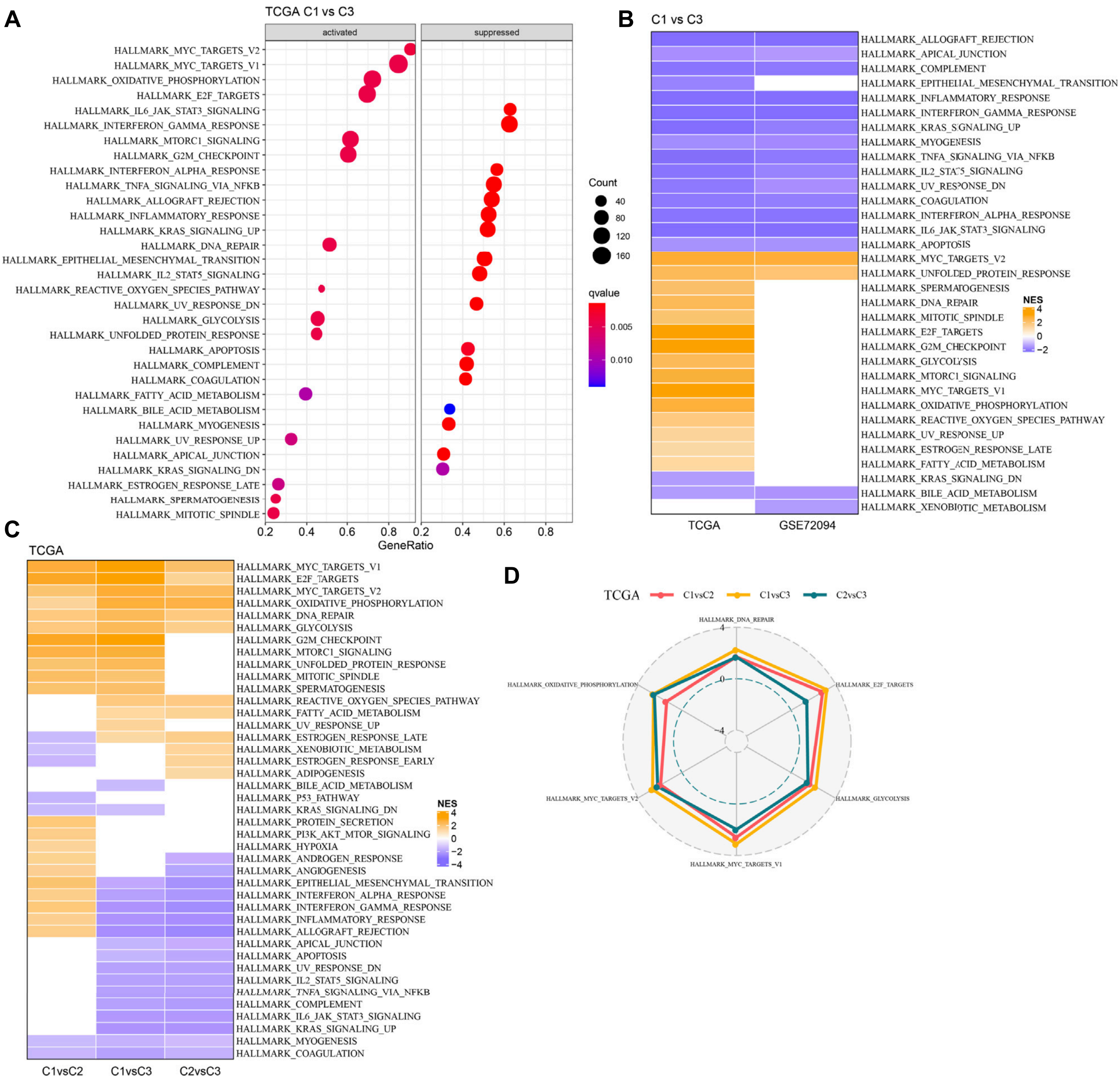


FIGURE 5
Pathway analysis between molecular subtypes. **(A)**. GSEA analysis results of C1 vs. C3 in TCGA cohort; **(B)** Bubble chart of GSEA analysis results of C1 vs. C3 subtypes in two LUAD cohorts; **(C)** Bubble chart of GSEA analysis results compared with different molecular subtypes in TCGA cohort; **(D)** Radar chart of C1 vs. C2 and C2 vs. C3 uniformly activated channels in TCGA queue.

the C1 subtype was significantly enriched in 32 pathways in the TCGA cohort and 18 pathways in the GSE72094 cohort (Figures 5A,B). Simultaneously, through the comparative analysis of abnormal pathways in C1 and C3 subtypes in various LUAD cohorts, it was found that the activated pathways mainly included some cell cycle-related pathways, such as HALLMARK_UNFOLDED_PROTEIN_RESPONSE, HALLMARK_MYC_TARGETS_V2, HALLMARK_DNA_REPAIR, HALLMARK_MITOTIC_SPINDLE, etc., while the inhibited pathways mainly included some immune-related pathways, such as HALLMARK_INFLAMMATORY_RESPONSE, HALLMARK_INTERFERON_GAMMA_RESPONSE, HALLMARK_ALLOGRAFT_REJECTION, HALLMARK_COMPLEMENT, HALLMARK_INTERFERON_ALPHA_RESPONSE, etc. (Figure 5B). Through the comparative analysis of the pathways in C1 and C2, C1 and C3 subtypes, and the differences

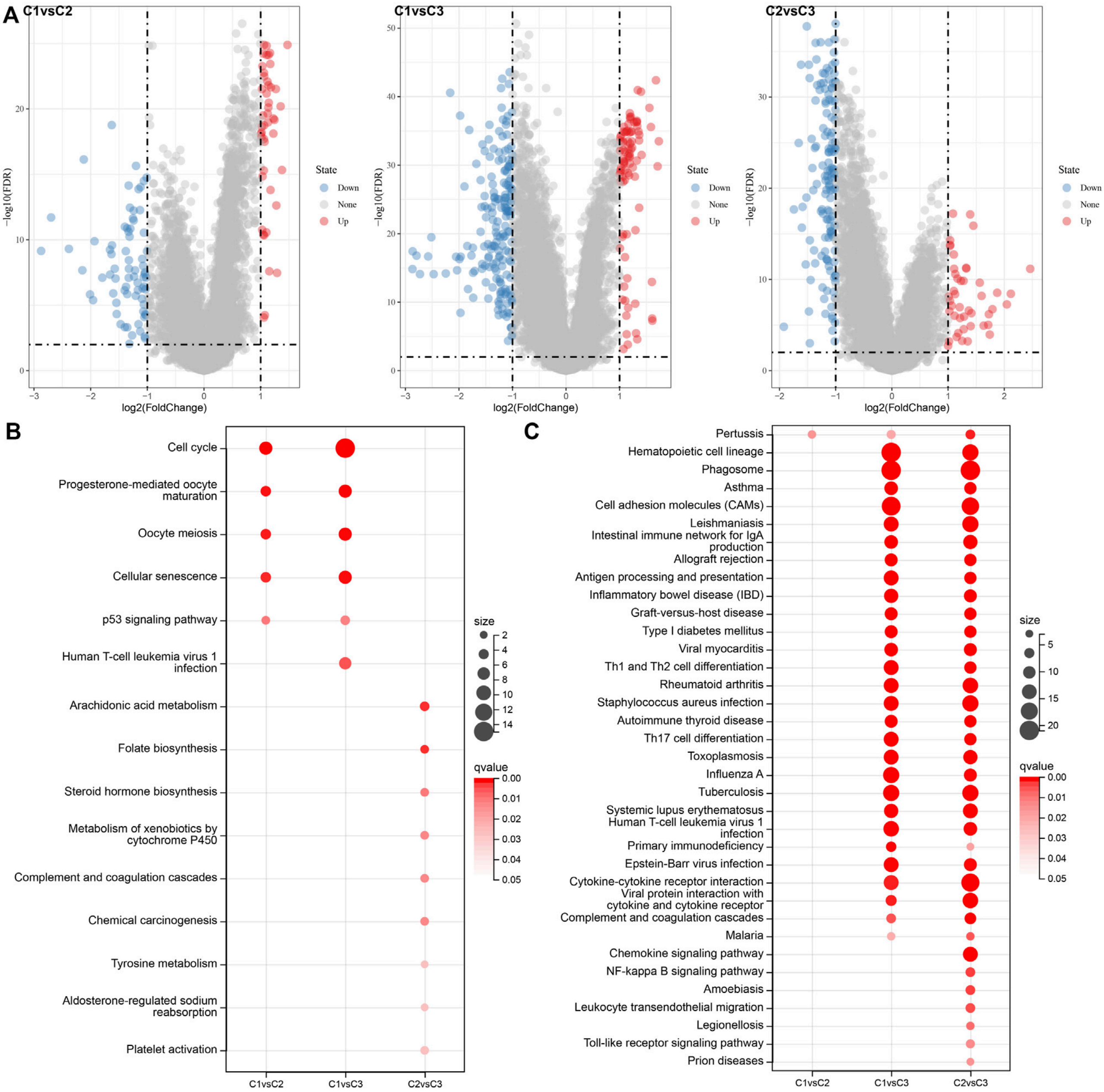


FIGURE 6
Differential expression analysis between molecular subtypes. **(A)** Volcano diagram of DEGs among TCGA molecular subtypes; **(B)** Bubble chart of KEGG function enrichment analysis results of differentially up-regulated genes among TCGA molecular subtypes; **(C)** Bubble chart of KEGG function enrichment analysis results of differentially downregulated genes among TCGA molecular subtypes.

between C2 and C3 subtypes in the TCGA cohort (Figures 5C,D), it was found that the cell cycle-related pathways in C1 patients were activated on the whole, while the immune-related pathways were inhibited. Therefore, we inferred that the necroptosis genes used for molecular typing might play a critical role in the cell cycle-related pathways and the tumor microenvironment.

Analysis of differentially expressed genes in molecular subtypes

In the analysis described above; three separate molecular subtypes were identified by the necroptosis genes with significant univariate prognosis. Next, the differentially expressed genes (DEGs) among C1 vs. C2, C1 vs. C3, and C2 vs. C3 subtypes

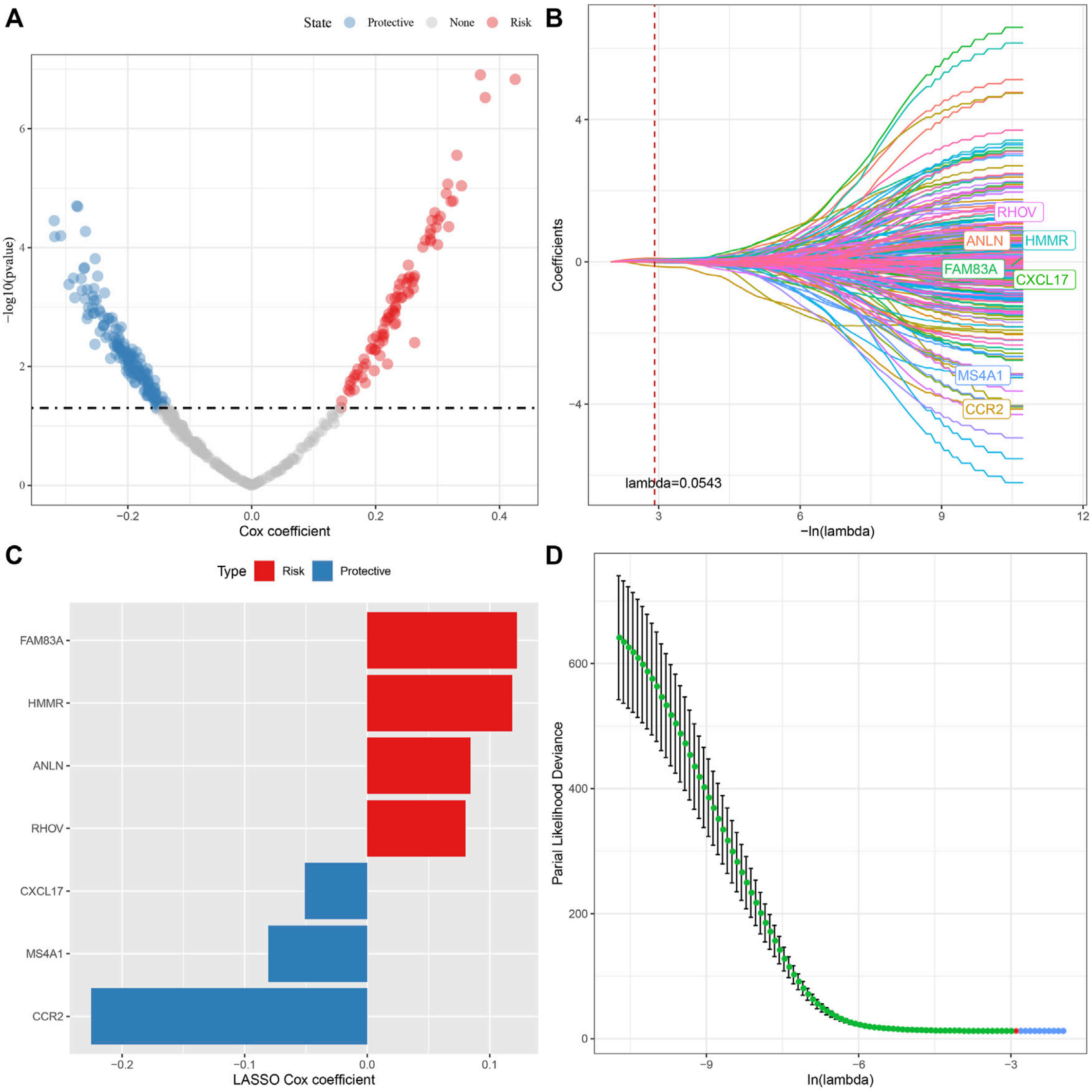


FIGURE 7
Lasso analysis of DEGs. **(A)** Analysis results of DEGs; **(B)** The locus of each independent variable changing with lambda; **(C)** Confidence interval under lambda; **(D)** Lasso coefficient distribution of the characteristics of genes linked with necroptosis.

were calculated by using the limma package. Firstly, there were 119 DEGs between the subtypes C1 and C2, including 46 highly expressed and 73 genes with low expression. Secondly, among the DEGs of C1 and C3 subtypes, there were 88 up-regulated genes and 183 down-regulated genes. Finally, among the DEGs between C2 and C3 subtypes, there were 45 up-regulated genes and 140 downregulated genes (Figure 6A). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of differentially up-regulated genes among C1 vs. C2, C1 vs. C3, and C2 vs. C3 subtypes was performed by the R software package clusterprofiler. The results showed that C1 vs. C2 and C1 vs. C3 subtypes were substantially enriched in some pathways linked with cell cycle such as cellular senescence, cell cycle, p53 signaling pathway, etc. While the C2 vs. C3 subtype was significantly enriched in metabolic-related pathways (Figure 6B). Similarly, based on the KEGG pathway enrichment analysis of differentially downregulated genes among C1 vs. C2, C1 vs. C3, and C2 vs. C3 subtypes, the results showed that there were fewer differential pathways among C1 vs. C2 subtypes, while there were more differential pathways among C1 vs. C3 and C2 v sC3 subtypes, especially among C2 vs. C3 subtypes, and the down-regulated genes among these subtypes were substantially enriched in some immune and inflammatory-related differential pathways (Figure 6C).

Identification of key necroptosis genes

405 genes were obtained by identifying DEGs among molecular subtypes. Next, these genes were assessed with univariate Cox regression analysis, along with the 242 genes that impacted the prognosis more ($p < 0.05$) were identified, including 84 “Risk” and 158 “Protective” genes (Figure 7A). Then, for Lasso regression, the “glmnet” R package was utilized to select the proportion of genes used to build risk models among the 242 genes with significant prognosis. Each independent variable’s change track was first examined. The number of independent variable coefficients that are progressively heading to 0 rose as the lambda value increased (Figure 7B). 10-fold cross-validation was utilized for creating the model, and the confidence interval under each lambda was assessed. The value of lambda = 0.0543 indicated the optimal output (Figure 7C). Therefore, we selected FAM83A, HMMR, ANLN, RHOV, CXCL17, MS4A1, and CCR2 as the related genes of necroptosis phenotype that affected the prognosis when lambda = 0.0543 (Figure 7D).

Establishment and verification of risk model

The prognostic RS related to apoptosis was calculated and normalized for each sample. At the same time, samples with RS greater than 0 were put in the RS-high group and samples having RS less than or equal to 0 were put in the RS-low group. A major difference was observed in the RS-high and -low groups ($p < 0.001$). Finally, 260 samples were put in the RS-high group, and 240 samples into the RS-low group. The RS distribution of patients in the TCGA cohort of the training set suggested that RS-high samples had a poor prognosis (Figure 8A). The ‘timeROC’ R package was utilized for assessing the Receiver Operating Characteristic (ROC) of RS prognosis classification, and the classification efficiency of prognosis prediction for 1, 3, and 5 years were analyzed respectively (Figure 8B). The outcomes revealed a high area under curve (AUC) values of prognosis prediction of the model for 1, 3, and 5 years, which were 0.73, 0.7, and 0.67, respectively. Finally, patients with higher RS showed worse overall survival in the training cohort (Figure 8C). For confirming the robustness of the clinical prognosis model prediction of necroptosis-related genes, it was verified in GSE72094 and GSE31210 cohorts. The RS of patients was calculated following the same method and samples were divided into high group when RS > 0 and samples were put in low group when RS < 0. The validation cohort had outcomes similarly to those of the training set. The prognosis of high RS was poor, while that of low RS was good (Figures 8D–G).

Riskscore in different clinicopathological characteristics

We discovered that the RS of patients with late T Stage, N Stage, M Stage, and Stage was considerably greater in comparison with that of patients with early stage by the comparison of RS distribution among the groups of clinicopathological features in the TCGA cohort. Additionally, we discovered that male patients had a higher RS. Between molecular subtypes, RS was compared and examined. When compared to the RS of C3 molecular subtype with a favorable prognosis, the RS of the C1 subtype with a much worse prognosis was significantly higher (Figure 9A). In addition, based on the comparative analysis of the prognosis differences between different clinicopathological characteristics groups in the TCGA cohort in the RS-high and -low groups defined by us, our risk groups also had good results in different clinical groups, proving the reliability of our risk groups (Figure 9B).

Characteristics of immune/pathways between riskscore groups

To clarify the variation in the immune microenvironment of patients in the RS group, the relative abundance of 22 immune cells in RS-high and -low groups in the TCGA cohort was compared. There were significant variations in 10 immune cells present in the RS-high and -low groups (Figure 10A). Such as, the abundance of resting CD4 memory T cells in the RS-low group were much higher in comparison with that in the RS-high group, while the abundance of activated CD4 memory T cells in the RS-low group was considerably lower when compared with the RS-high group. At the same time, the ESTIMATE was used for evaluating the immune cell infiltration. It was found that “the estimated immune subgroup” had higher immune infiltration (Figure 10B). This phenomenon was also observed in the GSE72094 cohort (Figures 10C,D).

Then, we studied the link of RS with 22 immune cell components in the TCGA queue and observed that RS and resting CD4 memory T cells, activated CD4 memory T cells, and resting dendritic cells along with nine others (Figure 10E).

To analyze the link of RS with the biological role of distinct samples, we chose the gene expression profile relating to the LUAD samples in the TCGA cohort and used the GSVA R package for single sample Gene Set Enrichment Analysis (ssGSEA). The score of individual samples on various functions was measured to get the ssGSEA score of individual functions related to each sample. After studying the link between these functions and calculating the RS, functional pathways greater than 0.45 were selected, from which we could see that RS and KEGG_CELL_Cycle and other cell cycle-related pathways showed a positive correlation (Figure 10F).

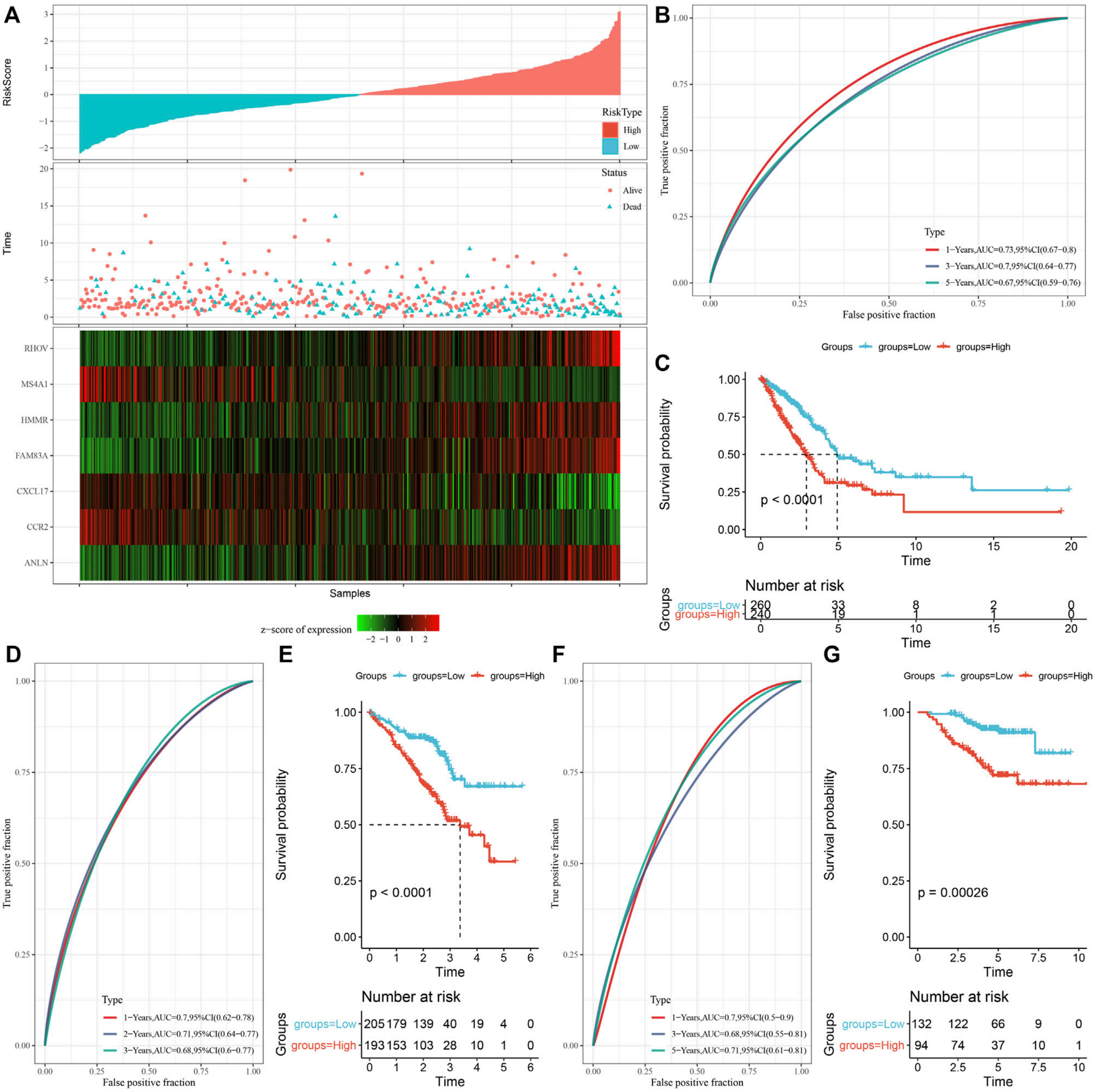


FIGURE 8
Establishment and verification of risk model. **(A)** RS, survival time, survival status, and expression of necroptosis-related genes in TCGA data set; **(B)** ROC curve and AUC classified by RS in TCGA data set; **(C)** KM survival curve distribution of RS in TCGA data set; **(D,E)** ROC curve and KM survival curve of RS in GSE72094 queue; **(F,G)** ROC curve and KM survival curve of RS in GSE31210 queue.

Differences in immunotherapy/chemotherapy between riskscore groups

In addition, whether there were differences in immunotherapy between RS groups in the TCGA cohort were analyzed. First, we did a comparison of the expression of immune checkpoints among RS groups and found that most immune checkpoint genes were differentially expressed in RS groups. On the whole, the differential expression of immune checkpoint genes, such as CTLA4, PDCD1, in the RS-low group was considerably increased in comparison with that in the RS-high group (Figure 11A). In addition, by evaluating the possible clinical impact of immunotherapy in the RS-high and -low groups, we observed that in the TCGA cohort, the RS-high

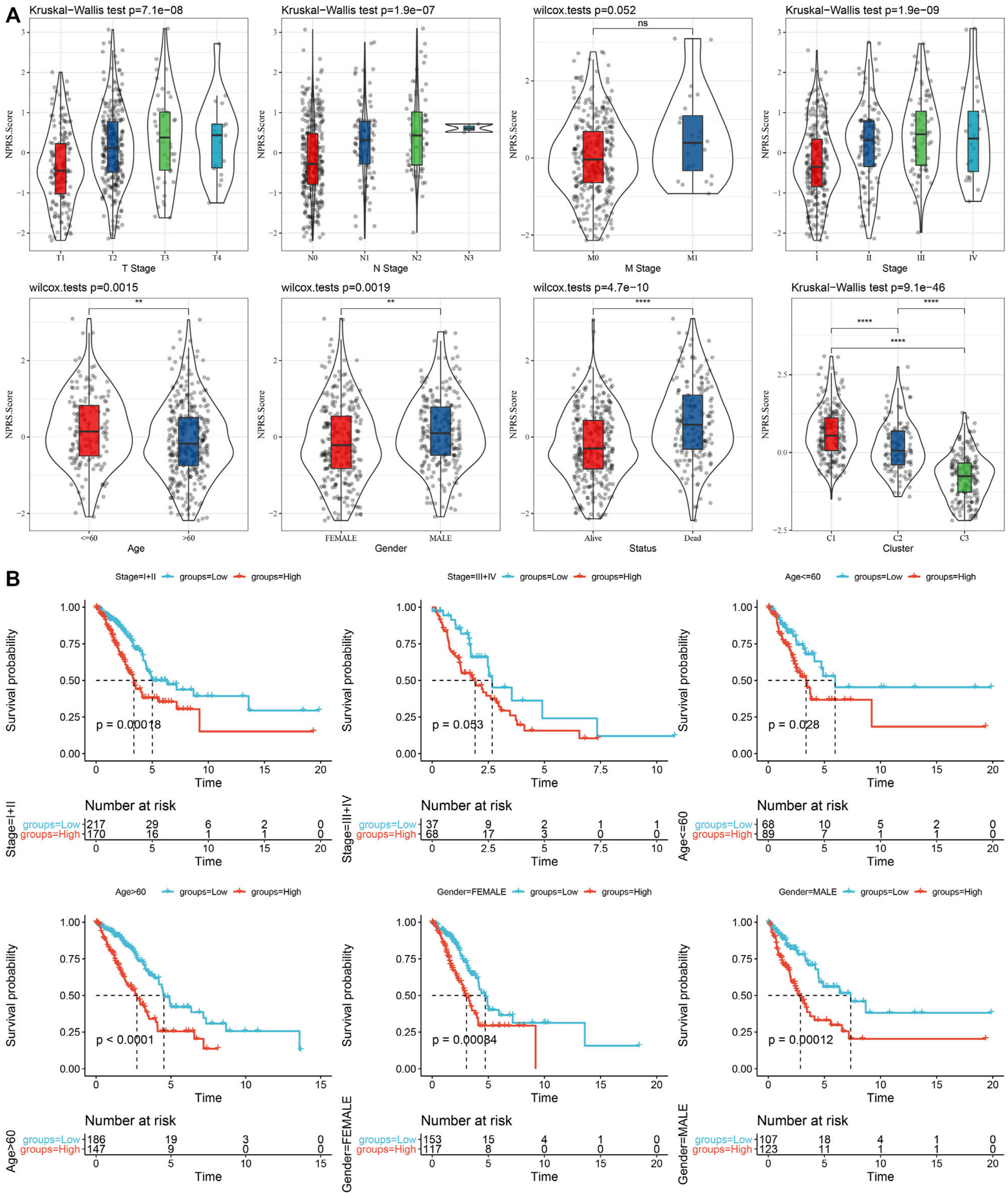


FIGURE 9
RS in different clinicopathological characteristics. **(A)** The difference of RS between different clinicopathological groups of the TCGA cohort; **(B)** KM curve between RS-high and -low groups among different clinicopathological groups of the TCGA cohort.

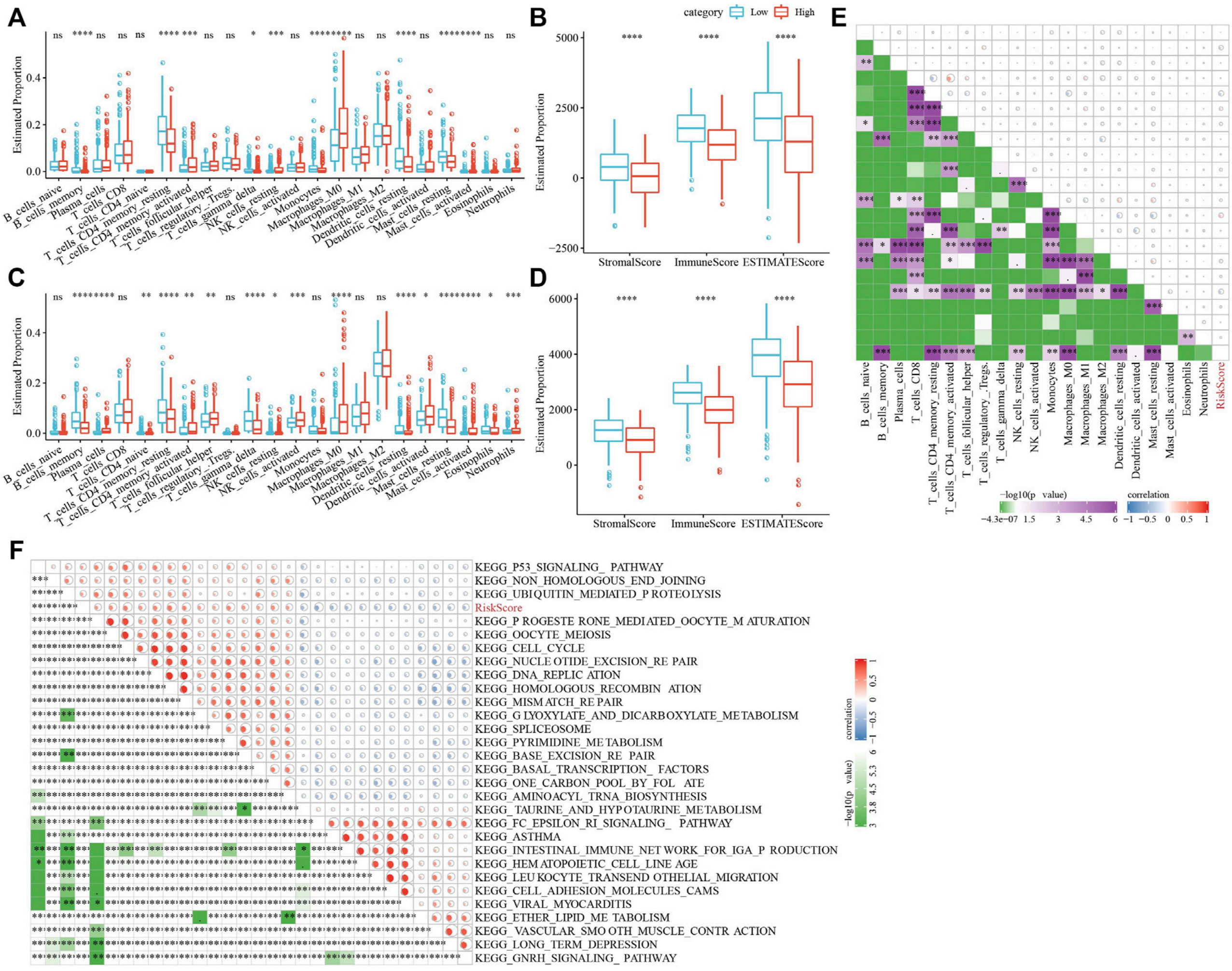


FIGURE 10
Characteristics of immune/pathways among the RS groups. **(A)** Proportion of immune cells components in TCGA cohort; **(B)** Proportion of immune cells components calculated by ESTIMATE software in TCGA cohort; **(C)** Proportion of immune cells components in TCGA cohort; **(D)** Proportion of immune cells components calculated by ESTIMATE software in TCGA cohort; **(E)** Correlation analysis between 22 immune cell components and RS in TCGA cohort; **(F)** The results of correlation analysis between KEGG pathway and RS whose correlation with RS is greater than 0.45. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$.

group had an increased TIDE score, suggesting that the possibility of immune escape for the RS-high group was more and that of benefiting from immunotherapy was less (Figure 11C). The scores of MDSC and T cell rejection were increased in the RS-high group, which might be a factor leading to the low benefit of immunotherapy in the RS-high group. Furthermore, the response of the RS group in the TCGA cohort to traditional chemotherapy drugs, such as docetaxel, vinorelbine, paclitaxel, and cisplatin was also analyzed. We discovered that the RS-high group showed more sensitivity to the stated drugs than the RS-low group (Figure 11E).

At the same time, the differences in immunotherapy and chemotherapy among RS groups in the GSE72094 cohort were analyzed, and the same phenomenon as that in the TCGA cohort was observed (Figures 11B,D,F).

Riskscore combined with clinicopathological characteristics for improving the prognosis model and survival prediction

Univariate and multivariate Cox regression analysis of RS and clinicopathological features showed that RS was the most significant prognostic factor (Figures 12A,B). For risk assessment quantification and survival probability of patients with LUAD, a nomogram was established (Figure 12C) in combination with RS,

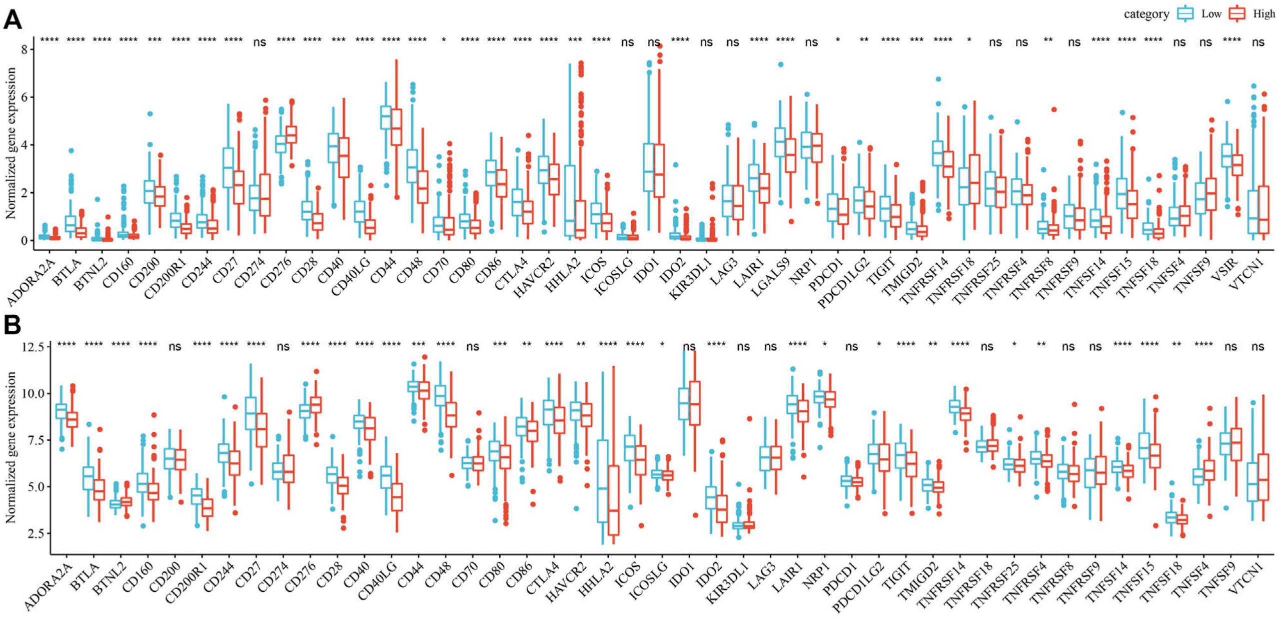


FIGURE 11
(Continued).

N Stage, T Stage, and other clinicopathological characteristics. The model results showed the greatest effect of RS on survival rate prediction. We evaluated the model for its prediction accuracy using a calibration curve, it could be observed that the predicted calibration curve of the three calibration points in 1, 3, and 5 years was near the standard curve (Figure 12D), showing the nomogram's good prediction ability. Moreover, the decision curve analysis (DCA) was also utilized for evaluating the model's reliability. It could be seen that the benefits of RS and nomogram were considerably increased when compared with that of the extreme curve. Compared with other clinicopathological characteristics, the nomogram showed the strongest ability to predict survival, followed by RS (Figure 12E).

Discussion

Necroptosis is a type of cell death related to the morphological characteristics of necrotic cells and its intrinsic signal transduction is like that of apoptotic cells. Nevertheless, necroptosis and apoptosis are different mechanisms that help in the inhibition of tumor development and metastasis (Fu et al., 2013; Lawlor et al., 2015; Newton, 2015). Numerous research conducted since the word "necroptosis" was first proposed have revealed that necroptosis can prevent tumor growth and metastasis, suggesting that it can be used for the treatment of cancer (Li et al., 2020a; Park et al., 2020; Tan et al., 2020). However, molecular typing of LUAD according to genes linked with necroptosis has not been reported. Based on necroptosis, cluster analysis was done using LUAD samples provided by the TCGA and GEO data sets, and we obtained three molecular subtypes C1, C2, and C3 of LUAD. C1 had a worse prognosis than C3, whereas C3 had a better prognosis. The matrix and immune cells enlisted and activated in the microenvironment associated with the tumor determine the tumor cells in LUAD. Immune cells and immune-related molecules also infiltrate the tumor microenvironment, which is where tumor cells proliferate, develop, and prepare for metastasis (Seong et al., 2020; Sprooten et al., 2020; Ma et al., 2021). Therefore, the variations in the immune microenvironment in subjects with different molecular subtypes were also observed, which showed that the immune score of the C3 subtype was increased in comparison with that of other subtypes, indicating that the C3 subtype had relatively high immune cell infiltration. This was supported by our prior study showing that C3 had a good prognosis and the overall survival rate of patients with a high immune score was more in comparison with that of patients having a low immune score. This finding indicates that from the beginning of tumor formation, LUAD patients with higher immune scores may have stronger adaptive immune responses than those with lower immune scores (Ma et al., 2021). Therefore, the higher immune cell content and an immune score of C3 may be one of the guarantees of a good prognosis.

Then, we calculated RS and constructed a risk model, in which RS-high samples had a worse prognosis. In addition, the

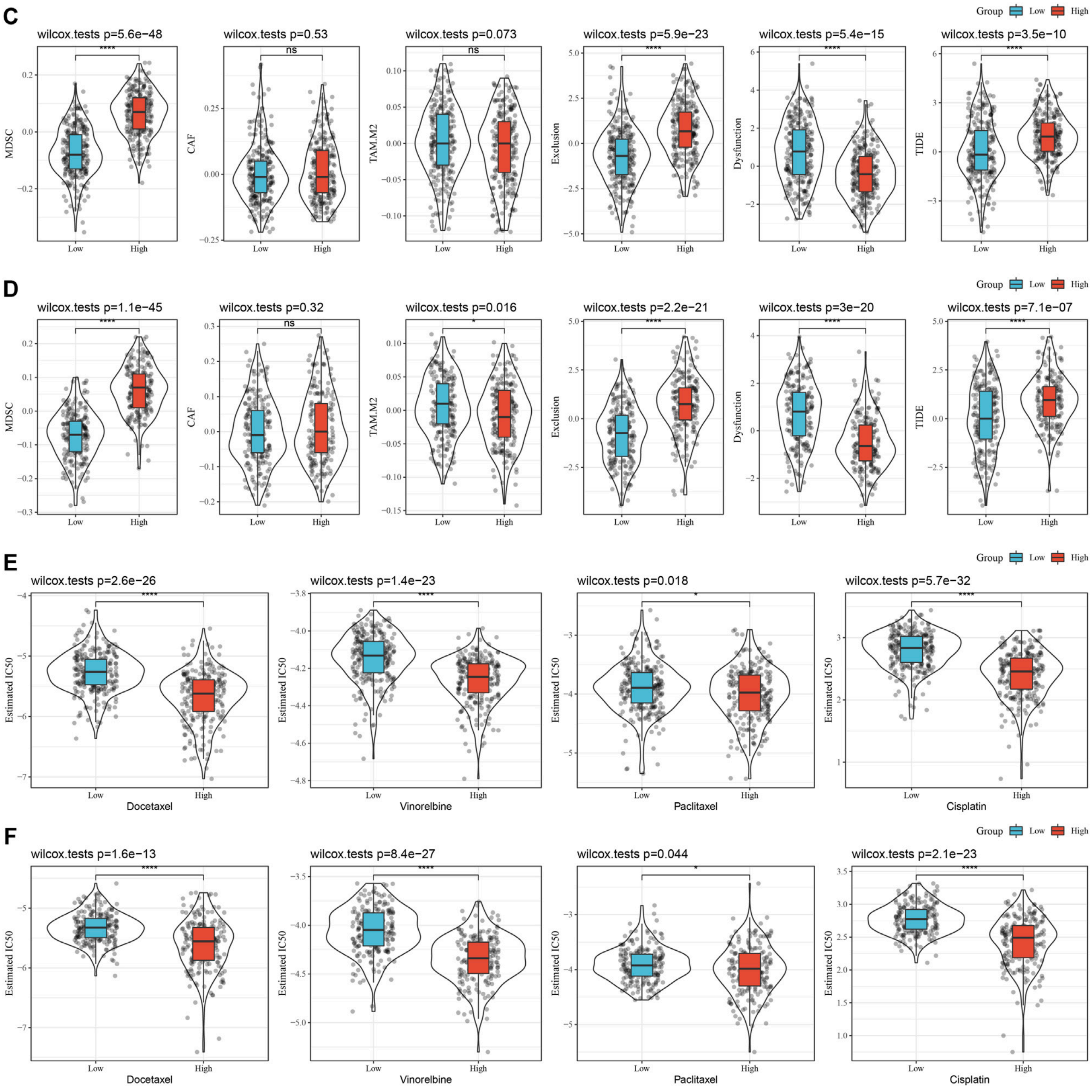


FIGURE 11
Differences in immunotherapy/chemotherapy between RS groups. **(A)** Immune checkpoints differentially expressed among different groups in the TCGA cohort; **(B)** Immune checkpoints differentially expressed among different groups in the GSE72094 cohort; **(C)** differences in TIDE analysis results among separate groups in TCGA queue; **(D)** variations in TIDE analysis results among different groups in GSE72094 queue; **(E)** Box diagram of estimated IC50 of docetaxel, vinorelbine, paclitaxel, and cisplatin in TCGA; **(F)** Box diagram of estimated IC50 of docetaxel, vinorelbine, paclitaxel, and cisplatin in GSE72094.

evaluation of potential clinical effects of immunotherapy in RS-high and -low groups showed that the RS-high group had a higher score and a higher possibility of immune escape. In other words, in comparison with the RS-low group, the patients of the RS-high group were observed to have a worse prognosis and up-regulated expression of immune checkpoints. They were more suitable for immunotherapy and were more likely to benefit from it.

Based on the role of necroptosis in the regulation of tumor immunity, we carried out the ssGSEA to find the immune status of various RS groups. Immune cells (resting CD4 memory T cells, memory B cells, and resting dendritic cells) were mostly active in

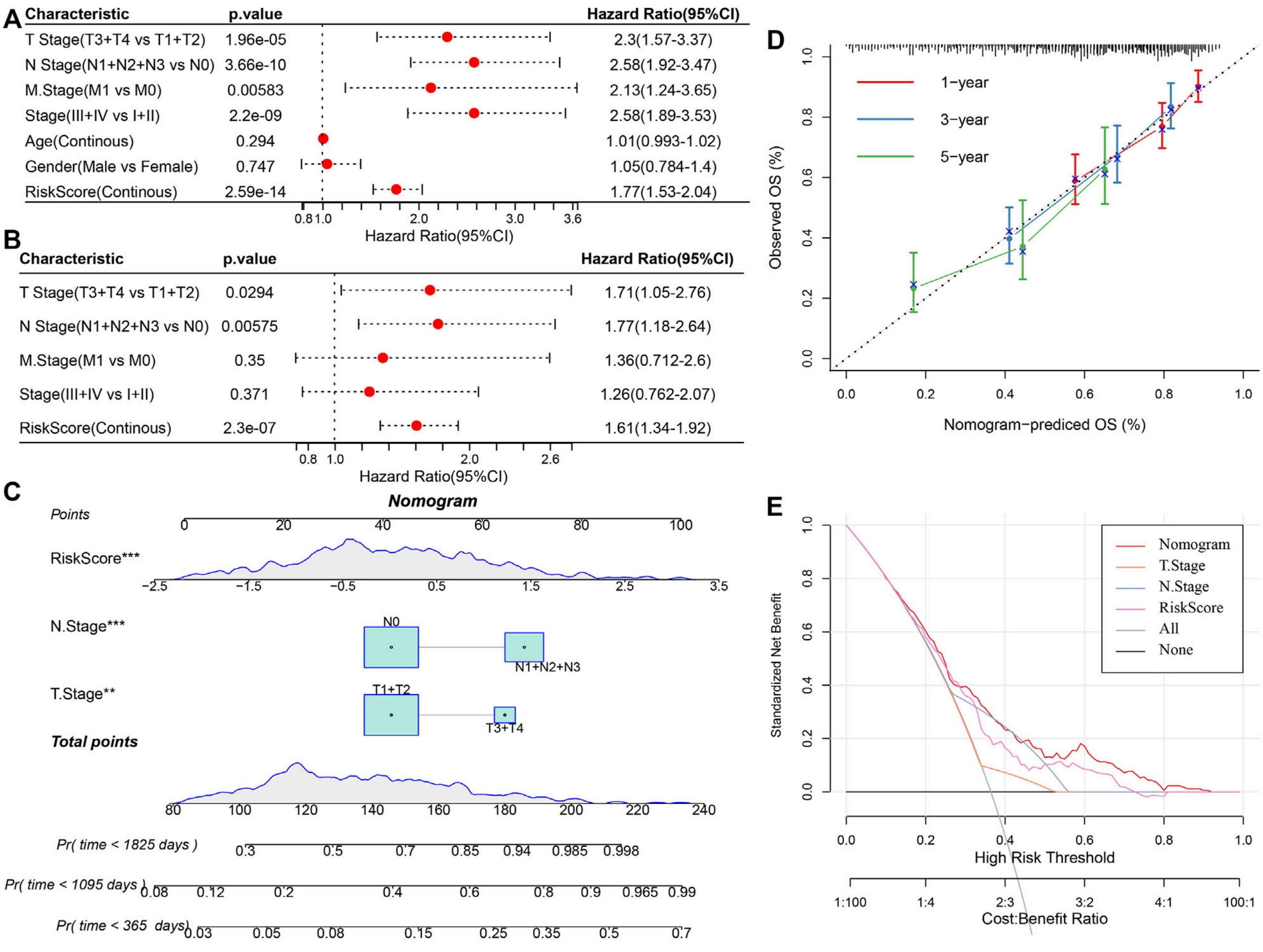


FIGURE 12
Improvement of prognosis model and survival prediction. **(A,B)** univariate and multivariate Cox analysis of RS and clinicopathological characteristics; **(C)** Nomogram model; **(D)** Calibration curve of nomogram in 1, 3, and 5 years; **(E)** Decision curve of the nomogram.

the RS-low group, among these, some were closely linked with necroptosis. But, CD8 T cells had no significance between high- and low-group. Necrotic cells could provide dendritic cells with tumor-specific antigens and inflammatory cytokines for antigen cross initiation (Sprooten et al., 2020). These outcomes indicate the possible involvement of necroptosis in the progression of LUAD by tumor immunity regulation.

In addition, the function of abnormal pathways in the C1 and C3 subtypes was analyzed, and the results showed that the activated pathways mainly included some cell cycle-related pathways, such as HALLMARK_MYC_TARGETS_V2. The relationship between cell cycles and necroptosis is inseparable. MYC pathway is one of the most significant signal pathways in the process of necroptosis. In addition, the MYC transcription factor has been shown in other studies to inhibit the formation of anti-necrotic protein of the RIPK1-RIPK3 complex (Seong et al., 2020), which fully demonstrates the reliability of our typing results. Different subtypes do have great differences in the process of necroptosis.

Studies have shown that FAM83A and FAM83A-AS1 are upregulated in LUAD in comparison with the adjacent healthy tissues. This high expression indicates poor survival and more advanced clinical stages (Wang et al., 2021a). Moreover, several studies have shown that FAM83A can be used as a prognostic characteristic and potential oncogene of LUAD (Zhang et al., 2019; Gan et al., 2020; Yu et al., 2020; Song et al., 2021). In this study, seven genes were identified as prognostic genes related to the phenotype of necroptosis, and FAM83A was one of them. In addition, these six genes (HMMR (Li et al., 2020b; Li et al., 2021), ANLN (Zhang et al., 2020; Deng et al., 2021), RHOV (Wang et al., 2021b; Zhang et al., 2021), CXCL17 (Liu et al., 2020; Wang et al., 2022), MS4A1 (Ma et al., 2020; Song et al., 2020), and CCR2 (Liu and Wu, 2021; Wan et al., 2021)) have also been studied to support their use as potential prognostic biomarkers and possible immunotherapeutic targets related to LUAD, but our research supports their involvement in the incidence and development of LUAD from the perspective of cell necroptosis.

The classification based on the correlation between necroptosis provides a novel insight for research on LUAD. In addition, nomograms were established in the LUAD cohort based on RS and clinicopathological characteristics. Compared with other clinicopathological characteristics, RS in this nomogram had significant advantages in accurately predicting the survival rate of LUAD and greatly enhanced the clinical application of gene risk typing linked with necroptosis. Therefore, the typing proposed in this study is new and meaningful, and it is found that necroptosis-related genes may be involved in it.

Nevertheless, this study has certain deficiencies and limitations. First, for external validation, the addition of more clinical databases is preferable. Moreover, further experimental evidence is still needed to confirm the conclusions of this paper. For example, experiments are needed to verify the expression differences of genes linked with necroptosis in three separate molecular subtypes. Finally, experiments should verify that the necroptosis-related genes in different subtypes have an impact on tumor progression and prognosis, and the specific study of their possible interaction and regulation mechanism needs to be further studied. To overcome the shortcomings of this research, we will recollect and expand clinical samples in the follow-up work, perform more external experiments for verifying the efficacy of this model, and conduct large-scale independent studies in the future to confirm the efficacy of this risk classification.

In summary, the predictive attributes of genes linked with necroptosis have the ability of independent prognostic prediction of LUAD patients, assist in elucidating the mechanism and process of necroptosis genes in LUAD, and provide LUAD patients with immunotherapy guidance, but additional experimental confirmation is still required in the future.

Conclusion

In a word, the stable molecular subtypes were identified by using the related genes of necroptosis through consensus clustering. Then, we chose a total of seven genes linked with the prognosis of necroptosis by analyzing the DEGs among the molecular subtypes and Lasso. Additionally, the RS model was created based on the prognosis-related genes of necroptosis. The model had strong robustness, which was independent of clinical-pathological characteristics, and played a stable predictive effect in independent data sets. Finally, we combined RS with clinicopathological characteristics to further improve the prognosis model and survival prediction. The model had high prediction accuracy and survival prediction ability.

Data availability statement

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

Author contributions

All authors contributed to this present work: GW and DF designed the study, ZZ acquired the data. GZ drafted the manuscript, WZ revised the manuscript. All authors read and approved the 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

References

Anusewicz, D., Orzechowska, M., and Bednarek, A. K. (2020). Lung squamous cell carcinoma and lung adenocarcinoma differential gene expression regulation through pathways of Notch, Hedgehog, Wnt, and ErbB signalling. *Sci. Rep.* 10 (1), 21128. doi:10.1038/s41598-020-77284-8

Arbour, K. C., and Riely, G. J. (2019). Systemic therapy for locally advanced and metastatic non-small cell lung cancer: A review. *Jama* 322 (8), 764–774. doi:10.1001/jama.2019.11058

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

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

Deng, F., Xu, Z., Zhou, J., Zhang, R., and Gong, X. (2021). ANLN regulated by miR-30a-5p mediates malignant progression of lung adenocarcinoma. *Comput. Math. Methods Med.* 2021, 9549287. doi:10.1155/2021/9549287

Fu, Z., Deng, B., Liao, Y., Shan, L., Yin, F., Wang, Z., et al. (2013). The anti-tumor effect of shikonin on osteosarcoma by inducing RIP1 and RIP3 dependent necroptosis. *BMC cancer* 13, 580. doi:10.1186/1471-2407-13-580

Gan, J., Li, Y., and Meng, Q. (2020). Corrigendum: Systematic analysis of expression profiles and prognostic significance for FAM83 family in non-small-cell lung cancer. *Front. Mol. Biosci.* 7, 653454. doi:10.3389/fmolb.2021.653454

Gong, S., Qu, X., Yang, S., Zhou, S., Li, P., and Zhang, Q. (2019). RFC3 induces epithelial-mesenchymal transition in lung adenocarcinoma cells through the Wnt/β-catenin pathway and possesses prognostic value in lung adenocarcinoma. *Int. J. Mol. Med.* 44 (6), 2276–2288. doi:10.3892/ijmm.2019.4386

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 (London, Engl.* 389 (10066), 299–311. doi:10.1016/S0140-6736(16)30958-8

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

Lawlor, K. E., Khan, N., Mildenhall, A., Gerlic, M., Croker, B. A., D'Cruz, A. A., et al. (2015). RIPK3 promotes cell death and NLRP3 inflammasome activation in the absence of MLKL. *Nat. Commun.* 6, 6282. doi:10.1038/ncomms7282

Li, M., Pan, M., Wang, J., You, C., Zhao, F., Zheng, D., et al. (2020). miR-7 reduces breast cancer stem cell metastasis via inhibiting RELA to decrease ESAM expression. *Mol. Ther. Oncolytics* 18, 70–82. doi:10.1016/j.omto.2020.06.002

Li, W., Pan, T., Jiang, W., and Zhao, H. (2020). HCG18/miR-34a-5p/HMMR axis accelerates the progression of lung adenocarcinoma. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 129, 110217. doi:10.1016/j.biopha.2020.110217

Li, X., Zuo, H., Zhang, L., Sun, Q., Xin, Y., and Zhang, L. (2021). Validating HMMR expression and its prognostic significance in lung adenocarcinoma based on data mining and bioinformatics methods. *Front. Oncol.* 11, 720302. doi:10.3389/fonc.2021.720302

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, F., and Wu, H. (2021). CC chemokine receptors in lung adenocarcinoma: The inflammation-related prognostic biomarkers and immunotherapeutic targets. *J. Inflamm. Res.* 14, 267–285. doi:10.2147/JIR.S278395

Liu, W., Xie, X., and Wu, J. (2020). Mechanism of lung adenocarcinoma spine metastasis induced by CXCL17. *Cell. Oncol.* 43 (2), 311–320. doi:10.1007/s13402-019-00491-7

Liu, Y., Chen, Q., Zhu, Y., Wang, T., Ye, L., Han, L., et al. (2021). Non-coding RNAs in necroptosis, pyroptosis and ferroptosis in cancer metastasis. *Cell Death Discov.* 7 (1), 210. doi:10.1038/s41420-021-00596-9

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

Ma, Q., Chen, Y., Xiao, F., Hao, Y., Song, Z., Zhang, J., et al. (2021). A signature of estimate-stromal-immune score-based genes associated with the prognosis of lung adenocarcinoma. *Transl. Lung Cancer Res.* 10 (3), 1484–1500. doi:10.21037/tlcr-21-223

Newton, K. (2015). RIPK1 and RIPK3: Critical regulators of inflammation and cell death. *Trends Cell Biol.* 25 (6), 347–353. doi:10.1016/j.tcb.2015.01.001

Nikoletopoulou, V., Markaki, M., Palikaras, K., and Tavernarakis, N. (2013). Crosstalk between apoptosis, necrosis and autophagy. *Biochim. Biophys. Acta* 1833 (12), 3448–3459. doi:10.1016/j.bbamcr.2013.06.001

Park, J. E., Lee, J. H., Lee, S. Y., Hong, M. J., Choi, J. E., Park, S., et al. (2020). Expression of key regulatory genes in necroptosis and its effect on the prognosis in non-small cell lung cancer. *J. Cancer* 11 (18), 5503–5510. doi:10.7150/jca.46172

Robinson, N., Ganesan, R., Hegedűs, C., Kovács, K., Kufer, T. A., and Virág, L. (2019). Programmed necrotic cell death of macrophages: Focus on pyroptosis, necroptosis, and parthanatos. *Redox Biol.* 26, 101239. doi:10.1016/j.redox.2019.101239

Seong, D., Jeong, M., Seo, J., Lee, J. Y., Hwang, C. H., Shin, H. C., et al. (2020). Identification of MYC as an antinecroptotic protein that stifles RIPK1-RIPK3 complex formation. *Proc. Natl. Acad. Sci. U. S. A.* 117 (33), 19982–19993. doi:10.1073/pnas.2000979117

Shi, J., Hua, X., Zhu, B., Ravichandran, S., Wang, M., Nguyen, C., et al. (2016). Somatic genomics and clinical features of lung adenocarcinoma: A retrospective study. *PLoS Med.* 13 (12), e1002162. doi:10.1371/journal.pmed.1002162

Song, C., Guo, Z., Yu, D., Wang, Y., Wang, Q., Dong, Z., et al. (2020). A prognostic nomogram combining immune-related gene signature and clinical factors predicts survival in patients with lung adenocarcinoma. *Front. Oncol.* 10, 1300. doi:10.3389/fonc.2020.01300

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

Sprooten, J., De Wijngaert, P., Vanmeerbeerk, I., Martin, S., Vangheluwe, P., Schlenner, S., et al. (2020). Necroptosis in immuno-oncology and cancer immunotherapy. *Cells* 9 (8), E1823. doi:10.3390/cells9081823

Tan, H. Y., Wang, N., Chan, Y. T., Zhang, C., Guo, W., Chen, F., et al. (2020). ID1 overexpression increases gefitinib sensitivity in non-small cell lung cancer by activating RIP3/MLKL-dependent necroptosis. *Cancer Lett.* 475, 109–118. doi:10.1016/j.canlet.2020.01.025

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48 (4), 812–830.e14. e14. doi:10.1016/j.immuni.2018.03.023

Vandenabeele, P., Galluzzi, L., Vanden Berghe, T., and Kroemer, G. (2010). Molecular mechanisms of necroptosis: An ordered cellular explosion. *Nat. Rev. Mol. Cell Biol.* 11 (10), 700–714. doi:10.1038/nrm2970

Wan, Y., Wang, X., Liu, T., Fan, T., Zhang, Z., Wang, B., et al. (2021). Prognostic value of CCR2 as an immune indicator in lung adenocarcinoma: A study based on tumor-infiltrating immune cell analysis. *Cancer Med.* 10 (12), 4150–4163. doi:10.1002/cam4.3931

Wang, G., Li, X., Yao, Y., Jiang, Z., Zhou, H., Xie, K., et al. (2021). FAM83A and FAM83A-AS1 both play oncogenic roles in lung adenocarcinoma. *Oncol. Lett.* 21 (4), 297. doi:10.3892/ol.2021.12558

Wang, K., Li, R., Zhang, Y., Qi, W., Fang, T., Yue, W., et al. (2022). Prognostic significance and therapeutic target of CXC chemokines in the microenvironment of lung adenocarcinoma. *Int. J. Gen. Med.* 15, 2283–2300. doi:10.2147/IJGM.S352511

Wang, Y., Zhang, G., and Wang, R. (2021). Six CT83-related genes-based prognostic signature for lung adenocarcinoma. *Comb. Chem. High. Throughput Screen.* 25, 1565–1575. doi:10.2174/1871520621666210713112630

Wei, J. W., Tafe, L. J., Linnik, Y. A., Vaickus, L. J., Tomita, N., and Hassanpour, S. (2019). Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci. Rep.* 9 (1), 3358. doi:10.1038/s41598-019-40041-7

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

Xin, S., Mao, J., Duan, C., Wang, J., Lu, Y., Yang, J., et al. (2022). Identification and quantification of necroptosis landscape on therapy and prognosis in kidney renal clear cell carcinoma. *Front. Genet.* 13, 832046. doi:10.3389/fgene.2022.832046

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

Yu, J., Hou, M., and Pei, T. (2020). FAM83A is a prognosis signature and potential oncogene of lung adenocarcinoma. *DNA Cell Biol.* 39 (5), 890–899. doi:10.1089/dna.2019.4970

Zhang, D., Jiang, Q., Ge, X., Shi, Y., Ye, T., Mi, Y., et al. (2021). RHOV promotes lung adenocarcinoma cell growth and metastasis through JNK/c-Jun pathway. *Int. J. Biol. Sci.* 17 (10), 2622–2632. doi:10.7150/ijbs.59939

Zhang, J., Sun, G., and Mei, X. (2019). Elevated FAM83A expression predicts poorer clinical outcome in lung adenocarcinoma. *Cancer Biomark.* 26 (3), 367–373. doi:10.3233/CBM-190520

Zhang, L., He, M., Zhu, W., Lv, X., Zhao, Y., Yan, Y., et al. (2020). Identification of a panel of mitotic spindle-related genes as a signature predicting survival in lung adenocarcinoma. *J. Cell. Physiol.* 235 (5), 4361–4375. doi:10.1002/jcp.29312

Zhou, X., Hua, D., Gao, C., Zhang, Y., Qiu, L., and Wang, L. (2019). Icotinib and pemetrexed in treatment of lung adenocarcinoma and the effects on prognostic survival rate of patients. *Oncol. Lett.* 18 (4), 4153–4159. doi:10.3892/ol.2019.10763

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fgene.2022.1038585

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Yufeng Lv,
Foresea Life Insurance Guangxi Hospital, China
Steven Mo,
YuanDong International Academy Of Life Sciences, China

*CORRESPONDENCE
Chunjiang Yu,
yuchj1972@163.com

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

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

RECEIVED 07 September 2022
ACCEPTED 31 October 2022
PUBLISHED 24 November 2022

CITATION
Wang L, Yu C, Tao Y, Yang X, Jiang Q, Yu H and Zhang J (2022), Transcriptome analysis reveals potential marker genes for diagnosis of Alzheimer's disease and vascular dementia.
Front. Genet. 13:1038585.
doi: 10.3389/fgene.2022.1038585

Transcriptome analysis reveals potential marker genes for diagnosis of Alzheimer's disease and vascular dementia

Li Wang[1†], Chunjiang Yu[2*†], Ye Tao[3], Xiumei Yang[4], Qiao Jiang[5], Haiyu Yu[6] and Jiejun Zhang[7]

[1]Department of Geriatrics, The Second Affiliated Hospital of the Harbin Medical University, Harbin, China, [2]Department of Neurology, The Second Affiliated Hospital of the Harbin Medical University, Harbin, China, [3]Department of Neurology, The First Hospital of SuiHua City, Suihua, China, [4]Department of Cardiovascularology, The Fifth Hospital of the Harbin City, Harbin, China, [5]Department of Neurology, The Fifth People's Hospital of the Dalian City, Dalian, China, [6]Rehabilitation Department of Jiamusi Center Hospital, Jiamusi, China, [7]Department of Neurology, Hebei Yanda Hospital, Hebei, China

Alzheimer's disease (AD) and vascular dementia (VD) are the two most common forms of dementia, share similar symptoms, and are sometimes difficult to distinguish. To investigate the potential mechanisms by which they differ, we identified differentially expressed genes in blood and brain samples from patients with these diseases, and performed weighted gene co-expression network analysis and other bioinformatics analyses. Weighted gene co-expression network analysis resulted in mining of different modules based on differences in gene expression between these two diseases. Enrichment analysis and generation of a protein-protein interaction network were used to identify core pathways for each disease. Modules were significantly involved in cAMP and AMPK signaling pathway, which may be regulated cell death in AD and VD. Genes of cAMP and neurotrophin signaling pathways, including *ATP1A3*, *PP2A*, *NCEH1*, *ITPR1*, *CAMKK2*, and *HDAC1*, were identified as key markers. Using the least absolute shrinkage and selection operator method, a diagnostic model for AD and VD was generated and verified through analysis of gene expression in blood of patients. Furthermore, single sample gene set enrichment analysis was used to characterize immune cell infiltration into brain tissue. That results showed that infiltration of DCs and pDCs cells was increased, and infiltration of B cells and TFH cells was decreased in the brain tissues of patients with AD and VD. In summary, classification based on target genes showed good diagnostic efficiency, and filled the gap in the diagnostic field or optimizes the existing diagnostic model, which could be used to distinguish between AD and VD.

KEYWORDS

Alzheimer's disease, vascular dementia, potential mechanism, WGCNA, immune cell infiltration

Introduction

Damage to neuronal structure may cause loss of nervous system function, which can lead to neurodegenerative disease. Alzheimer's disease (AD) is one of the most common neurodegenerative diseases worldwide (Lane et al., 2018). Alzheimer's disease is characterized by the presence of extracellular amyloid plaques caused by abnormal APP processing, resulting in β-amyloid peptide aggregation (Calsolaro and Edison, 2016). There is no cure for AD, disease progression cannot be reversed, and symptoms gradually worsen until patients lose their ability to care for themselves. Given the prolonged course of disease progression, AD results in incredible suffering for patients and their families, and places an enormous burden on healthcare systems. Previous studies have confirmed that the pathogenesis of AD can include genetic factors. Mutations in *APP*, *PSEN1*, and *PSEN2* have been shown to play key roles in familial AD (Lanoiselee et al., 2017). Advances in human disease research have shown that many complex diseases are caused by multiple genes. These genes interact to form a network that collectively influences the pathogenesis of diseases (Ding et al., 2019). Therefore, gene set risk assessment is viewed as a more accurate and effective method to study the genetic basis and mechanisms of complex diseases.

Vascular dementia (VD) is also a common form of dementia. The symptoms of VD are similar to those of AD, which can often complicate differential diagnosis (Uwagbai and Kalish, 2022). In some patients, VD and AD may coexist, resulting in a pathological condition known as mixed dementia. The etiology of dementia is complex, and treatment is difficult. To data, these biomarkers play a vital role for diagnosis and prognosis of AD or VD. Studies have demonstrated that identifying REPS1 as a candidate therapeutic biomarker in AD and VD (Luo et al., 2022). RBM8A (Zou et al., 2019) and YKL-40 (Mavroudis et al., 2021) were significantly associated with AD pathophysiology. Furthermore, toll-like receptor 2 (TLR2) is the hub gene that may participate in the course of VD (Wang et al., 2022). Therefore, exploration of the pathogenesis and biomarkers of VD/AD-induced dementia could deepen understanding of dementia, which may aid in diagnosis and improve choice of treatment strategies.

Studies have indicated that the multifactorial pathophysiology of dementia is not restricted to neuronal cells, and the immune system may play a key role (Heneka et al., 2015). For example, during the AD onset, T lymphocytes may infiltrate into brain tissue *via* the choroid plexus and participate in adaptive immune response. CD8 T lymphocytes were detected in the cerebrospinal fluid (CSF) of 11 patients with AD (Lueg et al., 2015). Other types of immune cells, including monocytes, macrophages, neutrophils, and T cells from the peripheral blood, were found to be broadly involved in the pathogenesis of AD (Polfliet et al., 2001; Ziegler-Heitbrock, 2007; Baik et al., 2014; Gate et al., 2020). Moreover, differences in levels of lymphocyte subsets were found in the brains of patients with different types of dementia, and a

TABLE 1 Clinical information of the sequencing data.

Library name	Sex	Age	Disease type	Blood pressure	Blood homocysteine	C Reactive protein	Fasting blood-glucose	Cholesterol total	Glycerin trilaurate	High density lipoprotein cholesterol	Low density lipoprotein cholesterol
A-D1	Woman	86	AD + DM	120/70	10.29	14.17	6.73	5.2	1.5	1.52	2.9
V1	Woman	83	VD	130/80	10.46	1.47	6	4.61	1.38	1.21	2.76
V2	Woman	83	VD + DM	165/89	11.16	1.15	5.51	2.52	0.71	1.29	0.83
V-D1	Woman	84	AD	152/88	9.01	10.89	7.5	4.88	2.22	1.14	2.93
A1	man	87	AD + DM	94/57	14.03	7.44	5.9	5.68	4.3	0.89	2
A-D2	man	76	VD + DM	160/100	13.83	9.93	7.48	5.26	2.16	1.26	3.86
V3	man	50	VD + DM	150/90	10.09	1.63	4.19	3.38	1.44	1.01	1.6
V-D2	man	73	VD	146/89	19.92	5.43	7.74	4.54	8.73	0.76	1.43
V7	Woman	61	VD	146/78	9.75	1.98	6.79	4.46	1.33	0.99	2.75

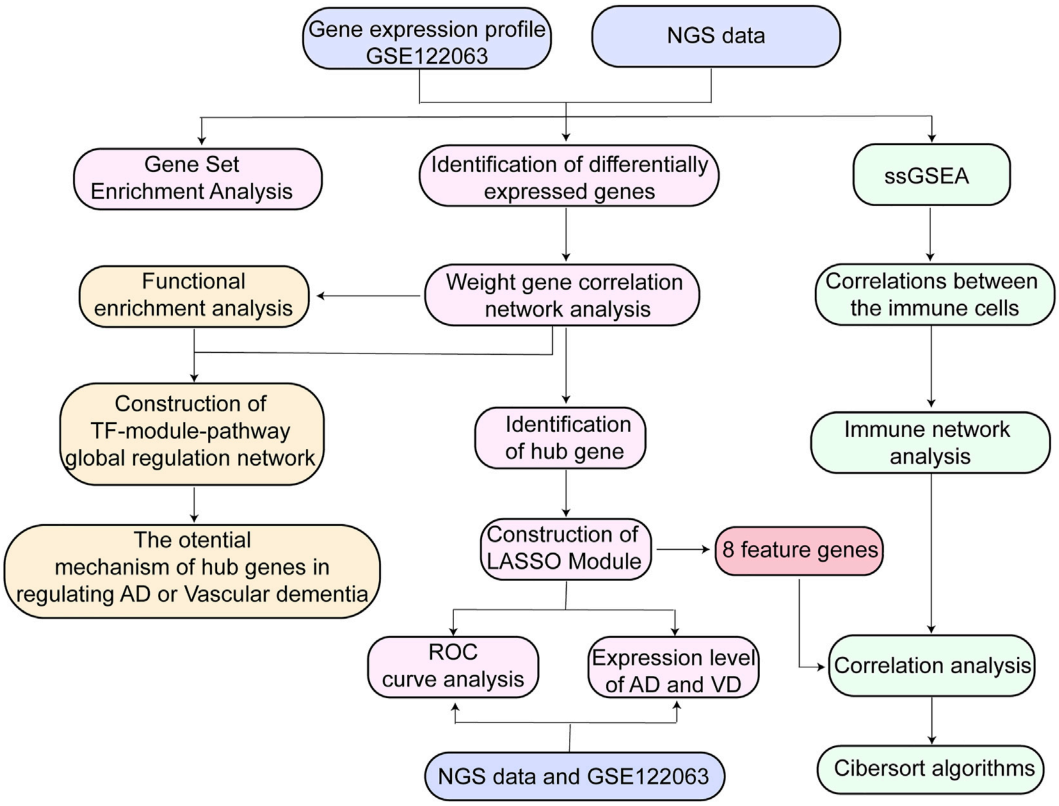


FIGURE 1
Flow chart of study design. AD, Alzheimer's disease; LASSO, least absolute shrinkage and selection operator; NGS data, Next Generation Sequencing data; ROC, receiver operating characteristic curve; ssGSEA, Single Sample Gene Set Enrichment Analysis; TF, transcription factor; VD, vascular dementia.

significant increase in classical natural killer (NK) cells was observed in VD (D'Angelo et al., 2020).

In this study, the expression profile of brain tissue samples from the Gene Expression Omnibus (GEO) data set and the blood expression profile of 3 patients with AD, 6 patients with VD, and 3 healthy donors were analyzed. The Weighted Gene Co-Expression Network Analysis (WGCNA) method and least absolute shrinkage and selection operator (LASSO) model were used to establish diagnostic gene signatures for AD and VD, and to identify potential therapeutic targets. Furthermore, we performed single-sample gene set enrichment analysis (ssGSEA) to quantify immune cell infiltration to provide a theoretical foundation for further research.

Methods and materials

Data collection and processing

High-throughput RNA Sequencing data were used to construct the blood RNA expression profiles of 3 patients with AD, 6 patients with VD, and 3 healthy donors. The AD and VD samples related clinical information were shown in Table 1. Public dataset GSE122063 was obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122063), which includes brain samples from 56 individuals with AD (44 female and 12 male), 36 individuals with VD (16 female and 20 male), and 44 healthy individuals (24 female and 20 male). The range of age was 60–91 years for healthy controls, 62–96 years for VD patients and 63–91 years for AD patients. Gene expression profiling was performed on frontal and temporal cortex tissue from patients with VD and AD, and healthy controls obtained from the University of Michigan Brain Bank. Controls and patients with AD had no infarcts in the autopsied hemisphere. In order to unify the different data, the normalizeBetweenArrays function in the limma package (Ritchie et al., 2015) was used to normalize the gene expression profiles. If a gene corresponds to multiple probes, the average expression value of these probes was chosen as the expression value of the gene. The workflow of the present study was shown in Figure 1.

All analyses in this study were based on the Bioinforcloud platform (http://www.bioinforcloud.org.cn), including expression profiles of GSE122063 and NGS data analysis. Bioinforcloud platform is a self-developed bioinformation analysis platform, which is a collection of data download, analysis and visualization of results, brought together various data processing and bioinformatics analysis methods. Furthermore, the DEGs were screened in two comparion-pairs (AD or VD compared to normal tissues in GSE122063 dataset and NGS data), which adjusted P values >0.05 were significant.

RNA purification and library preparation

Total RNA was extracted and a library was prepared according to the reagent manufacturer's instructions. RNA purity was verified using a kaiaoK5500®Spectrophotometer (Kaiao, Beijing, China), and the integrity of the RNA was evaluated using an RNA Nano 6000 Assay Kit on an Agilent 2,100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Two micrograms of total RNA from each sample was used as the input for library construction using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (#E7530L, NEB, USA).

Library clustering and sequencing

Clustering was performed on the HiSeqPE Cluster reagent kit v4-cBot-HS (Illumina) on the HiSeqPE cluster generation system according to the manufacturer's instructions. After clustering generation, they were sequenced on the Illumina platform of the library, and 150bp paired-end reads were generated.

Differential expression analysis

The expression profiles of 12 cases from the GSE122063 were selected using the Intersect function (Chen and Boutros, 2011) for analysis of co-expressed genes. Differentially expressed genes (DEGs) between AD and VD were screened using the limma package in R (Ritchie et al., 2015). Genes with adjusted $p < 0.05$ were considered to be significantly differentially expressed.

Identification of hub genes using WGCNA

To find co-expressed gene modules, we extracted DEGs from GSE122063 to perform co-expression network using WGCNA package in R (Langfelder and Horvath, 2008). First, the "dist" function was used to calculate the distance between the variables, and a hierarchical clustering analysis was performed using the "hclust" function. We calculated the power parameters using the "pickSoftThreshold" function, which in turn assessed the average connectivity and independence between the modules. The power is deemed proper when the independence exceeds 0.9. Co-expressed gene modules were identified by dynamic tree cutting methods, and hierarchical clustering was established. Subsequently, we calculated module-disease correlation using Pearson correlation analysis to obtain relevant modules with disease status (AD; VD). Furthermore, associations between genes and modules were defined as module memberships (MM), and gene significance (GS) was determined by a combination of phenotypic feature information and the co-expression. A gene was defined as a hub gene in the module if it had GS > 0.2 and MM > 0.9.

Functional enrichment analysis

The module eigengene were analyzed using Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis by the clusterProfiler package (Yu et al., 2012) in R. Results with $p < 0.05$ were considered significant. We screened biological processes (BP) and KEGG pathways related to AD or VD using Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) by the MSigDB c2.cp.kegg.v7.2.symbols.gmt gene set collection (Liberzon et al., 2015), P value <0.05 with the pathways were considered statistically significant.

Gene set variation analysis

We performed gene set variation analysis (GSVA) of the expression profile data sets GSE122063 and NGS data using the GSVA package in R (Hanzelmann et al., 2013). Individual samples were scored with the gene set using GSVA, and GSVA scores were obtained for each sample. The GSVA scores for gene sets were calculated for the GSE122063 and NGS data.

Construction of the protein–protein interaction network

Based on the interactions of human transcription factor (TFs) with their target genes in the TRRUST v2 database (Li et al., 2018), the hypergeometric test was used to predict potential TFs regulating functional modules. In addition, complex cellular functions were performed through the interactions between proteins. The PPI network was constructed using Cytoscape software (http://cytoscape.org/) (Shannon et al., 2003) according to STRING database (Szklarczyk et al., 2017).

Construction of LASSO model and receiver operating characteristic curve analysis

We used LASSO as predictive tool to select the best features of high-dimensional data (Ding et al., 2019). We extracted key genes to construct LASSO models to distinguish between AD and VD. Then, we calculated gene expression values for the regression coefficient, in which the formula weighted the expression of gene expression: signature index = ExpGene1*Coef1 + ExpGene2*Coef2 + ExpGene3*Coef3+. . ."Coef" is the regression coefficient of the gene, "Exp" represents the expression value of the gene. In addition, we randomly assigned samples in the GSE122063 dataset to the training set (75%) and to the test set (25%). To verify that the LASSO model could discriminate between AD and VD, ROC curve analysis was performed on the training and test sets using pROC package (Robin et al., 2011). To further validate the diagnostic efficiency of LASSO model, we validated the results in sequencing data.

Prediction of miRNA-target gene interactions

Interactions between the top 3 miRNAs with the largest |log FC| values and target genes were predicted using the TargetScan (Http://www.targetscan.org/vert_72) database (Lewis et al., 2005). Cytoscape software (Shannon et al., 2003) was used to visualize the network.

Single-sample gene set enrichment analysis

Relative immune cell infiltration levels in single sample were quantified using ssGSEA in R package GSVA (Hanzelmann et al., 2013). The degree of infiltration of the immune cells in the AD and VD samples were determined. Immunity network analysis was used to explore the correlation between immune cells. We also determined the correlation between feature genes and immune infiltration. The CIBERSORT algorithm (https://cibersort.stanford.edu/) was used to infer cell type proportions in the data from AD samples.

Results

Identification of differentially expressed gene in AD and VD

A total of 18,019 overlapping genes were detected between the profiles in the GSE122063 data set and our sequencing data (Figure 2A). These overlapping genes were further used for differentially expressed gene (DEG) analysis (Figure 2B). In the GSE122063 data set, there were 5,340 DEGs observed between the AD and VD groups, including 2,234 up-regulated genes and 3,106 down-regulated genes. In our NGS sequencing data set, there were 587 DEGs between the AD and VD groups, including 291 up-regulated genes and 296 down-regulated genes. These DEGs were able to discriminate between AD and VD (Figures 2C,D).

Gene modules associated with AD or VD

The WGCNA method was used to identify the core gene modules that differentiated between AD and VD (Figure 3). The results showed that when the minimum power was 3, the independence was greater than 0.90 (Figure 3A). As shown in Figure 3B, we identified four key gene modules that discriminated between AD and VD. The turquoise module was negatively correlated with AD and positively correlated with VD (r = -0.51, P = 2e-10 for AD and r = -0.19, p = 0.02 for VD). The blue module positively correlated with AD, but negatively correlated with VD (r = 0.36, P = 2e-05 for AD and r = -0.49, P = 1e-09 for VD). The brown module negatively correlated with VD (coefficient = -0.36, P = 2e-05) (Figure 3C). In the turquoise module, using GS AD > 0.2 and MM > 0.9 as thresholds, 292 genes were identified as up-regulated hub genes in AD and 24 genes were identified as down-regulated hub genes in VD. In the blue module, using GS AD > 0.2 and MM > 0.9 as thresholds, 4 genes were identified as up-regulated hub genes in AD and 5 genes were identified as down-regulated hub genes in VD. In the brown module, using GS > 0.2 and MM > 0.9 as thresholds, 8 hub genes were identified as poorly expressed in VD (Supplementary Table S1). The hub genes in the turquoise and blue modules were associated with both AD and VD, and hub genes in the brown module was associated with VD (Figure 3D).

Module genes functional enrichment analysis showed that turquoise module genes were significantly involved in biological processes related to neurotransmitters and synaptic regulation such as modulation of chemical synaptic transmission, regulation of trans-synaptic signaling, synapse organization, and vesicle-mediated transport in synapse. The blue module genes were significantly involved in biological processes related to glial cells and nerve sheath cells such as myelination, glial cell differentiation, ensheathment of neurons, and glial cell differentiation. The two modules were associated with KEGG pathways related to cAMP signaling pathway, neurotrophin signaling pathway, GnRH signaling pathway, and ECM- receptor interaction. (Figures 3E,F). Above all, the pathways of module genes may be play a vital role and that promote the development and progress in AD or VD.

Validation of critical pathways in AD and VD

The overlapping genes among the hub genes and the genes identified in KEGG pathway analysis were evaluated further

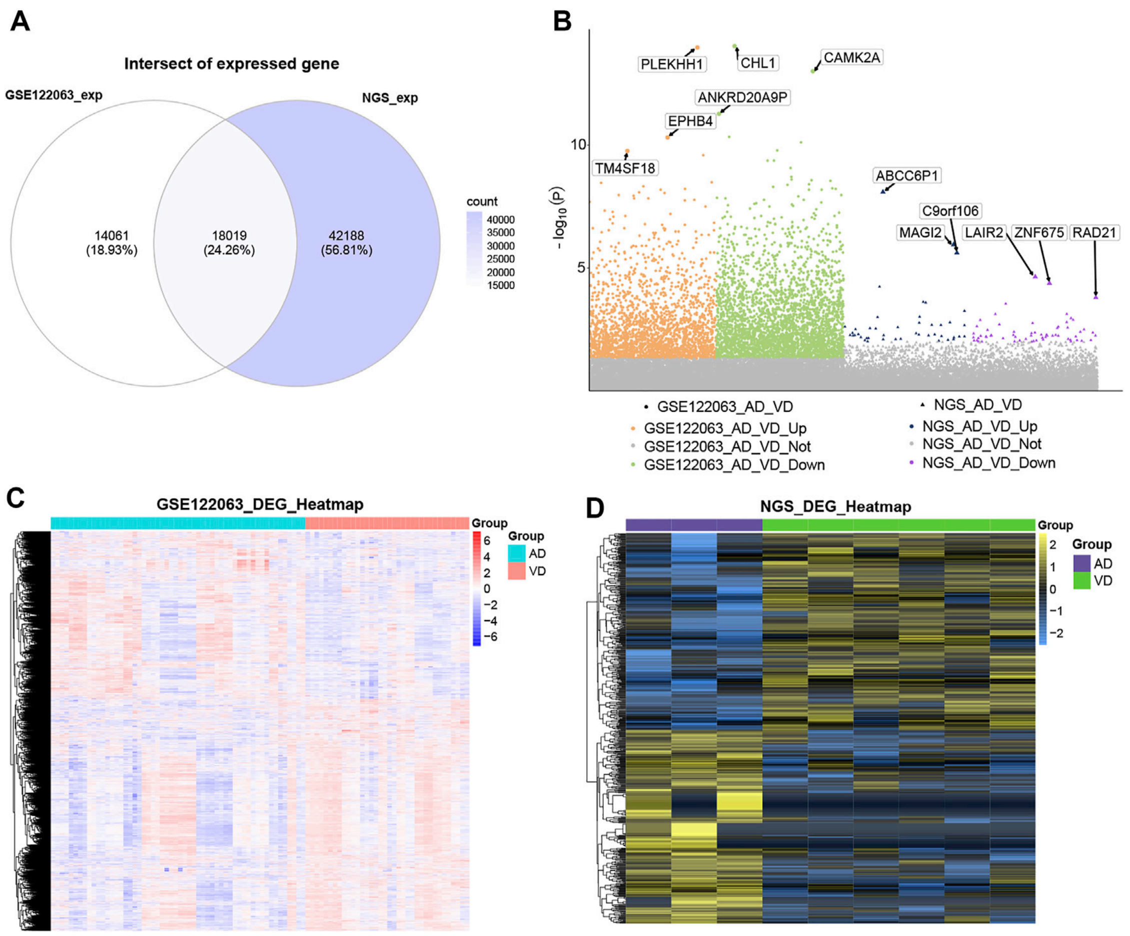


FIGURE 2
Differential expression analysis. **(A)** The venn diagram showed the genes overlaped in the blood expression profile of GSE122063 data set and sequencing data. **(B)** Manhattan map of differential gene expression. **(C)** The heat map of DEGs in GSE122063 data set. **(D)** The heat map of DEGs in Next Generation Sequencing data (NGS). AD, Alzheimer's disease; VD, vascular dementia.

(Supplementary Table S2). A total of 21 hub genes were selected as target genes in the data set (Figure 4A). Then, we constructed a TF-module genes-pathway global regulation network containing 9 TFs and 6 hub genes (Figure 4B). Finally, the mechanisms of different modules in progression of AD or VD were explored (Figures 4C,D).

LASSO model can predict AD and VD

Eight target genes were identified with non-zero regression coefficients as optimal features from 21 target genes in the training set using the LASSO method and 10-fold cross-validation (Figure 5A). Principal component analysis (PCA) showed that the target genes could distinguish AD from VD (Figure 5B). The accuracy of the 8 feature genes based on LASSO model was 0.986 in the training set and 0.960 in the test set. This demonstrated that the model was robust (Figures 5C,D). The results using our sequencing data agreed with the results from the GSEA dataset (Figure 5E). Moreover, the expression of the 8 feature genes was significantly higher in patients with VD than in patients with AD (Figure 5F). The accuracy of the 8 feature genes for discrimination between AD and VD was 0.845 (Figure 5G). In conclusion, we screened 8 feature genes for discrimination of AD and VD by LASSO model, including WNT10B, PPP2CA, NCEH1, MAP2K4, ITPR1, GRIA4, GABBR2 and ATP1A3.

Immune cells infiltration in AD and VD

The GSE122063 data and our NGS data were used to investigate the immune cell types in the AD and VD samples. Dendritic cells (DCs) and plasmacytoid dendritic cells (pDC) were present at significantly greater levels in the AD and VD

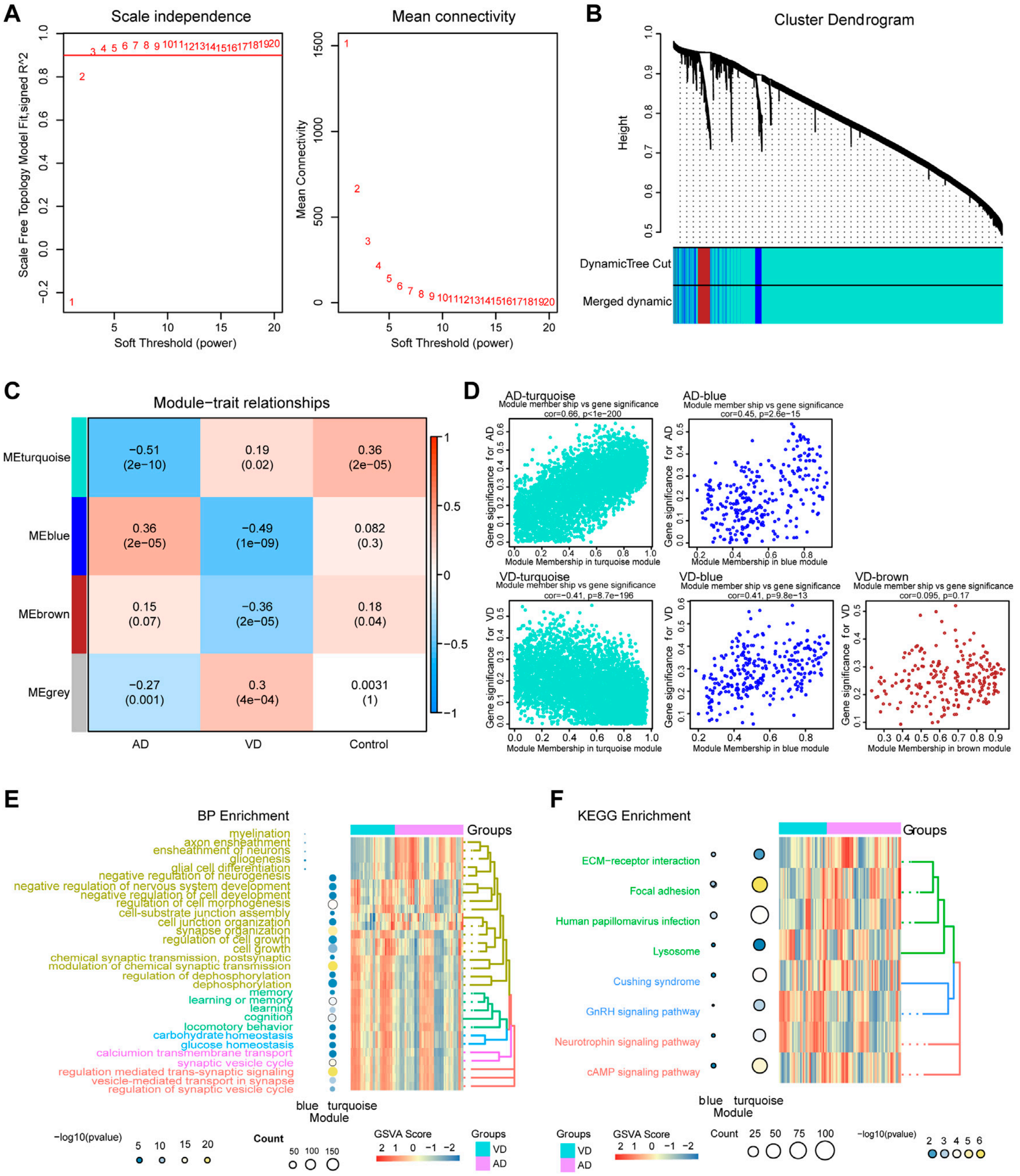


FIGURE 3
Weighted Gene Co-Expression Network Analysis. **(A)** Definition of power related to modules. **(B)** Recognition module. **(C)** The turquoise module negatively correlated with AD and positively correlated with VD, blue module was positively correlated with AD, but the opposite of VD, brown module was negatively correlated with VD. Red, positive correlstion; Blue, negative correlation. **(D)** Module membership and gene significance strongly correlated with each other within each module. **(E)** Biological processes involving genes of the different modules. **(F)** KEGG pathways involving genes of the different modules. AD, Alzheimer's disease; VD, vascular dementia; KEGG, Kyoto Encyclopedia of Genes and Genomes.

samples (Figure 6A). Correlation analysis between the 24 immune cell types showed that increased infiltration of B cells was significantly correlated with AD, and infiltration of aDCs was significantly correlated with VD (Figure 6B). We performed correlation analysis on immune cells using CIBERSORT (proportion). The results showed that pDCs were positively associated with neutrophils (Figure 6C). In addition, we also clustered immune cells based on abundance,

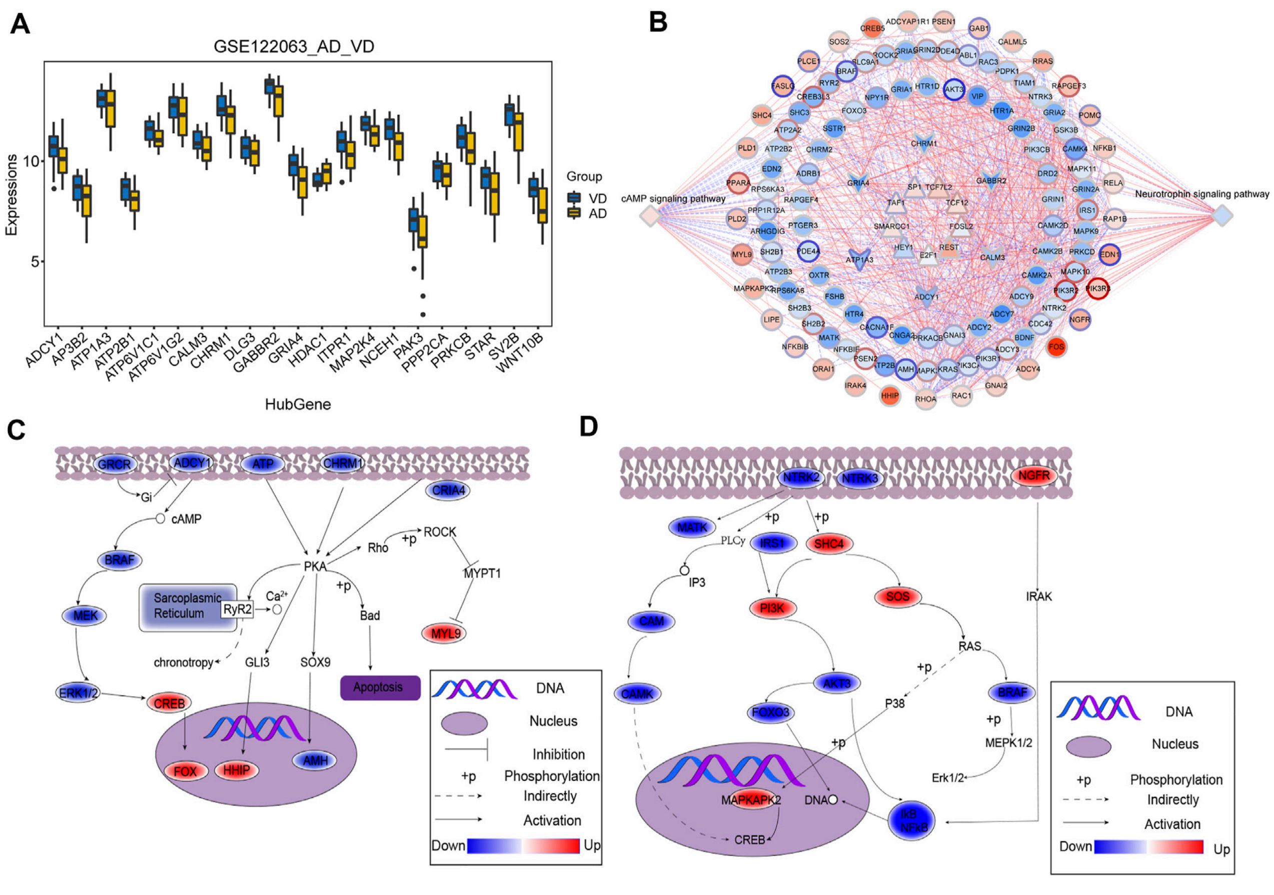


FIGURE 4
Integrated regulation of AD or VD by change in hub gene expression of modules. **(A)** The expression of target genes in the GSE122063 data set. The thick black bar in the middle indicates the interquartile range, and the black line extending from it represents the 95% confidence interval. **(B)** Integrated regulatory network of cAMP signaling pathways and Neurotrophin signaling pathways. **(C,D)** Maps of gene-pathway correlations in both blue and turquoise module. AD, Alzheimer's disease; VD, vascular dementia.

resulting in four clusters (Figure 6D). As shown in Figures 6B,E cells and T follicular helper (TFH) cells were correlated with seven featured genes. We found that plasma cells represented the highest proportion of infiltrated immune cells (Figure 6F).

Discussion

Progress in modern biotechnology and big data analysis has resulted in expansion of biomedical research of diseases beyond clinical symptoms and manifestations. Research has increasingly targeted the regulatory mechanisms of diseases at the molecular level. Previous studies have shown that onset and progression of AD were not caused by a single gene or a few mutations, but by disruption of a comprehensive gene regulation network (Raikwar et al., 2018; Ding et al., 2019).

In this study, using module mining analysis of data sets, we built a module-related biological network. Module mining using WGCNA resulted in identification of three modules associated with AD and VD. Among these, turquoise module genes were up-regulated in AD and down-regulated in VD. Blue module genes were down-regulated in AD and up-regulated in VD. Brown module genes were down-regulated in VD. The results showed that the blue module was enriched in biological processes associated with glial cells and nerve sheath cells, and the turquoise module was associated with cell cycle, synapse, and neurotransmission. Abnormal glial cell function has been shown to play an important role in the pathophysiology of AD (Herculano-Houzel, 2014). Under certain conditions, microglia express proinflammatory factors that may accelerate development of AD (Heppner et al., 2015). In AD, oxidative damage results in changes in cell cycle regulation. Cell cycle dysfunction may play an important role in neuronal dysfunction in AD, and may represent a potential therapeutic target (Bonda et al., 2010). These findings have been shown to be associated with development of AD and VD (Ding et al., 2019). In general,

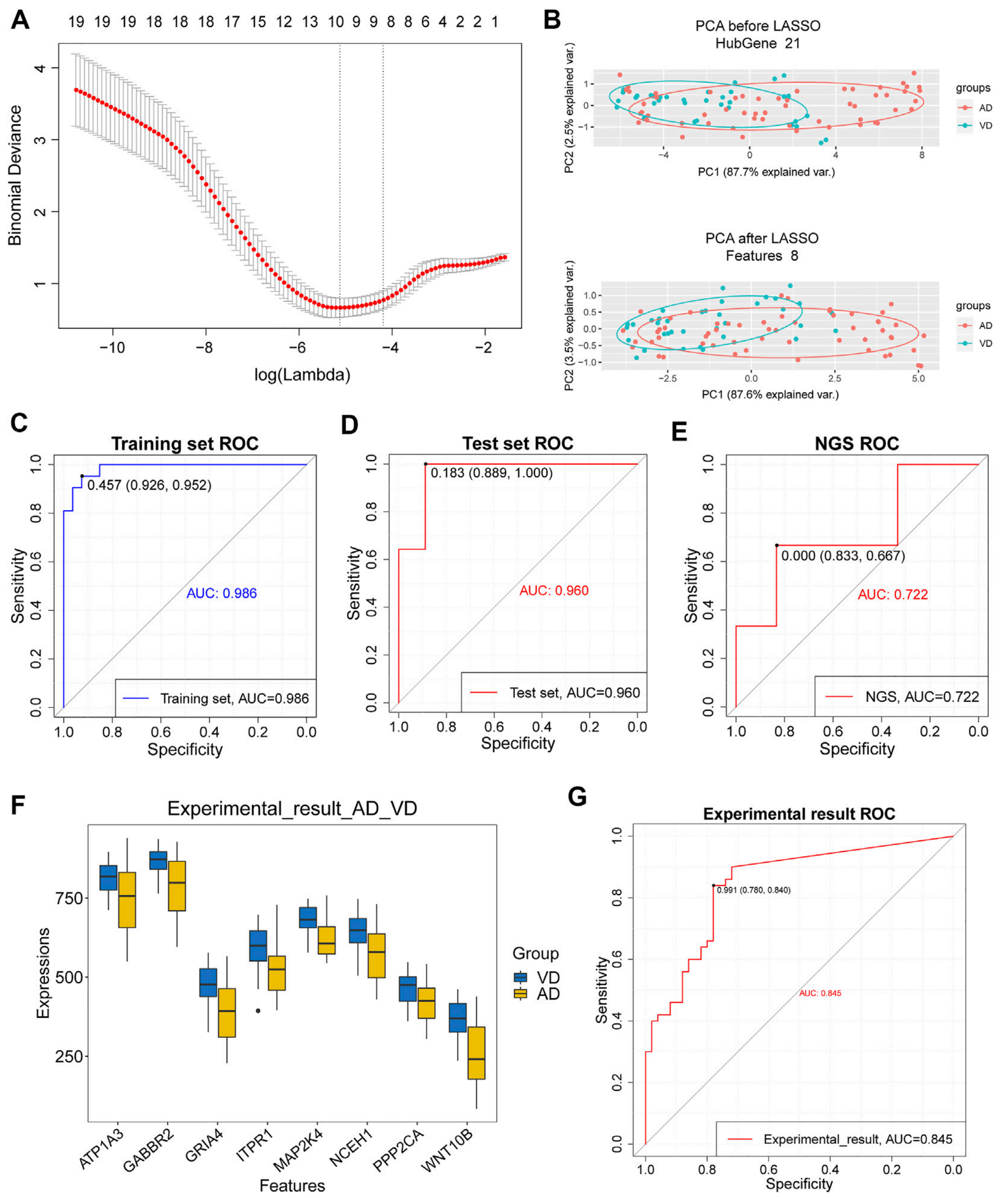


FIGURE 5
Assessment of models for identification of Alzheimer's disease (AD) and vascular dementia (VD). **(A)** 10-fold cross-validation for tuning parameter selection in the LASSO model. **(B)** PCA prior to and after LASSO variable reduction. LASSO, least absolute shrinkage and selection operator; PCA, principal component analysis. **(C,D)** ROC curve for patients with AD and patients with VD in the training and test sets. **(E)** ROC curve for patients with AD and patients with VD in the NGS dataset. **(F)** Gene expression levels in patients with AD and patients with VD in the NGS and GSE122063 data sets. The thick black bar in the middle indicates the interquartile range, and the black line extending from it represents the 95% confidence interval. **(G)** ROC curves for patients with AD and patients with VD in the NGS and GSE122063 data sets.

the genes of modules were involved in pathways play a vital role and which may promote the occurrence of disease course in AD and VD.

We identified six hub genes that regulate key cellular signaling pathways. According to previous studies, ATP1A3 (Shrivastava et al., 2020), PP2A (Wang et al., 2019), NCEH1 (Ding et al., 2019), ITPR1 (Uddin et al., 2018), and CAMKK2 (Sabbir, 2018) have been shown to be associated with onset of AD. The target genes identified in GSE122063 data set were highly expressed in VD, except for HDAC1. In addition, reduced

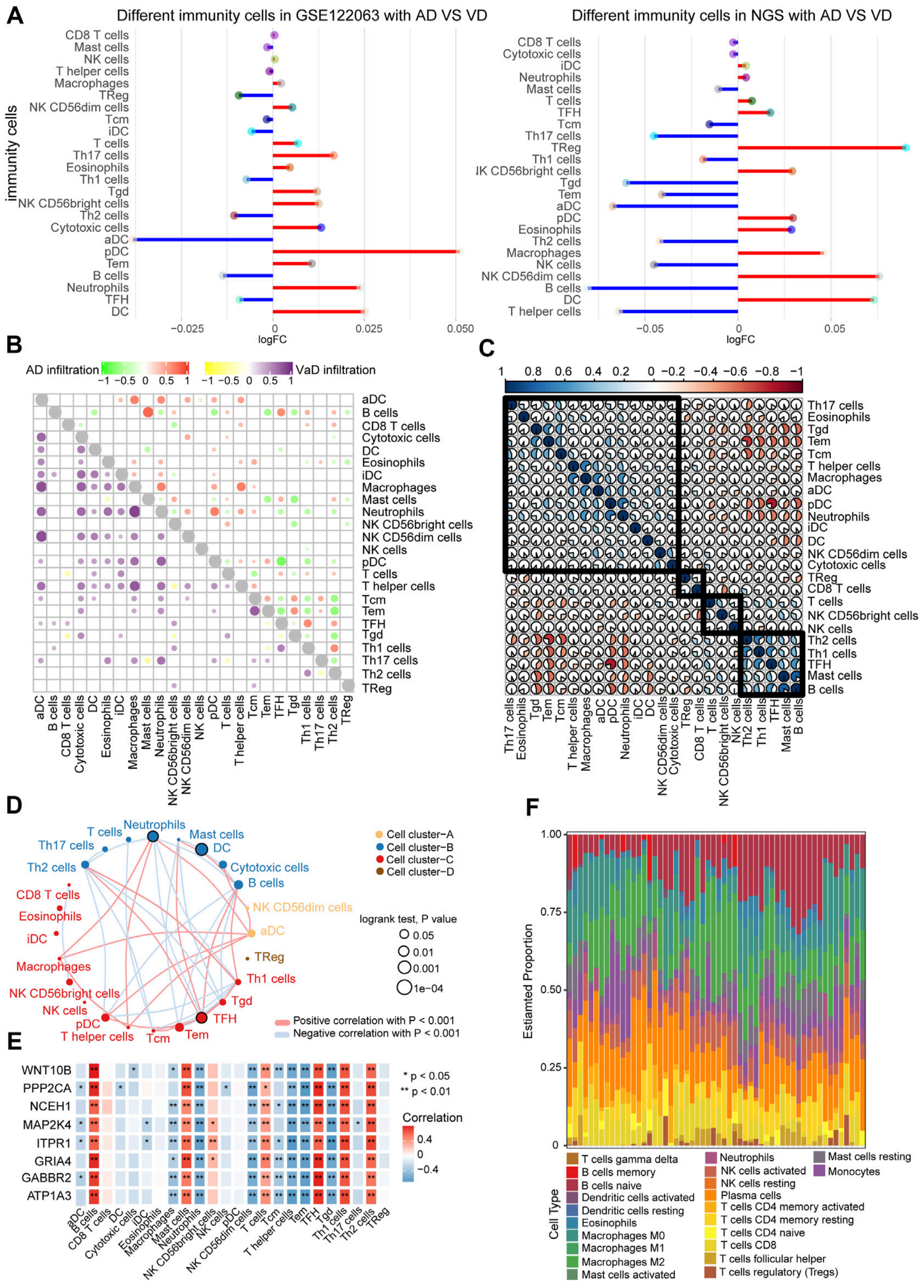


FIGURE 6
Correlation between immune cells in AD and VD. **(A)** Expression of differentially abundant immune cells in the GSE122063 and NGS datasets. **(B)** Correlation between immune cell types in AD and VD. Red and purple represent positive correlations, and green and yellow indicate negative correlations. **(C)** Correlation between immune cells. The blue section indicates activation, and the orange section indicates inhibition. **(D)** Network of immune cell types (abundance). Circles represents the prognostic effect of the cell type, and the thickness of the line indicates the strength of correlations between the cell types. **(E)** Correlation between immune cell types and the eight featured genes. **(F)** Estimated proportions of immune cell types in AD. AD, Alzheimer's disease; VD, vascular dementia.

cAMP signaling through PKA has been shown to be a key feature of AD pathology, and local increases in cAMP signaling may contribute to AD pathology (Kelly, 2018). Neurotrophin plays an important role in central and peripheral neuron survival and differentiation. Inhibition of axonal neurotrophin transport may also contribute to development of AD (Wu et al., 2009). Our results showed that 9 TFs regulated these pathways through six interacting hub genes. A comprehensive regulatory landscape network map was constructed. ATP1A3, PP2A, NCEH1, ITPR1, CAMKK2, and

Eight feature genes were identified using LASSO regression that may be involved in development of AD. Studies have shown that the AMPA receptor (GRIA4) was significantly up-regulated in the hippocampus of patients with AD (Jacob et al., 2007). MAP2K4 was exhibited brain-specific gene and to play essential roles in the regulation of cell proliferation in AD (Wu et al., 2021), while MAP2K4 was related with the condition and prognosis of endometrial carcinoma (Zhang et al., 2022). ITPR1 (Seo et al., 2020) and GABBR2 (Yin et al., 2021) may be associated with AD, and prostate cancer (Choi et al., 2022). Furthermore, PPP2CA as a candidate gene that it may affect the risk of AD (Vazquez-Higuera et al., 2011). NCEH1 and WNT10B, and ATP1A3 have been rarely reported to be associated with AD, but WNT10B has an important role in progression of colorectal cancer (Shi et al., 2019) and hepatocellular carcinoma (Zhou et al., 2020). Therefore, suggesting that WNT10B, ITPR1, GABBR2, ATP1A3, NCEH1, MAP2K4, PPP2CA, and GRIA4 may play a vital role in AD and VD, while also need more studies to further validate the expression of hub genes. Furthermore, the LASSO model based on target genes showed good diagnostic value, which was validated using our sequencing data.

Studies have reported that age-related immunoadaptive recombination causes lymphocyte immunity as a whole to begin having a role in an intermediate metastable state, and the dominant role of immune factors in the pathogenesis of VD and AD (Nuvakhova and Rachin, 2020). In the present, to quantify the extent to which the immune cells infiltrated into brain tissue, we used ssGSEA. The results showed that infiltration of B cells and TFH cells was significantly higher in AD and VD. Nuclear factor of activated B cells has been shown to be involved in physiological inflammatory processes, and was a promising target for treatment of AD (Seo et al., 2018). We also detected decreased levels of B and T lymphocytes in AD and VD, though the decreases were not statistically significant (Busse et al., 2017). Follicular helper CD4 T cells are specialized helpers of B cells (Crotty, 2011). Regulatory T cells were significantly reduced in VD patients, and the T cells were significantly increased in AD patients, possibly due to the inflammation triggered by Aβ (Ziegler-Heitbrock, 2007). Recently, neuroinflammation and tissue-resident immune cells are increasingly recognized as key factors in the pathogenesis of AD (Guzman-Martinez et al., 2019; Lutshumba et al., 2021). Therefore, we speculated that immune cell interactions may promote development of AD and VD.

In conclusion, we used WGCNA analysis to mine modules related to AD and VD, and identified target genes that may regulate AD and VD. Using LASSO modeling, we showed that these target genes could distinguish between AD and VD. Furthermore, modules of WGCNA were significantly involved in cAMP signaling pathway, suggesting genes of pathways may be promote the cell death in AD and VD. However, this study had some limitations. First, the study was based primarily on bioinformatics analysis, while the experiments were not validated, so we only offer theoretical conclusions. Second, our sequencing data can discriminate was validated between AD and VD, while the samples size was relatively small, so studies with large sample sizes are warranted to affirm our findings. Therefore, this study provided a theoretical basis for discrimination between AD and VD, and provided new insight for future studies.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Second Affiliated Hospital of Harbin Medical University (KY2022-094). The patients/participants provided their written informed consent to participate in this study.

Author contributions

YT, XY, and QJ analyzed the data and wrote the manuscript. LW and CY contributed to the study design, and reviewed and edited the manuscript. HY and JZ contributed to data collection, data interpretation, and manuscript writing. All authors read and approved the final manuscript.

Funding

This study was supported by postdoctoral scientific research developmental fund of Heilongjiang Province (LBH-Q15100 and LBH-Q18095).

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.2022.1038585/full#supplementary-material

SUPPLEMENTARY TABLE S1
Hub genes.

SUPPLEMENTARY TABLE S2
21 hub genes regulating KEGG pathway.

References

Baik, S. H., Cha, M. Y., Hyun, Y. M., Cho, H., Hamza, B., Kim, D. K., et al. (2014). Migration of neutrophils targeting amyloid plaques in Alzheimer's disease mouse model. *Neurobiol. Aging* 35, 1286–1292. doi:10.1016/j.neurobiolaging.2014.01.003

Bonda, D. J., Bajic, V. P., Spremo-Potparevic, B., Casadesus, G., Zhu, X., Smith, M. A., et al. (2010). Review: Cell cycle aberrations and neurodegeneration. *Neuropathol. Appl. Neurobiol.* 36, 157–163. doi:10.1111/j.1365-2990.2010.01064.x

Busse, M., Michler, E., von Hoff, F., Dobrowolny, H., Hartig, R., Frodl, T., et al. (2017). Alterations in the peripheral immune system in dementia. *J. Alzheimers Dis.* 58, 1303–1313. doi:10.3233/JAD-161304

Calsolaro, V., and Edison, P. (2016). Neuroinflammation in Alzheimer's disease: Current evidence and future directions. *Alzheimers Dement.* 12, 719–732. doi:10.1016/j.jalz.2016.02.010

Chen, H., and Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable venn and euler diagrams in R. *BMC Bioinforma.* 12, 35. doi:10.1186/1471-2105-12-35

Choi, S., Lee, S., Han, Y. H., Choi, J., Kim, I., Lee, J., et al. (2022). miR-31-3p functions as a tumor suppressor by directly targeting GABBR2 in prostate cancer. *Front. Oncol.* 12, 945057. doi:10.3389/fonc.2022.945057

Crotty, S. (2011). Follicular helper CD4 T cells (TFH). *Annu. Rev. Immunol.* 29, 621–663. doi:10.1146/annurev-immunol-031210-101400

D'Angelo, C., GolDeck, D., Pawelec, G., Gaspari, L., Di Iorio, A., and Paganelli, R. (2020). Exploratory study on immune phenotypes in Alzheimer's disease and vascular dementia. *Eur. J. Neurol.* 27, 1887–1894. doi:10.1111/ene.14360

Ding, J., Kong, W., Mou, X., and Wang, S. (2019). Construction of transcriptional regulatory network of Alzheimer's disease based on PANDA algorithm. *Interdiscip. Sci.* 11, 226–236. doi:10.1007/s12539-018-0297-0

Gate, D., Saligrama, N., Leventhal, O., Yang, A. C., Unger, M. S., Middeldorp, J., et al. (2020). Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* 577, 399–404. doi:10.1038/s41586-019-1895-7

Guzman-Martinez, L., Maccioni, R. B., Andrade, V., Navarrete, L. P., Pastor, M. G., and Ramos-Escobar, N. (2019). Neuroinflammation as a common feature of neurodegenerative disorders. *Front. Pharmacol.* 10, 1008. doi:10.3389/fphar.2019.01008

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

Heneka, M. T., Golenbock, D. T., and Latz, E. (2015). Innate immunity in Alzheimer's disease. *Nat. Immunol.* 16, 229–236. doi:10.1038/ni.3102

Heppner, F. L., Ransohoff, R. M., and Becher, B. (2015). Immune attack: The role of inflammation in alzheimer disease. *Nat. Rev. Neurosci.* 16, 358–372. doi:10.1038/nrn3880

Herculano-Houzel, S. (2014). The glia/neuron ratio: How it varies uniformly across brain structures and species and what that means for brain physiology and evolution. *Glia* 62, 1377–1391. doi:10.1002/glia.22683

Jacob, C. P., KoutsiliEri, E., Bartl, J., NeuEn-Jacob, E., Arzberger, T., ZaNderN., et al. (2007). Alterations in expression of glutamatergic transporters and receptors in sporadic Alzheimer's disease. *J. Alzheimers Dis.* 11, 97–116. doi:10.3233/jad-2007-11113

Kelly, M. P. (2018). Cyclic nucleotide signaling changes associated with normal aging and age-related diseases of the brain. *Cell. Signal.* 42, 281–291. doi:10.1016/j.cellsig.2017.11.004

Lane, C. A., Hardy, J., and Schott, J. M. (2018). Alzheimer's disease. *Eur. J. Neurol.* 25, 59–70. doi:10.1111/ene.13439

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

Lanoiselee, H. M., Nicolas, G., Wallon, D., Rovelet-Lecrux, A., Lacour, M., Rousseau, S., et al. (2017). APP, PSEN1, and PSEN2 mutations in early-onset alzheimer disease: A genetic screening study of familial and sporadic cases. *PLoS Med.* 14, e1002270. doi:10.1371/journal.pmed.1002270

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20. doi:10.1016/j.cell.2004.12.035

Li, L., Wang, L., Li, H., Han, X., Chen, S., Yang, B., et al. (2018). Characterization of LncRNA expression profile and identification of novel LncRNA biomarkers to diagnose coronary artery disease. *Atherosclerosis* 275, 359–367. doi:10.1016/j.atherosclerosis.2018.06.866

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Lueg, G., Gross, C. C., Lohmann, H., Johnen, A., Kemmling, A., Deppe, M., et al. (2015). Clinical relevance of specific T-cell activation in the blood and cerebrospinal fluid of patients with mild Alzheimer's disease. *Neurobiol. Aging* 36, 81–89. doi:10.1016/j.neurobiolaging.2014.08.008

Luo, J., Chen, L., Huang, X., Xie, J., Zou, C., Pan, M., et al. (2022). REPS1 as a potential biomarker in Alzheimer's disease and vascular dementia. *Front. Aging Neurosci.* 14, 894824. doi:10.3389/fnagi.2022.894824

Lutshumba, J., Nikolajczyk, B. S., and Bachstetter, A. D. (2021). Dysregulation of systemic immunity in aging and dementia. *Front. Cell. Neurosci.* 15, 652111. doi:10.3389/fncel.2021.652111

Mavroudis, I., Chowdhury, R., Petridis, F., Karantali, E., Chatzikonstantinou, S., Balmus, I. M., et al. (2021). YKL-40 as a potential biomarker for the differential diagnosis of Alzheimer's disease. *Med. Kaunas.* 58, 60. doi:10.3390/medicina58010060

Nuvakhova, M. B., and Rachin, S. A. (2020). Age-related changes in the immune system and cognitive disorders in vascular dementia and Alzheimer's disease. *Zh. Nevrol. Psikhiatr. Im. S. S. Korsakova* 120, 154–159. doi:10.17116/jnevro2020120081154

Polfliet, M. M., Goede, P. H., Van Kesteren-Hendrikx, E. M., VaN Rooije, N. N., Dijkstra, C. D., and Van den Berg, T. K. (2001). A method for the selective depletion of perivascular and meningeal macrophages in the central nervous system. *J. Neuroimmunol.* 116, 188–195. doi:10.1016/s0165-5728(01)00282-x

Raikwar, S. P., Thangavel, R., Dubova, I., Ahmed, M. E., Selvakumar, P. G., Kempuraj, D., et al. (2018). Neuro-immuno-gene- and genome-editing-therapy for Alzheimer's disease: Are we there yet? *J. Alzheimers Dis.* 65, 321–344. doi:10.3233/JAD-180422

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

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

Sabbir, M. G. (2018). Loss of Ca(2+)/calmodulin dependent protein kinase kinase 2 leads to aberrant transferrin phosphorylation and trafficking: A potential biomarker for Alzheimer's disease. *Front. Mol. Biosci.* 5, 99. doi:10.3389/fmolb.2018.00099

Seo, E. J., Fischer, N., and Efferth, T. (2018). Phytochemicals as inhibitors of NF-κB for treatment of Alzheimer's disease. *Pharmacol. Res.* 129, 262–273. doi:10.1016/j.phrs.2017.11.030

Seo, J., Byun, M. S., Yi, D., Lee, J. H., Jeon, S. Y., Shin, S. A., et al. (2020). Genetic associations of *in vivo* pathology influence Alzheimer's disease susceptibility. *Alzheimers Res. Ther.* 12, 156. doi:10.1186/s13195-020-00722-2

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

Shi, L., Xi, J., Xu, X., Peng, B., and Zhang, B. (2019). MiR-148a suppressed cell invasion and migration via targeting WNT10b and modulating beta-catenin signaling in cisplatin-resistant colorectal cancer cells. *Biomed. Pharmacother.* 109, 902–909. doi:10.1016/j.biopha.2018.10.080

Shrivastava, A. N., Triller, A., and Melki, R. (2020). Cell biology and dynamics of Neuronal Na(+)/K(+)-ATPase in health and diseases. *Neuropharmacology* 169, 107461. doi:10.1016/j.neuropharm.2018.12.008

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., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi:10.1093/nar/gkw937

Uddin, M. S., Stachowiak, A., Mamun, A. A., Tzvetkov, N. T., Takeda, S., Atanasov, A. G., et al. (2018). Autophagy and Alzheimer's disease: From molecular mechanisms to therapeutic implications. *Front. Aging Neurosci.* 10, 04. doi:10.3389/fnagi.2018.00004

Uwagbai, O., and Kalish, V. B. (2022). *Vascular dementia*. StatPearls: Treasure Island FL.

Vazquez-Higuera, J. L., Mateo, I., Sanchez-Juan, P., Rodriguez-Rodriguez, E., Pozueta, A., Calero, M., et al. (2011). Genetic variation in the tau protein phosphatase-2A pathway is not associated with Alzheimer's disease risk. *BMC Res. Notes* 4, 327. doi:10.1186/1756-0500-4-327

Wang, X., Qi, Y., Zhou, X., Zhang, G., and Fu, C. (2019). Corrigendum to 'Alteration of scaffold: Possible role of MACF1 in Alzheimer's disease pathogenesis'. *Med. Hypotheses* 136, 109509. doi:10.1016/j.mehy.2019.109509

Wang, Y., Lv, S., Zhou, X., Niu, X., Chen, L., Yang, Z., et al. (2022). Identification of TLR2 as a key target in neuroinflammation in vascular dementia. *Front. Genet.* 13, 860122. doi:10.3389/fgene.2022.860122

Wu, A. T. H., Lawal, B., Wei, L., Wen, Y. T., Tzeng, D. T. W., and Lo, W. C. (2021). Multiomics identification of potential targets for alzheimer disease and antrocin as a therapeutic candidate. *Pharmaceutics* 13, 1555. doi:10.3390/pharmaceutics13101555

Wu, C., Cui, B., He, L., Chen, L., and Mobley, W. C. (2009). The coming of age of axonal neurotrophin signaling endosomes. *J. Proteomics* 72, 46–55. doi:10.1016/j.jprot.2008.10.007

Yin, P., Xue, Y., Wang, T., Zhong, D., and Li, G. (2021). The therapeutic targets of fingolimod (FTY720) are involved in pathological processes in the frontal cortex of Alzheimer's disease patients: A network pharmacology study. *Front. Aging Neurosci.* 13, 609679. doi:10.3389/fnagi.2021.609679

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, X., Huang, Y., and Pang, X. (2022). Value of serum p53, PKD1, and MAP2K4 in evaluating the condition and prognosis of endometrial carcinoma. *Am. J. Transl. Res.* 14, 5059–5067.

Zhou, F., Lei, Y., Xu, X., Zhou, H., Liu, H., Jiang, J., et al. (2020). LINC00355: 8 promotes cell proliferation and migration with invasion via the MiR-6777-3p/Wnt10b axis in Hepatocellular Carcinoma. *J. Cancer* 11, 5641–5655. doi:10.7150/jca.43831

Ziegler-Heitbrock, L. (2007). The CD14+ CD16+ blood monocytes: Their role in infection and inflammation. *J. Leukoc. Biol.* 81, 584–592. doi:10.1189/jlb.0806510

Zou, D., Li, R., Huang, X., Chen, G., Liu, Y., Meng, Y., et al. (2019). Identification of molecular correlations of RBM8A with autophagy in Alzheimer's disease. *Aging (Albany NY)* 11, 11673–11685. doi:10.18632/aging.102571

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 November 2022
DOI 10.3389/fgene.2022.1017520

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Xichun Liu,
The Affiliated Hospital of Qingdao University, China
Zeyan Li,
Shandong University, China

*CORRESPONDENCE
Haixia Wang,
deanyang928@163.com

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

RECEIVED 12 August 2022
ACCEPTED 22 September 2022
PUBLISHED 25 November 2022

CITATION
Wang H (2022), Network pharmacology- and molecular docking-based approaches to unveil the pharmacological mechanisms of dihydroartemisinin against esophageal carcinoma.
Front. Genet. 13:1017520.
doi: 10.3389/fgene.2022.1017520

Network pharmacology- and molecular docking-based approaches to unveil the pharmacological mechanisms of dihydroartemisinin against esophageal carcinoma

Haixia Wang*

Pharmacy Department, West Hospital of the Second Hospital of Shanxi Medical University, Taiyuan, China

Objective: Dihydroartemisinin (DHA) is an active metabolite of artemisinin and its derivatives, which is a potent drug extensively applied in clinical treatment of malaria. The antitumor properties of DHA have received increasing attention. However, there is no systematic summary on the pharmacological mechanisms of DHA against esophageal carcinoma (ESCA). The present study implemented network pharmacology- and molecular docking-based approaches to unveil the pharmacological mechanisms of DHA against ESCA.

Methods: DHA targets were accessed through integrating the SwissTargetPrediction, HERB, as well as BATMAN-TCM platforms. In TCGA-ESCA dataset, genes with differential expression were screened between 161 ESCA and 11 normal tissue specimens. DHA targets against ESCA were obtained through intersection. Their biological significance was evaluated with functional enrichment analysis. A prognostic signature was established *via* uni- and multivariate cox regression analyses. DHA-target interactions were predicted *via* molecular docking. Molecular dynamics simulation was implemented to examine the stability of DHA binding to potential targets. **Results:** The study predicted 160 DHA targets as well as 821 genes with differential expression in ESCA. Afterwards, 16 DHA targets against ESCA were obtained, which remarkably correlated to cell cycle progression. The ADORA2B- and AURKA-based prognostic signature exhibited the reliability and independency in survival prediction. The stable docking of DHA-ADORA2B and DHA-AURKA was confirmed.

Abbreviations: ESCA, esophageal carcinoma; ESCC, squamous cell carcinoma; EAC, adenocarcinoma; DHA, dihydroartemisinin; TCGA, The Cancer Genome Atlas; PPI, Protein-protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes; OS, overall survival; ROC, receiver operating characteristic; AUC, area under the curve; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene set enrichment analysis; RMSD, Root Mean Square Deviation; RMSF, Root Mean Square Float.

Conclusion: Collectively, this study systematically revealed the basis and mechanism of DHA against ESCA through targeting multi-target and multi-pathway mechanisms, and thus offered theoretical and scientific basis for the clinical application of DHA.

KEYWORDS

dihydroartemisinin, esophageal carcinoma, network pharmacology, molecular docking, ADORA2B, AURKA

Introduction

Esophageal carcinoma (ESCA) remains one of the most lethal cancers across the globe (Sung et al., 2021), with two primary histological subtypes: squamous cell carcinoma (ESCC) as well as adenocarcinoma (EAC) (Lu et al., 2022). The overall 5-year survival rate is merely 15–25% (Hulshof et al., 2021). The features and etiology of ESCA may vary on the basis of region or ethnicity (Liu et al., 2017). ESCC occupies the predominant subtype of ESCA, notably in the Asian and African regions, which correlates to dietary habits as well as exposure to carcinogens (Zhang et al., 2022). Oppositely, EAC mainly develops from esophageal epithelial intestinal metaplasia caused by chronic gastroesophageal reflux diseases, which is the majority of ESCA in Western countries (Zhang et al., 2022). Currently, endoscopic or surgical resection is appropriate for early-stage patients, while radio- or/and chemotherapies are mainly reserved for those with advanced stage or metastases (Mukherjee et al., 2021; Yang et al., 2022). Chemotherapeutic agents (cisplatin, 5-fluorouracil, doxorubicin, etc.) may cause a range of dose-limiting toxicities (Zhang et al., 2021). Molecular targeted therapy has achieved unprecedented progress in cancer therapy (Lu et al., 2022). Nevertheless, only trastuzumab against HER2, as well as ramucirumab targeting VEGFR-2 have been recommended for the treatment of ESCA (Xu et al., 2021; Safran et al., 2022). Immunotherapy comprising immune checkpoint inhibitors, vaccines, monoclonal antibodies as well as adoptive cellular immunotherapy represents a novel therapeutic option of ESCA (Janjigian et al., 2021; Shitara et al., 2021). Despite the remarkable potential of immunotherapy in cancer therapy, individual patients' response greatly varies, and only a small percentage of cases respond to immunotherapy (Luo et al., 2021; Sun et al., 2021). Hence, achievement of the goal of effective therapeutic options remains challenging. Rather than exploring a single ESCA-causing gene and therapeutic agents that act only on a single target, the entire drug-disease network should be considered, aiming to determine multi-target agents to lower side effects.

Traditional Chinese medicine artemisinin is extracted from Artemisia annua L, and dihydroartemisinin (DHA) with high water solubility and antimalarial activity is the first-generation derivative of this compound that is an effective and fast-acting antimalarial agent with low toxicity (Zhu et al., 2020). DHA displays the considerable potential for preventing or treating ESCA because of its low toxicity and known safety, and preclinical research has proposed the molecular mechanisms and pharmacological effect underlying DHA against ESCA (Zhu et al., 2020). Zhao et al. reported that DHA exhibits an anti-proliferative effect against ESCC cells, which is capable of down-regulating mTOR cascade pathway partly *via* binding to AKT1 and p70S6K (Zhu et al., 2020). Pyroptosis is a novel form of pro-inflammatory programmed cell death, and its inducers enable to strengthen antitumor effects (Niu et al., 2022). Song et al. proposed that DHA may trigger pyroptosis of ESCC cells *via* impacting the activity of PKM2-caspase-8/3-GSDME signaling (Jiang et al., 2021). Administration of chemotherapeutic agents is usually accompanied by resistance. DHA may sensitize ESCC to cisplatin through attenuating the activity of Sonic Hedgehog pathway (Cui et al., 2020). There is an interplay between autophagy and epithelial-mesenchymal transition during tumor progression (Chen et al., 2021; Niu et al., 2021). He at al. found that the migratory capacity and the epithelial-mesenchymal transition process of ESCA cells is mitigated by DHA *via* triggering autophagy (Chen et al., 2020). PKM2 is a key regulator of glycolysis, and DHA lowers glycolysis of ESCA through down-regulating PKM2 (Li et al., 2019). Additionally, hTERT has been confirmed as a therapeutic target of DHA against ESCC (Li et al., 2021). Altogether, DHA exerts an anti-ESCA effect through targeting multi-target and multi-pathway mechanisms.

In the current study, we developed a new integrated strategy to probe out the key targets and mechanisms of DHA in the treatment of ESCA on the basis of network pharmacology and molecular docking, which offered the theoretical and scientific basis for the clinical application of DHA against ESCA.

Materials and methods

Dihydroartemisinin target prediction

The chemical structure of DHA was acquired from the PubChem chemical information resource (https://pubchem.ncbi.nlm.nih.gov) (Kim et al., 2021). Potential targets of DHA were searched from the SwissTargetPrediction (http://www.swisstargetprediction.ch) (Daina et al., 2019), as well as HERB (http://herb.ac.cn/) (Fang et al., 2021) web servers. Additionally, DHA targets were screened through the BATMAN-TCM (http://

bionet.ncpsb.org/batman-tcm) platform in accordance with the screening conditions of score cutoff = 5 and adjusted p-value < 0. 05 (Liu et al., 2016). DHA targets were matched and corrected on the basis of the universal Protein Resource (UniProt, (2021); http://sparql.uniprot.org/). Afterwards, DHA targets obtained from three platforms were merged, and deduplicated.

The cancer genome atlas (TCGA)-esophageal carcinoma data acquisition

RNA transcriptome profiles as well as clinicopathological and survival information of ESCA cases were downloaded from TCGA-ESCA dataset (https://cancergenome.nih.gov/). In total, 161 ESCA and 11 normal tissue specimens were enrolled for subsequent analysis.

Differential expression analysis

In TCGA-ESCA dataset, limma package was adopted for screening differentially expressed genes (DEGs) between ESCA and normal tissue specimens. Adjusted p-value < 0.01 together with |log2 fold-change|>1 were set as the screening criteria. Shared genes between DHA targets and DEGs were intersected through the Venny 2.1 online tool, named DHA targets against ESCA.

Protein-protein interaction (PPI) analysis

DHA targets against ESCA were uploaded to the Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/) (Szklarczyk et al., 2015). In accordance with hide unconnected targets along with highest confidence of 0. 900, PPI network of DHA targets against ESCA was obtained. Hub genes among DHA targets against ESCA were further evaluated with cytoHubba plugin of Cytoscape tool 3.7.2 (Chin et al., 2014). Degree of each hub gene was also computed.

Establishment and verification of a prognostic signature

In TCGA-ESCA dataset, univariate cox regression analysis was implemented to screen which DHA targets against ESCA significantly correlated to ESCA cases' overall survival (OS). Prognostic DHA targets against ESCA with p-value <0.05 were enrolled in a multivariate cox regression model. The prognostic signature was conducted on the basis of the linear combination of regression coefficients derived from the multivariate Cox regression model multiplied with expression values. The risk score of each case was computed. TCGA-ESCA cases were equally separated into discovery set as well as validation set. In each set, ESCA cases were further assigned to high- and low-risk subpopulations. Kaplan-Meier curves of OS between two subpopulations were plotted, and OS difference was estimated with log-rank test. Alive and dead status of cases in each set was displayed. Receiver operating characteristic (ROC) curve was drawn for evaluation of the prediction accuracy of the prognostic signature. Area under the curve (AUC) was also computed. TNMplot.com web tool was employed for comparison of survival difference between high and low expression subpopulations of DHA targets against ESCA in TCGA-ESCA dataset (Bartha and Győrffy, 2021). Expression values of DHA targets against ESCA across distinct pathological stages were displayed with ggpubr package, and the difference was evaluated using one-way analysis of variance.

Uni- and multivariate cox regression analysis

Clinicopathological factors comprising gender, pathological stage, TNM staging as well as prognostic signature were enrolled for estimating their associations with TCGA-ESCA cases' OS by use of uni- and multivariate cox regression analysis.

Functional enrichment analysis

Biological features of DHA targets against ESCA were probed out utilizing clusterProfiler package (Yu et al., 2012). Gene Ontology (GO) covering biological process, cellular component and molecular function, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were separately implemented.

Gene set enrichment analysis (GSEA)

The potentially altered biological processes between high and low expression subpopulations of DHA targets against ESCA were determined with GSEA tool (Subramanian et al., 2005). The "c5. go.v7.5.1. symbols" and "c2. cp.kegg.v7.5.1. symbols" gene sets acquired from the Molecular Signatures Database Molecular Signatures Database (Liberzon et al., 2015) were utilized as reference sets.

Molecular docking

The 3D structures of DHA targets against ESCA were accessed from the PDB database and saved in pdb format. PyMOL software was used for protein pretreatment. In addition, the 3D structure of DHA was created using

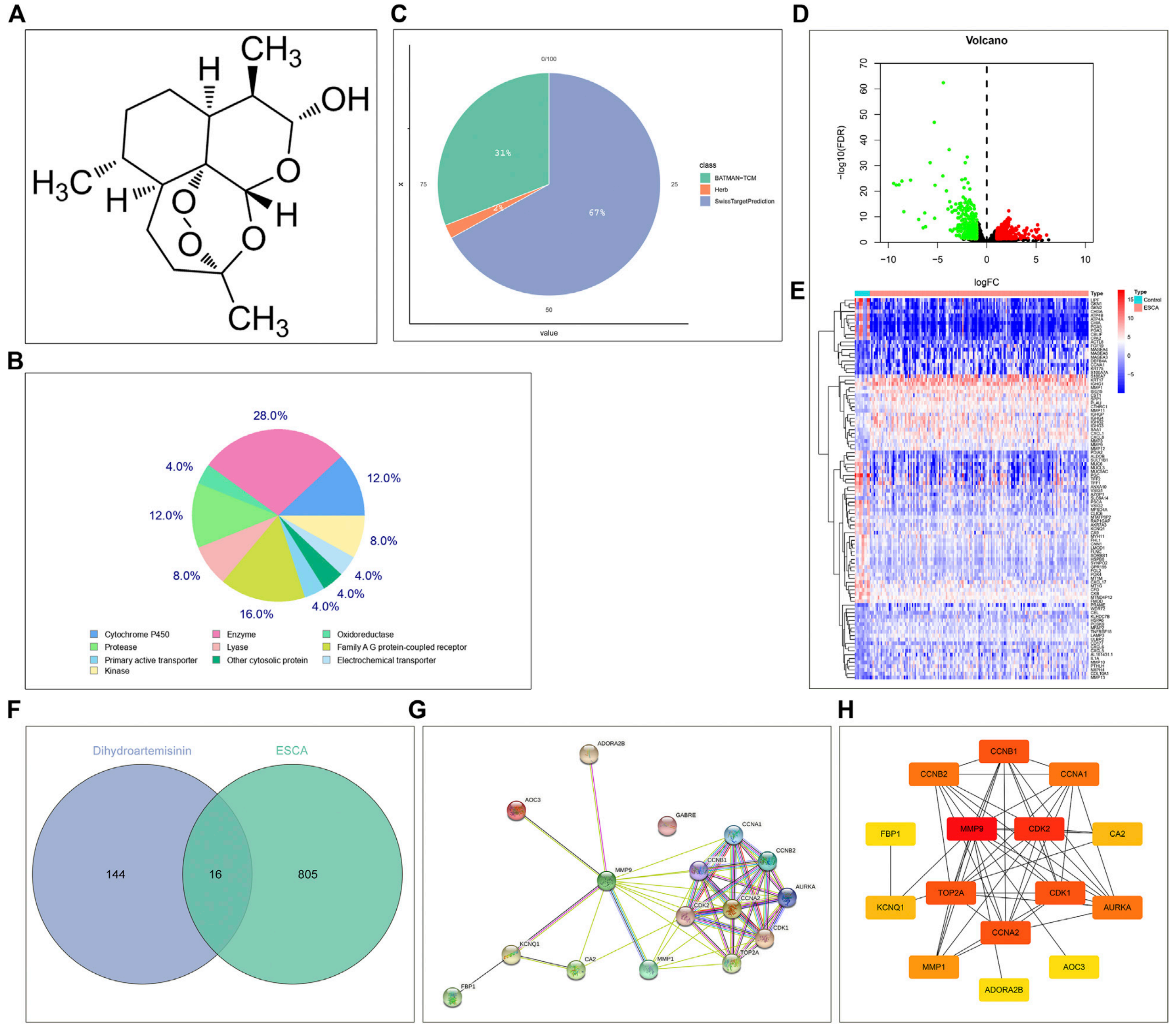


FIGURE 1
Network pharmacology determines DHA targets against ESCA. **(A)** The 2D chemical structure of DHA from the PubChem chemical information resource. **(B)** The target class distribution for the first 25 DHA targets predicted by the SwissTargetPrediction platform. **(C)** The distribution of DHA targets predicted by the SwissTargetPrediction, HERB, as well as BATMAN-TCM platforms. **(D)** Volcano diagram of the DEGs between 161 ESCA and 11 normal tissue specimens in TCGA-ESCA dataset. Red, up-regulation; green, down-regulation. **(E)** Heatmap of the expression values of the first 100 most significantly DEGs between 161 ESCA and 11 normal tissue specimens in TCGA-ESCA dataset. **(F)** Venn diagram of the shared genes between DHA targets and DEGs. **(G,H)** The PPI network of DHA targets against ESCA through the **(G)** STRING online tool or **(H)** cytoHubba plugin of Cytoscape tool.

Chem3D software and saved in mol*2 format. The pre-processed targets and DHA were imported into AutoDockTools 1.5.6 software for molecular docking, and the results were saved in pdbqt format. Vina scripts were run for computing molecular binding energies as well as visualizing molecular docking results. Additionally, Discovery Studio 2019 software was implemented for searching for docking sites as well as calculating LibDockScore for flexible binding. The molecular docking results were imported into PyMOL software for visualizing the molecular docking conformation.

Molecular dynamics simulation

Molecular dynamics simulation was implemented by use of Discovery Studio 2019 software. The molecular structure of

TABLE 1 DHA targets against ESC ranked by degree.

Gene symbol	Log2 fold-change	Adjusted *p*-value	Degree
MMP9	3.239293	2.35E-05	13
CA2	−2.37148	8.01E-08	3
FBP1	−1.21561	0.00363	1
CCNA1	4.166111	0.004252	8
CCNA2	1.499277	9.59E-08	9
CDK2	1.037182	6.92E-05	10
AOC3	−2.21232	7.01E-09	1
MMP1	2.947104	7.96E-05	6
TOP2A	2.52532	3.19E-08	9
AURKA	1.827825	2.06E-09	8
CDK1	1.828233	1.24E-10	9
CCNB1	1.837112	7.43E-10	9
CCNB2	1.771093	4.36E-10	8
KCNQ1	−3.12088	1.83E-15	3
ADORA2B	1.488384	0.001348	1
GABRE	2.197392	0.006224	1

DHA was added to the CHARMM force field by “simulation” module, and DHA targets against ESCA were solvated by “solvation” module. The molecular dynamics simulation parameters were set by “standard dynamics cascade module for the targets added with the solvent system. The system temperature was elevated from 50 K to 300 K under 100 ns of analog sampling; the time step was set to 1 ns; and other parameters were set to default values. The Root Mean Square Deviation (RMSD) as well as Root Mean Square Float (RMSF) values were analyzed.

Results

Network pharmacology determines dihydroartemisinin targets against esophageal carcinoma

From the PubChem chemical information resource, we accessed the 2D chemical structure of DHA, as illustrated in Figure 1A. Through integrating the SwissTargetPrediction, HERB, as well as BATMAN-TCM platforms, potential therapeutic targets of DHA were predicted. In accordance with the SwissTargetPrediction platform, the target class distribution for the first 25 DHA targets was visualized (Figure 1B). Enzyme was the predominant target class, occupying 28.0%. In total, 111, 3, and 51 DHA targets were separately predicted by the SwissTargetPrediction, HERB, as well as BATMAN-TCM platforms (Figure 1C). Under merging and deduplication, 160 DHA targets were finally acquired (Supplementary Table S1).

In TCGA-ESCA dataset, 821 DEGs (comprising 429 genes with up-regulation as well as 392 genes with down-regulation) between 161 ESCA and 11 normal tissue specimens were screened in line with the screening condition of adjusted p-value < 0.01 together with |log2 fold-change|>1 (Figure 1D). The specific information of DEGs was listed in Supplementary Table S2. The first 100 most significantly DEGs were depicted in Figure 1E. Venn diagram illustrated 16 DHA targets against ESCA shared by DHA targets and DEGs (Figure 1F), comprising MMP9, CA2, FBP1, CCNA1, CCNA2, CDK2, AOC3, MMP1, TOP2A, AURKA, CDK1, CCNB1, CCNB2, KCNQ1, ADORA2B, and GABRE. Their closely interactions were investigated in the PPI network (Figures 1G,H and Table 1).

Establishment and verification of a reliable prognostic signature for esophageal carcinoma

The prognostic implication of DHA targets against ESCA was investigated in TCGA-ESCA dataset. Univariate cox regression analysis demonstrated that AURKA, and ADORA2B significantly correlated to ESCA cases’ OS. AURKA, and ADORA2B were utilized for establishing a multivariate cox regression model. The risk score was computed on the basis of the coefficients and expression values of AURKA, and ADORA2B. TCGA-ESCA cases were equally divided into discovery set and validation set. In discovery set, we assigned cases into high- and low-risk subpopulations. Survival difference between subpopulations was assessed. High-risk subpopulation exhibited worse OS outcome in comparison to low-risk subpopulation (Figure 2A). Survival status

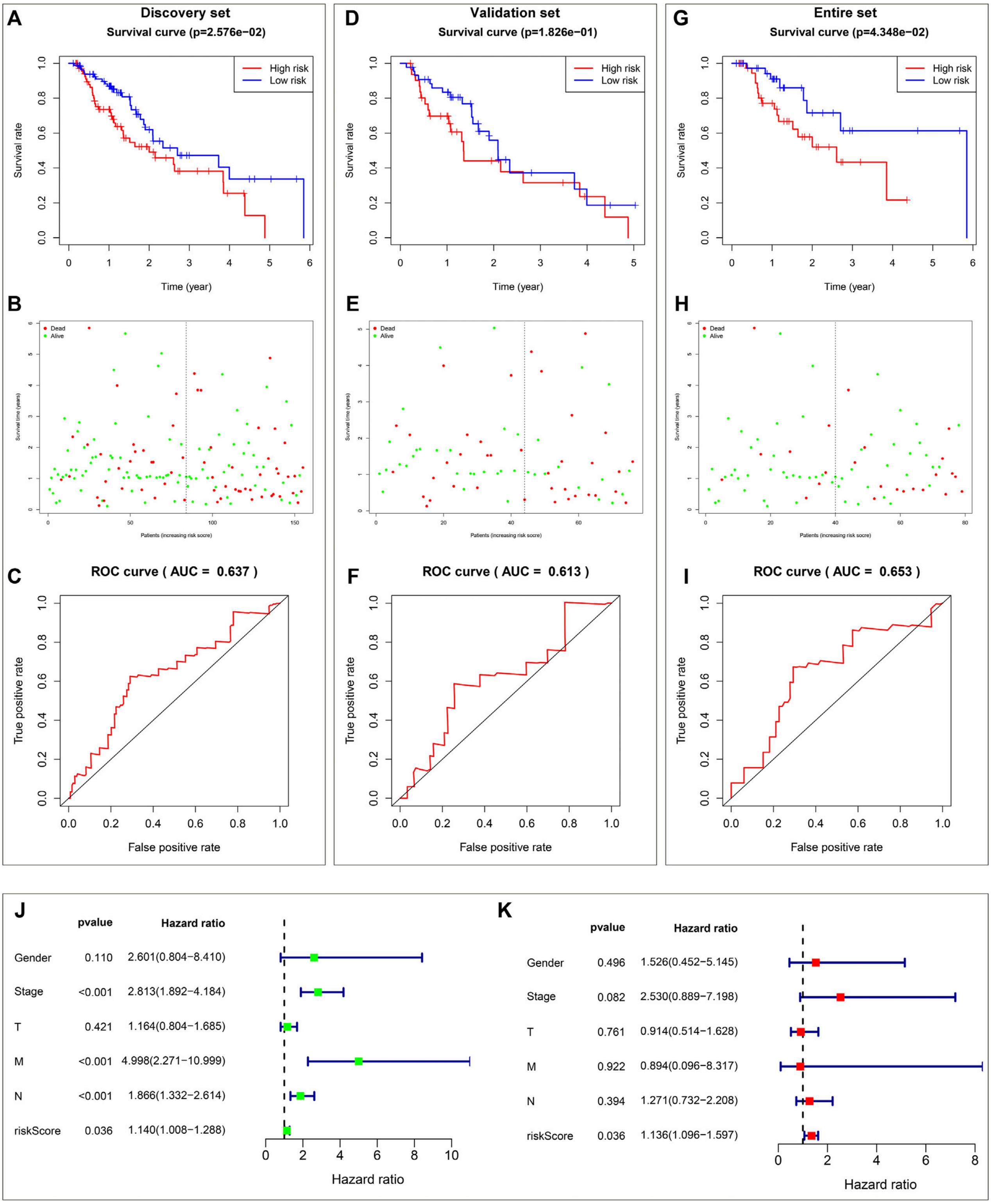


FIGURE 2
Establishment and verification of a reliable prognostic signature for ESCA cases in TCGA-ESCA dataset. **(A)** Kaplan-Meier OS curves and log-rank test between high- and low-risk subpopulations in discovery set. **(B)** Distribution of alive and dead status in high- and low-risk subpopulations in discovery set. **(C)** ROC curve on the basis of the prognostic signature in discovery set. **(D)** Kaplan-Meier OS curves and log-rank test between high- and low-risk subpopulations in validation set. **(E)** Distribution of alive and dead status in high- and low-risk subpopulations in validation set. **(F)**

(Continued)

FIGURE 2 (Continued)
ROC curve of the prognostic signature in validation set. **(G)** Kaplan-Meier OS curves and log-rank test between high- and low-risk subpopulations in entire set. **(H)** Distribution of alive and dead status in high- and low-risk subpopulations in entire set. **(I)** ROC curve based on the prognostic signature in entire set. **(J)** Forest diagram of univariate cox regression analysis for the associations of clinicopathological factors and the prognostic signature with ESCA cases' OS. **(K)** Forest diagram of multivariate cox regression analysis for the associations of clinicopathological factors and the prognostic signature with ESCA cases' OS.

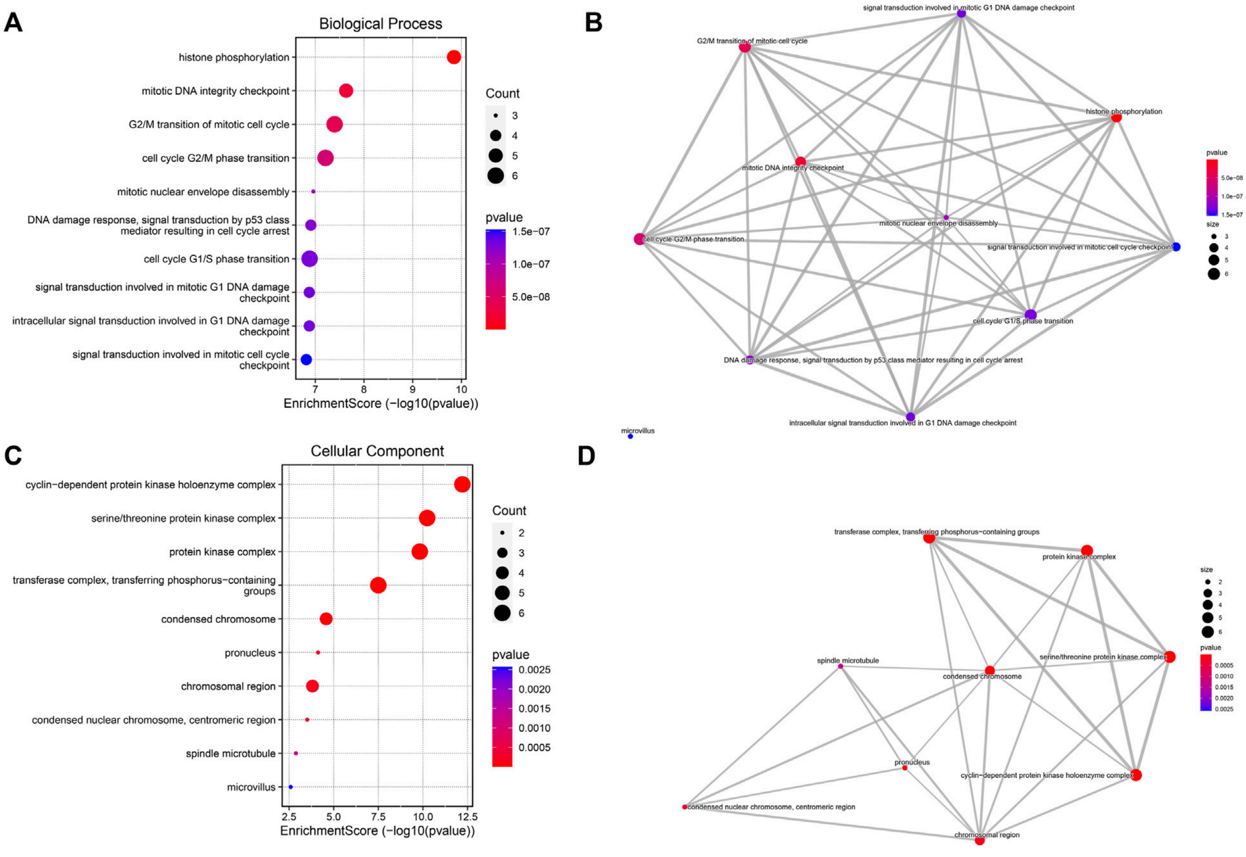


FIGURE 3
Biological features of DHA targets against ESCA. **(A)** The first ten biological process terms enriched by DHA targets against ESCA. **(B)** Interactions between the first ten biological process terms. **(C)** The first ten cellular component terms enriched by DHA targets against ESCA. **(D)** Interactions between the first ten cellular component terms. **(E)** The first ten molecular function terms enriched by DHA targets against ESCA. **(F)** Interactions between the first ten molecular function terms. **(G)** The first ten KEGG pathways enriched by DHA targets against ESCA. **(H)** Interactions between the first ten KEGG pathways.

was also compared between subpopulations. More dead cases were investigated in high-risk subpopulation (Figure 2B). For evaluating whether the prognostic signature enabled to estimate ESCA cases' OS, ROC curve was plotted. In Figure 2C, the AUC value was 0.637, indicative of the excellent performance of the prognostic signature in estimating ESCA cases' OS. The prognostic signature was further verified in validation set and entire set. As expected, it exhibited the remarkable advantage in ESCA cases' prognosis prediction (Figures 2D–I), indicating the clinical generalizability of this signature.

The prognostic signature independently predicts esophageal carcinoma patients' prognosis

Utilizing univariate cox regression analysis, we investigated the associations of clinicopathological factors and the prognostic signature with ESCA cases' OS. As illustrated in Figure 2J, the prognostic signature along with pathological stage, N stage and M stage were linked to worse ESCA cases' OS. Multivariate cox regression analysis was conducted for estimating which factors

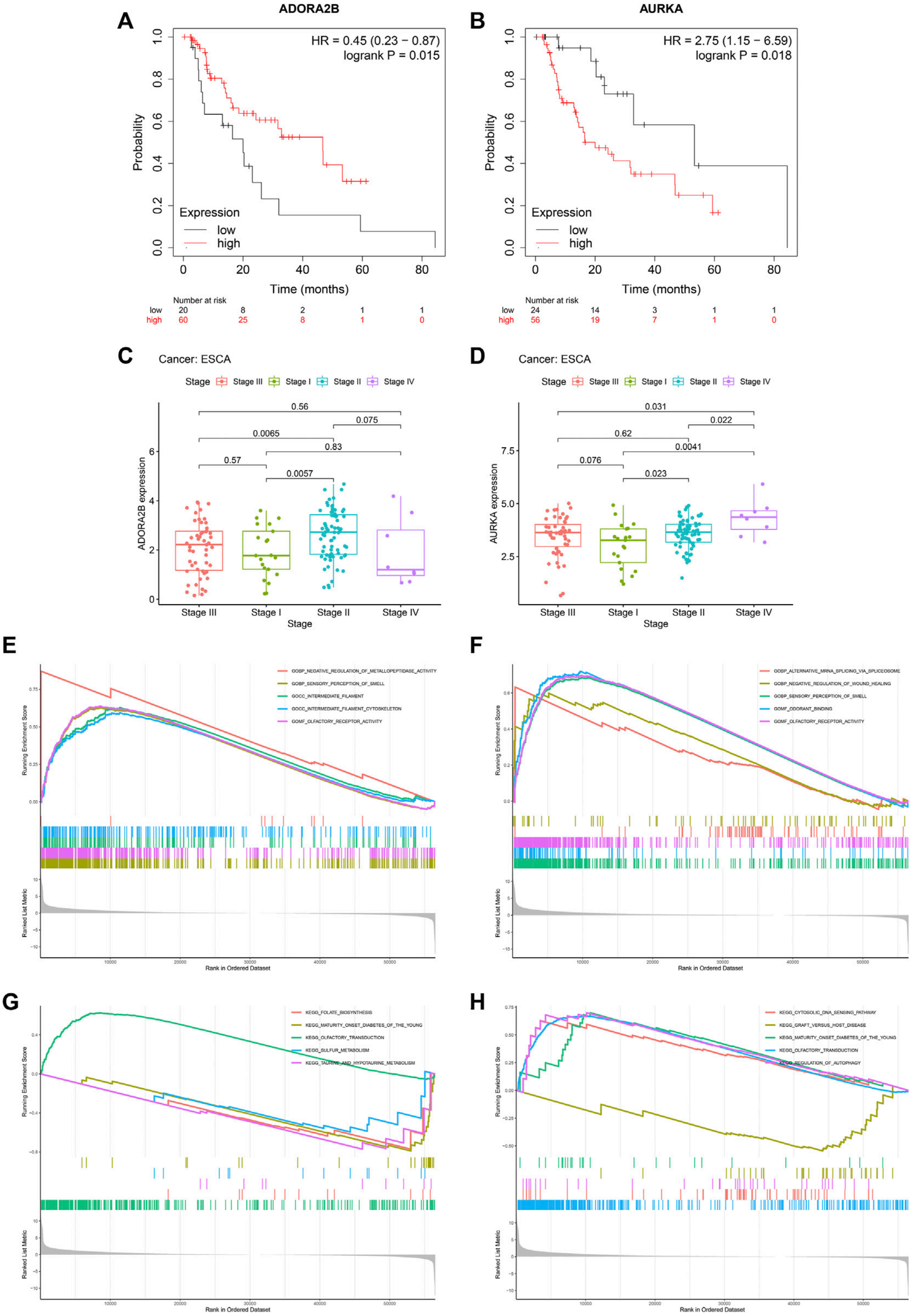


FIGURE 4
DHA targets against ESCA (ADORA2B and AURKA) correlate to prognosis and pathological stage of ESCA as well as tumorigenic pathways. **(A)** Kaplan-Meier OS curves between high and low ADORA2B expression subpopulations in TCGA-ESCA dataset. **(B)** Kaplan-Meier OS curves between high and low AURKA expression subpopulations in TCGA-ESCA dataset. **(C,D)** ADORA2B and AURKA expression across distinct pathological stages. **(E,F)** GO enrichment results of ADORA2B and AURKA. **(G,H)** KEGG pathway enrichment results of ADORA2B and AURKA.

TABLE 2 Molecular docking between DHA and potential targets.

Domain	Compound	Vina ($kcal \cdot mol^{-1}$)	RMSD	DS(LibDockScore)	Hydrogen bond interaction	Hydrophobic interaction
MMP9(1GKC)	DHA	−7.2	1.221	92.2116	TYR:420,TYR:423,PRO:421	LEU:188,VAL:398,HIS:401,HIS:405,HIS:411,LEU:187
CA2(3F6U)	DHA	−5.9	2.163	76.5684	TRP:215,GLU:192	HIS:57,CYS:220,CYS:191
FBP1(7CVH)	DHA	−6.4	1.678	99.8336	ALA:190,SER:46,SER:47	ARG:50,PRO:189,ALA:52,ILE:191,LYS:43,LEU:187
CCNA1(2G9X)	DHA	−9.0	1.488	80.4867	HIS:296	LEU:297,HIS:71,LYS:300
CCNA2(1OI9)	DHA	−9.6	1.068	74.8124	HIS:71,LYS:300,SER:0	HIS:296,LEU:297,ILE:70
CDK2 (1GIH)	DHA	−9.6	0.978	79.1237	GLN:131	ILE:10,LEU:134,ALA:31,VAL:18,VAL:64,ALA:144,PHE:80
AOC3(2Y74)	DHA	−8.7	1.791	94.64	ARG:488	LEU:608,PHE:610,PHE:704
MMP1(1SU3)	DHA	72.6	1.825	86.2719	PRO:90,SER:239,TYR:240,ASN:315	PHE:316,TYR:237,VAL:319
TOP2A (5NNE)	DHA	−4.0	1.153	66.8962	GLN:85	VAL:87,ILE:146,LEU:92
AURKA (5DOS)	DHA	−8.9	1.397	86.7565	ASN:261,ASP:274,LYS:162	MG:401,ALA:273,LEU:194,LEU:210,LEU:263,LEU:139,ALA:160,VAL:147
CDK1 (6GU6)	DHA	−8.8	1.877	50.3516	-	VAL:174,PRO:62
CCNB1(4Y72)	DHA	−10.4	1.311	92.661	ASP:146,TYR:15,GLY:13	ARG:158,LEU:149,VAL:164,VAL:165
CCNB2(5LQF)	DHA	−9.4	1.340	85.0795	GLU:221,LYS:324,ARG:188,ARG:307	LEU:193,HIS:320,LEU:303,PRO:189
KCNQ1(6B8M)	DHA	7.8	1.842	71.392	ALA:545	LYS:541,VAL:544,LYS:548,ALA:348
ADORA2B (5UEN)	DHA	−8.4	1.811	88.3329	ALA:84	ALA:66,VAL:62,LEU:65,PHE:171,VAL:174,CYS:80,CYS:169,ILE:69
GABRE(6DW0)	DHA	10.6	1.637	80.0045	ASP:56,GLN:185	HIS:55,LYS:278,PRO:277,LEU:183,PRO:184,MET:49

RMSD, root mean square deviation; DHA, dihydroartemisinin.

independently predicted ESCA cases' OS. As a result, the prognostic signature acted as an independent risk factor (Figure 2K).

Dihydroartemisinin targets against esophageal carcinoma correlate to cell cycle progression

Biological features of DHA targets against ESCA were probed out utilizing GO and KEGG enrichment approach. For biological process, histone phosphorylation, mitotic DNA integrity checkpoint, G2/M transition of mitotic cell cycle, cell cycle G2/M phase transition, mitotic nuclear envelope disassembly, DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest, cell cycle G1/S phase transition, signal transduction involved in mitotic G1 DNA damage checkpoint, intracellular signal transduction involved in G1 DNA damage checkpoint and signaling transduction involved in mitotic cell cycle checkpoint were significantly linked to DHA targets against ESCA (Figures 3A,B). Cellular components of cyclin-dependent protein kinase holoenzyme complex, serine/threonine protein kinase complex, protein kinase complex, transferase complex, transferring phosphorus-containing groups, condensed chromosome, pronucleus, chromosomal region, condensed nuclear chromosome, centromeric region, spindle microtubule, and microvillus were remarkably enriched by DHA targets against ESCA (Figures 3C,D). Also, they correlated to molecular functions of histone kinase activity, cyclin-dependent serine/threonine kinase regulator activity, protein kinase regulator activity, kinase regulator activity, cyclin-dependent protein serine/threonine kinase activity, cyclin-dependent protein kinase activity, cyclin binding, protein heterodimerization activity, metalloendopeptidase activity as well as protein serine/threonine kinase activity (Figures 3E,F). As illustrated in KEGG pathway enrichment results, progesterone-mediated oocyte maturation, cell cycle, cellular senescence, oocyte meiosis, p53 signaling pathway, hepatitis B, viral carcinogenesis, human T-cell leukemia virus one infection, AMPK signaling pathway as well as FoxO signaling pathway (Figures 3G,H). Altogether, DHA targets against ESCA might participate in mediating cell cycle progression.

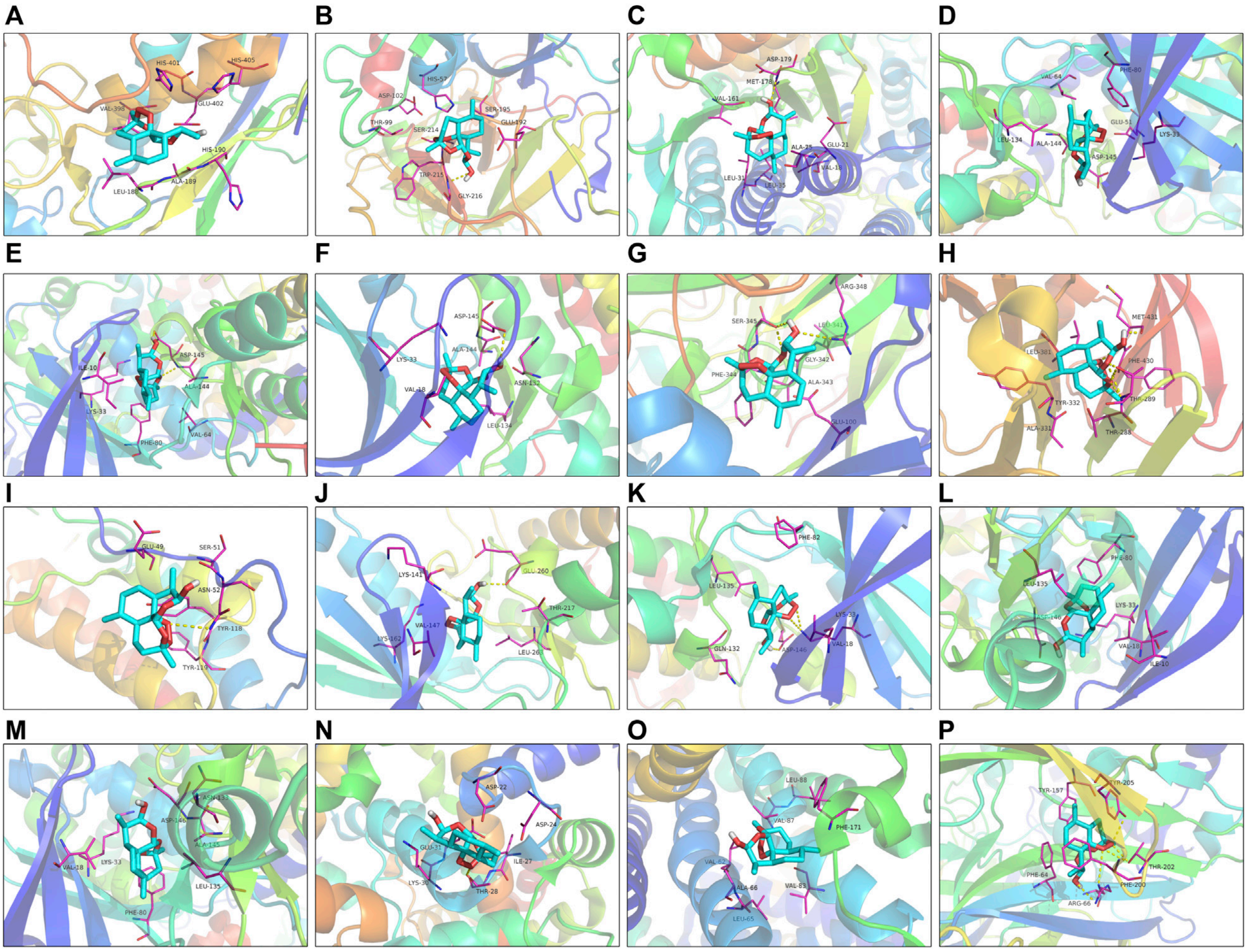


FIGURE 5
The 3D molecular docking models between DHA and potential targets. **(A)** MMP9; **(B)** CA2; **(C)** FBP1; **(D)** CCNA1; **(E)** CCNA2; **(F)** CDK2; **(G)** AOC3; **(H)** MMP1; **(I)** TOP2A; **(J)** AURKA; **(K)** CDK1; **(L)** CCNB1; **(M)** CCNB2; **(N)** KCNQ1; **(O)** ADORA2B; and **(P)** GABRE.

Dihydroartemisinin targets against esophageal carcinoma (AURKA and ADORA2B) correlates to prognosis and pathological stage of esophageal carcinoma

Among DHA targets against ESCA, DORA2B, and AURKA significantly correlated to ESCA prognosis. Up-regulated ADORA2B was linked to better OS, while up-regulated AURKA was correlated to worse OS in TCGA-ESCA dataset (Figures 4A,B). Further analysis for the associations of ADORA2B and AURKA with pathological stage of ESCA was carried out. In Figure 4C, stage II cases exhibited higher ADORA2B expression. Additionally, AURKA expression displayed positive associations to pathological stage (Figure 4D), indicating that AURKA was linked to ESCA progression.

Mechanisms underlying ADORA2B and AURKA

Mechanisms underlying ADORA2B and AURKA were further probed out utilizing GSEA. For GO enrichment results, negative regulation of metallopeptidase activity, sensory perception of smell, intermediate filament, intermediate filament cytoskeleton and olfactory receptor activity exhibited positive interactions to ADORA2B (Figure 4E). Meanwhile, alternative mRNA splicing *via* spliceosome, negative regulation of wound healing, sensory perception of smell, odorant binding and olfactory receptor activity displayed positive correlations to AURKA (Figure 4F). For KEGG pathway enrichment results, folate biosynthesis, maturity onset diabetes of the young, olfactory transduction, sulfur metabolism as well as taurine and hypotaurine metabolism were positively linked to

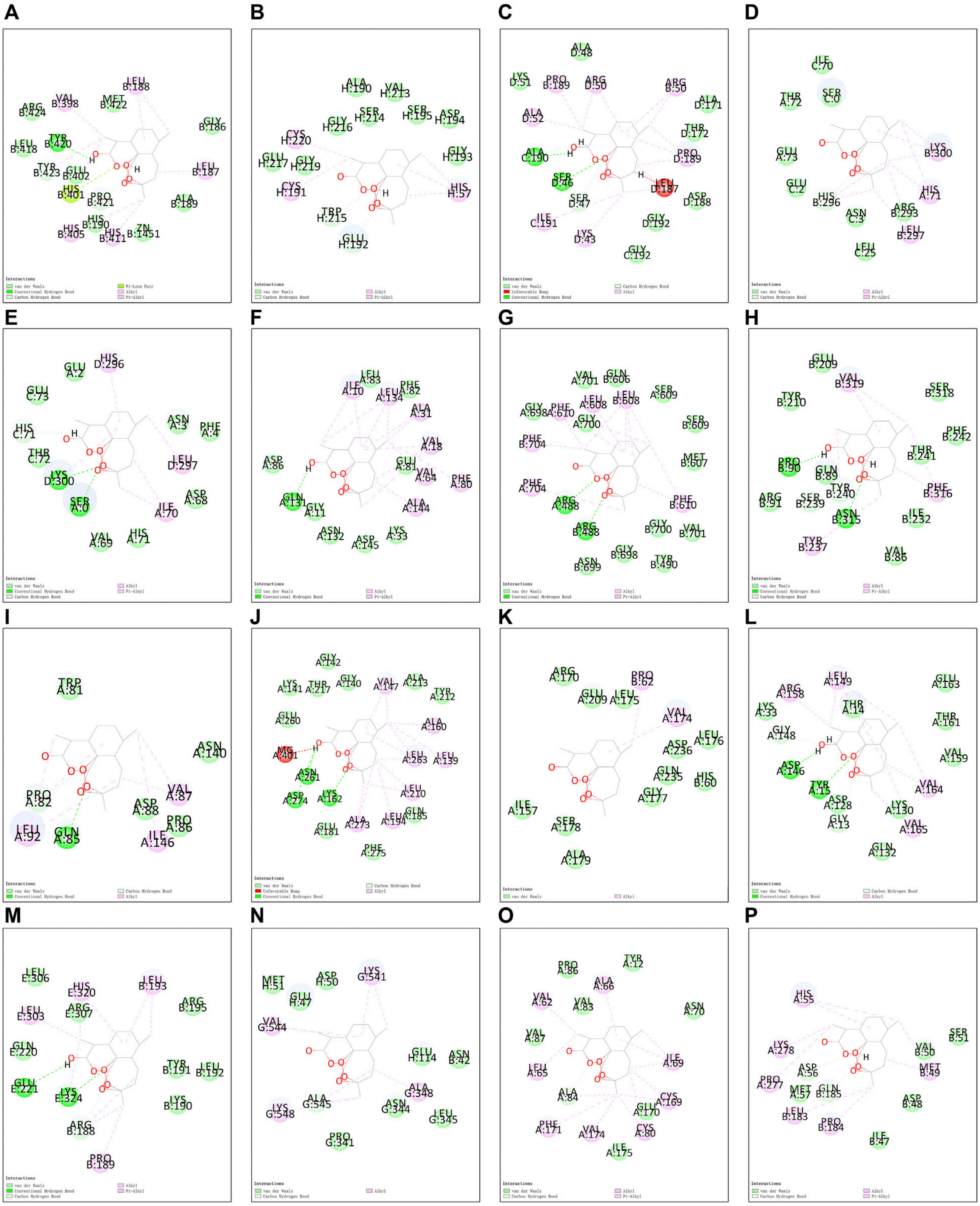


FIGURE 6
The 2D molecular docking models between DHA and potential targets. **(A)** MMP9; **(B)** CA2; **(C)** FBP1; **(D)** CCNA1; **(E)** CCNA2; **(F)** CDK2; **(G)** AOC3; **(H)** MMP1; **(I)** TOP2A; **(J)** AURKA; **(K)** CDK1; **(L)** CCNB1; **(M)** CCNB2; **(N)** KCNQ1; **(O)** ADORA2B; and **(P)** GABRE.

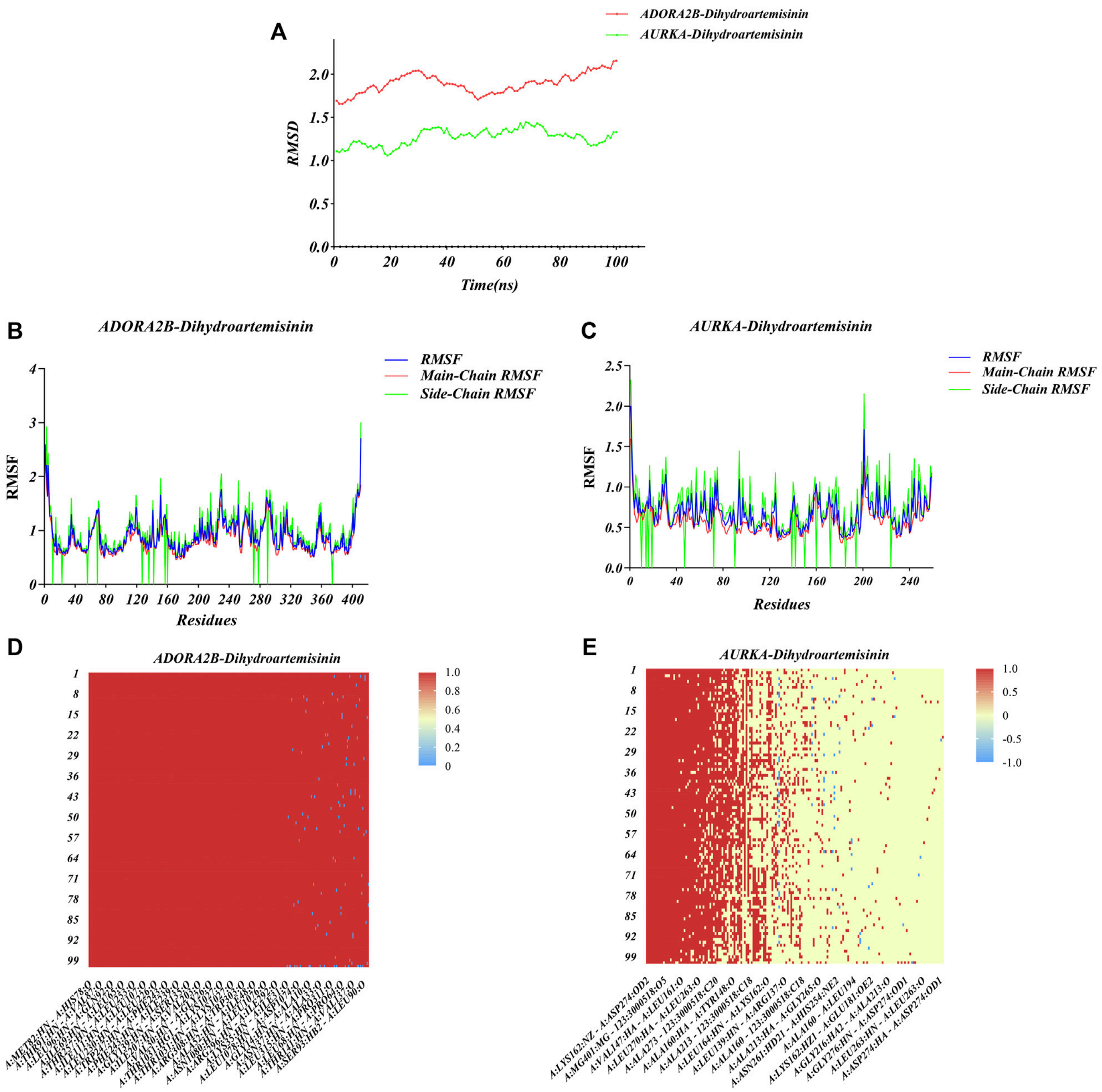


FIGURE 7
Molecular dynamics simulation of ADORA2B-DHA and AURKA-DHA complexes. **(A)** Alterations in RMSD values during the process of molecular dynamics simulation of ADORA2B-DHA (red) and AURKA-DHA (green) complexes. **(B,C)** Alterations in RMSF values during the process of molecular dynamics simulation of **(B)** ADORA2B-DHA and **(C)** AURKA-DHA complexes. **(D,E)** Heatmaps of hydrogen bonding interactions of **(D)** ADORA2B-DHA and **(E)** AURKA-DHA complexes.

ADORA2B (Figure 4G). Also, cytosolic DNA sensing pathway, graft *versus* host disease, maturity onset diabetes of the young, olfactory transduction, and regulation of autophagy exhibited positive correlations to AURKA (Figure 4H).

Molecular docking between dihydroartemisinin and potential targets

Utilizing Chem3D, we obtained the 3D structure of DHA with mol*2 format, and downloaded the 3D structure of

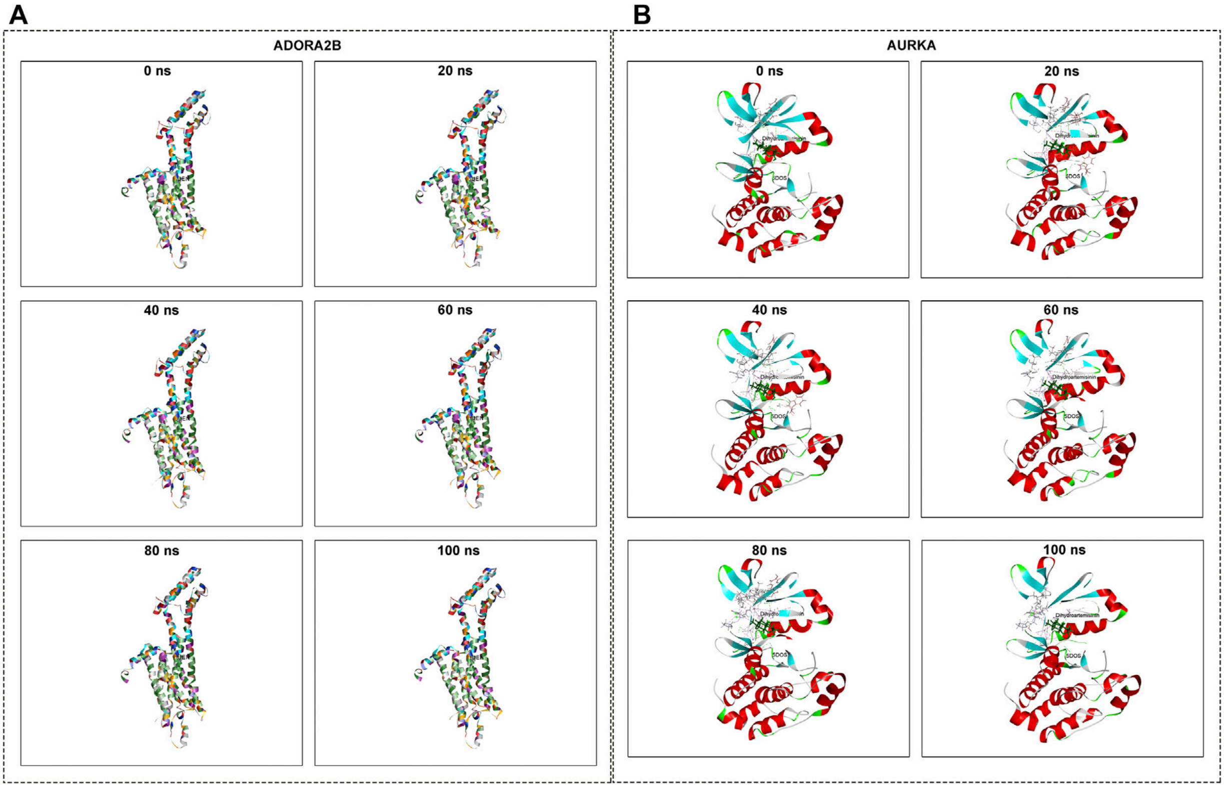


FIGURE 8
Molecular dynamics simulation of ADORA2B-DHA and AURKA-DHA complexes. **(A)** ADORA2B-DHA complex; **(B)** AURKA-DHA complex.

potential targets from the PDB database with pdb format. Through AutoDockTools 1.5.6 software, the 3D structure of DHA and potential targets were converted to pdbqt format, and thus searched for the active pocket. Afterwards, we run the Vina script to compute the binding energies of DHA and potential targets. As listed in Table 2, the binding energies of the docking bodies formed by DHA targets against ESCA (MMP9, CA2, FBP1, CCNA1, CCNA2, CDK2, AOC3, MMP1, TOP2A, AURKA, CDK1, CCNB1, CCNB2, KCNQ1, ADORA2B, and GABRE) and DHA were all $<-5.0\ \mathrm{kcal\ mol^{-1}}$, indicative of the stable docking. Additionally, we computed LibDockScore to dock DHA and its potential targets by use of Discovery Studio 2019 software. As a result, in each potential target, docking site of DHA can be found, and the LibDockScore of the docking model formed by DHA and each target was >50. The 3D and 2D molecular docking models between DHA and potential targets (MMP9, CA2, FBP1, CCNA1, CCNA2, CDK2, AOC3, MMP1, TOP2A, AURKA, CDK1, CCNB1, CCNB2, KCNQ1, ADORA2B, and GABRE) were separately displayed as Figures 5A–P and Figures 6A–P.

Molecular dynamics simulation

The conformations of ADORA2B-DHA and AURKADHA molecular docking were used for subsequent molecular dynamics simulation analysis. The ADORA2B-DHA complex was added by 16,738 water molecules, 44 sodium as well as 53 chloride, while the AURKA-DHA complex was added by 5,688 water molecules, 15 sodium as well as 20 chloride. We further investigated the structural stability of ADORA2B-DHA and AURKADHA complexes during molecular dynamics simulation. Meanwhile, the RMSD values of the two complexes were computed during 100 ns molecular dynamics simulation. The ADORA2B-DHA and AURKADHA complexes were stable following 100 ns molecular dynamics simulation, as illustrated in Figure 7A. The RMSD value of ADORA2B-DHA complex mainly fluctuated from 1.65612 to 2.15719, and the mean RMSD value was 1.89383; while the RMSD value of AURKA-DHA complex mainly fluctuated from 1.05842 to 1.44505, and its mean RMSD value was 1.26881 (Figure 7A). The RMSD fluctuation values of the two complexes were all within a reasonable range, demonstrating that the ADORA2B-DHA and AURKA-DHA complexes were in a stable state during the process of molecular dynamics simulation. For analyzing the volatility of distinct amino acids in the two complexes during the process of molecular dynamics simulation, we also computed the RMSF values of all amino acids during the simulation. ADORA2B-DHA complex fluctuated greatly around amino acids Ser6, Ala7, Phe8, Leu315, and Phe316 (Figure 7B); AURKA-DHA complex fluctuated greatly around amino acids Gln127, Trp128, Ala129, Phe329, Glu330, Ala331, and Asn332 (Figure 7C). The heatmaps of hydrogen bonds during the ADORA2B-DHA and AURKADHA molecular docking were separately exhibited in Figures 7D,E. Hydrogen bonding interactions were present in most conformations of ADORA2B-DHA and AURKADHA molecular docking, demonstrating that these hydrogen bonds were persistent and stable. The molecular dynamics simulation results of ADORA2B-DHA and AURKA-DHA complexes were separately displayed in Figures 8A,B.

Discussion

At present, the paradigms of development of single-targeted therapeutic agents are challenging, primarily because of lack of effectiveness as well as resistance. Therefore, natural compounds selectively acting on two or more targets exhibit higher efficacy in comparison to single-target drugs. The current study employed network pharmacology as well as molecular docking approaches to predict multi-targets and multi-pathways of DHA against ESCA.

Previous studies have proposed that DHA exhibits the anti-ESCA property through targeting distinct molecules as well as pathways (Li et al., 2019; Ma et al., 2020). Thus, we presented a systematic analysis for unveiling DHA targets against ESCA. Through integrating the SwissTargetPrediction, HERB and BATMAN-TCM platforms, 160 DHA targets were finally acquired. Meanwhile, 821 DEGs between 161 ESCA and 11 normal tissue specimens in TCGA-ESCA dataset were screened, which might participate in ESCA initiation as well as progression. Following the intersection of DHA targets and DEGs, we eventually determined 16 DHA targets against ESCA, comprising MMP9, CA2, FBP1, CCNA1, CCNA2, CDK2, AOC3, MMP1, TOP2A, AURKA, CDK1, CCNB1, CCNB2, KCNQ1, ADORA2B, and GABRE. These DHA targets against ESCA remarkably correlated to cell cycle progression. Evidence demonstrates that DHA triggers cell cycle arrest during tumor progression (Lin et al., 2016; Saidi et al., 2021).

Among them, ADORA2B as well as AURKA correlated to ESCA cases’ OS. AURKA is a membrane of serine/threonine kinase family, and its activation is essential for cell division process by modulating mitosis (Du et al., 2021). AURKA accelerates ESCC progression *via* improving the activity of distinct pathways such as EGFR-PI3K-Akt (Du et al., 2020; Shi et al., 2021). Nevertheless, whether ADORA2B participates in ESCA remains uncharted. On the basis of ADORA2B as well as AURKA, a prognostic signature was developed for survival prediction. Despite the reliability and independency of the prognostic signature in ESCA cases’ OS, large prospective ESCA cohorts are required for verifying it.

Molecular docking uses mapping software to place small molecule compounds on the binding region of macromolecular targets, and then calculates parameters to predict the binding ability and binding mode of the two (Tao et al., 2020). Through the strength of the binding ability, the possible mechanisms of action of the small molecule compound can be preliminarily inferred, and the interaction mode between the small molecule compound and the potential target can be quickly and accurately described, which provides a scientific basis for the preparation of derivatives, and thus shortens the drug development cycle and reduces research and development costs. In this study, through use of molecular docking, the stable binding of DHA to 16 potential targets were verified, and thus speculated the main anti-ESCA targets and interactions. Molecular dynamics simulation can simulate the combination of the two in the natural state, and can restore some reaction processes that cannot be detected by current technical means. We employed molecular dynamics simulations for confirming the stability of the binding of DHA to protein targets (ADORA2B and AURKA). Despite this, experiments are essential for confirming ADORA2B-DHA and AURKA-DHA interactions in ESCA.

Conclusion

Collectively, the current research unveiled that DHA exerted an anti-ESCA effect through targeting multiple targets (especially ADORA2B and AURKA) as well as multiple pathways (especially cell cycle progression) on the basis of network pharmacology and molecular docking approaches, and thus provided the theoretical basis for the pharmacological research of DHA against ESCA.

Data availability statement

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

Author contributions

HW conceived and designed the study and conducted all the experiments and data analysis, and wrote the manuscript. All authors reviewed and approved the manuscript.

Conflict of interest

The author declares 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.2022.1017520/full#supplementary-material

SUPPLEMENTARY TABLE S1
DHA targets predicted by the SwissTargetPrediction, HERB, as well as BATMAN-TCM platforms.

SUPPLEMENTARY TABLE S2
The specific information of 821 DEGs between 161 ESCA and 11 normal tissue specimens in TCGA-ESCA dataset.

References

Bartha, Á., and Győrffy, B. (2021). TNMplot.com: A web tool for the comparison of gene expression in normal, tumor and metastatic tissues. *Ijms* 22 (5), 2622. doi:10.3390/ijms22052622

Chen, L., Niu, X., Qiao, X., Liu, S., Ma, H., Shi, X., et al. (2021). Characterization of interplay between autophagy and ferroptosis and their synergistical roles on manipulating immunological tumor microenvironment in squamous cell carcinomas. *Front. Immunol.* 12, 739039. doi:10.3389/fimmu.2021.739039

Chen, X., He, L. Y., Lai, S., and He, Y. (2020). Dihydroartemisinin inhibits the migration of esophageal cancer cells by inducing autophagy. *Oncol. Lett.* 20 (4), 1. doi:10.3892/ol.2020.11955

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, S11. doi:10.1186/1752-0509-8-s4-s11

Cui, W., Fang, T., Duan, Z., Xiang, D., Wang, Y., Zhang, M., et al. (2020). Dihydroartemisinin sensitizes esophageal squamous cell carcinoma to cisplatin by inhibiting sonic Hedgehog signaling. *Front. Cell Dev. Biol.* 8, 596788. doi:10.3389/fcell.2020.596788

Daina, A., Michielin, O., and Zoete, V. (2019). SwissTargetPrediction: Updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* 47 (W1), W357–w364. doi:10.1093/nar/gkz382

Du, R., Huang, C., Chen, H., Liu, K., Xiang, P., Yao, N., et al. (2020). SDCBP/MDA-9/syntenin phosphorylation by AURKA promotes esophageal squamous cell carcinoma progression through the EGFR-PI3K-Akt signaling pathway. *Oncogene* 39 (31), 5405–5419. doi:10.1038/s41388-020-1369-2

Du, R., Huang, C., Liu, K., Li, X., and Dong, Z. (2021). Targeting AURKA in cancer: Molecular mechanisms and opportunities for cancer therapy. *Mol. Cancer* 20 (1), 15. doi:10.1186/s12943-020-01305-3

Fang, S., Dong, L., Liu, L., Guo, J., Zhao, L., Zhang, J., et al. (2021). Herb: A high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res.* 49 (D1), D1197–d1206. doi:10.1093/nar/gkaa1063

Hulshof, M., Geijsen, E. D., Rozema, T., Oppedijk, V., Buijsen, J., Neelis, K. J., et al. (2021). Randomized study on dose escalation in definitive chemoradiation for patients with locally advanced esophageal cancer (ARTDECO study). *Jco* 39 (25), 2816–2824. doi:10.1200/jco.20.03697

Janjigian, Y. Y., Shitara, K., Moehler, M., Garrido, M., Salman, P., Shen, L., et al. (2021). First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): A randomised, open-label, phase 3 trial. *Lancet* 398 (10294), 27–40. doi:10.1016/s0140-6736(21)00797-2

Jiang, M., Wu, Y., Qi, L., Li, L., Song, D., Gan, J., et al. (2021). Dihydroartemisinin mediating PKM2-caspase-8/3-GSDME axis for pyroptosis in esophageal squamous cell carcinoma. *Chemico-Biological Interact.* 350, 109704. doi:10.1016/j.cbi.2021.109704

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49 (D1), D1388–d1395. doi:10.1093/nar/gkaa971

Li, Q., Ma, Q., Xu, L., Gao, C., Yao, L., Wen, J., et al. (2021). Human telomerase reverse transcriptase as a therapeutic target of dihydroartemisinin for esophageal squamous cancer. *Front. Pharmacol.* 12, 769787. doi:10.3389/fphar.2021.769787

Li, S., Huang, P., Gan, J., Ling, X., Du, X., Liao, Y., et al. (2019). Dihydroartemisinin represses esophageal cancer glycolysis by down-regulating pyruvate kinase M2. *Eur. J. Pharmacol.* 854, 232–239. doi:10.1016/j.ejphar.2019.04.018

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lin, R., Zhang, Z., Chen, L., Zhou, Y., Zou, P., Feng, C., et al. (2016). Dihydroartemisinin (DHA) induces ferroptosis and causes cell cycle arrest in head and neck carcinoma cells. *Cancer Lett.* 381 (1), 165–175. doi:10.1016/j.canlet.2016.07.033

Liu, X., Zhang, M., Ying, S., Zhang, C., Lin, R., Zheng, J., et al. (2017). Genetic Alterations in esophageal tissues from squamous dysplasia to carcinoma. *Gastroenterology* 153 (1), 166–177. doi:10.1053/j.gastro.2017.03.033

Liu, Z., Guo, F., Wang, Y., Li, C., Zhang, X., Li, H., et al. (2016). BATMAN-TCM: A bioinformatics analysis tool for molecular mechANism of traditional Chinese medicine. *Sci. Rep.* 6, 21146. doi:10.1038/srep21146

Lu, Z., Wang, J., Shu, Y., Liu, L., Kong, L., Yang, L., et al. (2022). Sintilimab versus placebo in combination with chemotherapy as first line treatment for locally advanced or metastatic oesophageal squamous cell carcinoma (ORIENT-15): Multicentre, randomised, double blind, phase 3 trial. *Bmj* 377, e068714. doi:10.1136/bmj-2021-068714

Luo, H., Lu, J., Bai, Y., Mao, T., Wang, J., Fan, Q., et al. (2021). Effect of camrelizumab vs placebo added to chemotherapy on survival and progression-free survival in patients with advanced or metastatic esophageal squamous cell carcinoma. *Jama* 326 (10), 916–925. doi:10.1001/jama.2021.12836

Ma, Q., Liao, H., Xu, L., Li, Q., Zou, J., Sun, R., et al. (2020). Autophagy-dependent cell cycle arrest in esophageal cancer cells exposed to dihydroartemisinin. *Chin. Med.* 15, 37. doi:10.1186/s13020-020-00318-w

Mukherjee, S., Hurt, C., Radhakrishna, G., Gwynne, S., Bateman, A., Gollins, S., et al. (2021). Oxaliplatin/capecitabine or carboplatin/paclitaxel-based preoperative chemoradiation for resectable oesophageal adenocarcinoma (NeoSCOPE): Long-term results of a randomised controlled trial. *Eur. J. Cancer* 153, 153–161. doi:10.1016/j.ejca.2021.05.020

Niu, X., Chen, L., Li, Y., Hu, Z., and He, F. (2022). Ferroptosis, necroptosis, and pyroptosis in the tumor microenvironment: Perspectives for immunotherapy of SCLC. *Seminars Cancer Biol.* doi:10.1016/j.semcancer.2022.03.009

Niu, X., Huang, B., Qiao, X., Liu, J., Chen, L., and Zhong, M. (2021). MicroRNA-1-3p suppresses malignant phenotypes of ameloblastoma through down-regulating lysosomal associated membrane protein 2-mediated autophagy. *Front. Med.* 8, 670188. doi:10.3389/fmed.2021.670188

Safran, H. P., Winter, K., Ilson, D. H., Wigle, D., DiPetrillo, T., Haddock, M. G., et al. (2022). Trastuzumab with trimodality treatment for oesophageal adenocarcinoma with HER2 overexpression (NRG oncology/RTOG 1010): A multicentre, randomised, phase 3 trial. *Lancet Oncol.* 23 (2), 259–269. doi:10.1016/s1470-2045(21)00718-x

Saidi, H., Murtaza, B., Khan, A. S., Koceir, E. A., Hichami, A., and Khan, N. A. (2021). DHA induces Jurkat T-cell arrest in G2/M phase of cell cycle and modulates the plasma membrane expression of TRPC3/6 channels. *Biochimie* 181, 169–175. doi:10.1016/j.biochi.2020.12.005

Shi, K., Zhang, J. Z., Yang, L., Li, N. N., Yue, Y., Du, X. H., et al. (2021). Protein deubiquitylase USP3 stabilizes Aurora A to promote proliferation and metastasis of esophageal squamous cell carcinoma. *BMC Cancer* 21 (1), 1196. doi:10.1186/s12885-021-08934-x

Shitara, K., Özgüroğlu, M., Bang, Y. J., Di Bartolomeo, M., Mandalà, M., Ryu, M. H., et al. (2021). Molecular determinants of clinical outcomes with pembrolizumab versus paclitaxel in a randomized, open-label, phase III trial in patients with gastroesophageal adenocarcinoma. *Ann. Oncol.* 32 (9), 1127–1136. doi:10.1016/j.annonc.2021.05.803

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

Sun, J. M., Shen, L., Shah, M. A., Enzinger, P., Adenis, A., Doi, T., et al. (2021). Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): A randomised, placebo-controlled, phase 3 study. *Lancet* 398 (10302), 759–771. doi:10.1016/s0140-6736(21)01234-4

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

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi:10.1093/nar/gku1003

Tao, Q., Du, J., Li, X., Zeng, J., Tan, B., Xu, J., et al. (2020). Network pharmacology and molecular docking analysis on molecular targets and mechanisms of Huashi Baidu formula in the treatment of COVID-19. *Drug Dev. Industrial Pharm.* 46 (8), 1345–1353. doi:10.1080/03639045.2020.1788070

UniProt (2021). The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–d489. doi:10.1093/nar/gkaa1100()

Xu, R. H., Zhang, Y., Pan, H., Feng, J., Zhang, T., Liu, T., et al. (2021). Efficacy and safety of weekly paclitaxel with or without ramucirumab as second-line therapy for the treatment of advanced gastric or gastroesophageal junction adenocarcinoma (RAINBOW-Asia): A randomised, multicentre, double-blind, phase 3 trial. *Lancet Gastroenterology Hepatology* 6 (12), 1015–1024. doi:10.1016/s2468-1253(21)00313-7

Yang, Y., Li, B., Yi, J., Hua, R., Chen, H., Tan, L., et al. (2022). Robot-assisted versus conventional minimally invasive esophagectomy for resectable esophageal squamous cell carcinoma. *Ann. Surg.* 275 (4), 646–653. doi:10.1097/sla.0000000000005023

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

Zhang, X., Liang, H., Li, Z., Xue, Y., Wang, Y., Zhou, Z., et al. (2021). Perioperative or postoperative adjuvant oxaliplatin with S-1 versus adjuvant oxaliplatin with capecitabine in patients with locally advanced gastric or gastro-oesophageal junction adenocarcinoma undergoing D2 gastrectomy (RESOLVE): An open-label, superiority and non-inferiority, phase 3 randomised controlled trial. *Lancet Oncol.* 22 (8), 1081–1092. doi:10.1016/s1470-2045(21)00297-7

Zhang, X., Wang, Y., and Meng, L. (2022). Comparative genomic analysis of esophageal squamous cell carcinoma and adenocarcinoma: New opportunities towards molecularly targeted therapy. *Acta Pharm. Sin. B* 12 (3), 1054–1067. doi:10.1016/j.apsb.2021.09.028

Zhu, L., Chen, X., Zhu, Y., Qin, J., Niu, T., Ding, Y., et al. (2020). Dihydroartemisinin inhibits the proliferation of esophageal squamous cell carcinoma partially by targeting AKT1 and p70S6K. *Front. Pharmacol.* 11, 587470. doi:10.3389/fphar.2020.587470

Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fgene.2022.1043870

OPEN ACCESS

EDITED BY
Xing Niu,
China Medical University, China

REVIEWED BY
Ji Lv,
Maternity and Child Health Hospital of Qinhuangdao, China
Xingling Wang,
Liaoning Cancer Hospital, China Medical University, China

*CORRESPONDENCE
Yan Wang,
wwyy3196@ccmu.edu.cn

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

RECEIVED 14 September 2022
ACCEPTED 23 November 2022
PUBLISHED 08 December 2022

CITATION
Ji C, He Y and Wang Y (2022), Identification of necroptosis subtypes and development of necroptosis-related risk score model for in ovarian cancer.
Front. Genet. 13:1043870.
doi: 10.3389/fgene.2022.1043870

Identification of necroptosis subtypes and development of necroptosis-related risk score model for in ovarian cancer

Chen Ji, Yue He and Yan Wang*

Department of Gynecological Oncology, Beijing Obstetrics and Gynecology Hospital, Beijing Maternal and Child Health Care Hospital, Capital Medical University, Beijing, China

Background: ith the ongoing development of targeted therapy, non-apoptotic cell death, including necroptosis, has become a popular topic in the field of prevention and treatment. The purpose of this study was to explore the effect of necroptosis-related genes (NRGs) on the classification of ovarian cancer (OV) subtypes and to develop a necroptosis-related risk score (NRRS) classification system.

Methods: 74 NRGs were obtained from the published studies, and univariate COX regression analysis was carried out between them and OV survival. Consensus clustering analysis was performed on OV samples according to the expression of NRGs related to prognosis. Furthermore, the NRRS model was developed by combining Weighted Gene Co-Expression Network Analysis (WGCNA) with least absolute shrinkage and selection operator (Lasso)-penalized Cox regression and multivariate Cox regression analysis. And the decision tree model was constructed based on the principle of random forest screening factors principle.

Results: According to the post-related NRGs, OV was divided into two necroptosis subtypes. Compared with Cluster 1 (C1), the overall survival (OS) of Cluster 2 (C2) was significantly shorter, stromal score and immune score, the infiltration level of tumor associated immune cells and the expression of 20 immune checkpoints were significantly higher. WGCNA identified the blue module most related to necroptosis subtype, and 12 genes in the module were used to construct NRRS. NRRS was an independent prognostic variable of OV. The OS of samples with lower NRRS was significantly longer, and tumor mutation burden and homologous recombination defect were more obvious.

Conclusion: This study showed that necroptosis plays an important role in the classification, prognosis, immune infiltration and biological characteristics of OV subtypes. The evaluation of tumor necroptosis may provide a new perspective for OV treatment.

KEYWORDS

necroptosis, ovarian cancer, classification of subtypes, risk model, prognosis, immune infiltration

Introduction

Ovarian cancer (OV) is the deadliest cancer in the female reproductive tract (Whelan et al., 2022). Worldwide, 313,959 new diagnosed ovarian cancer and 207,252 succumb to this disease in 2020 (Sung et al., 2021). OV is usually confined to the peritoneal cavity throughout its course, with occasional distant metastases. Due to vague and non-specific signs and symptoms, and limited screening methods, the initial diagnosis is usually delayed until extensive intraperitoneal diffusion occurs through the adjacent peritoneal surface, ascites and rich lymphatic vessels (Achimas-Cadariu et al., 2022). According to statistics, about 3/4 of OV patients are diagnosed with advanced stage, and the prognosis is disappointing (Zhang et al., 2022). OV also faces a large number of unsolved problems such as difficult choice of treatment strategies and high recurrence rate (Rakina et al., 2022). Surgical treatment is currently recognized as the best method for the treatment of ovarian cancer, and platinum-paclitaxel chemotherapy as adjuvant therapy can significantly improve the effectiveness of ovarian cancer treatment. (van Stein et al., 2021; Wood and Ledermann, 2022). However, OV a highly heterogeneous at the molecular level, therefore, molecular targeted therapy is considered as a less toxic but more effective treatment in OV (Guan and Lu, 2018). For example, the combined application of PARP inhibitors, anti-VEGF monoclonal antibody and ICI has become a research hotspot (Revythis et al., 2022). Better understanding the biological characteristics and molecular heterogeneity of OV in order to formulate or improve treatment strategies and improve quality of life is an urgent demand (Zhang et al., 2022).

Apoptotic cell death plays an important role in OV (Hou et al., 2019). Ongoing development of targeted therapy allows non-apoptotic cell death to become popular in the field of prevention and treatment, including ferroptosis, alkaliptosis, autophagy, necroptosis, pyroptosis, immunogenic cell death as well as other cell death modes (Chen et al., 2022a). Necroptosis is a form of programmed necrosis, which differs from apoptosis as caspases activation is not involved in its progression. Instead, it is mediated by external signals, which trigger the activation of Mixed-Lineage Kinase Domain-Like (MLKL) signaling cascade, Receptor Interacting Protein 1 (RIP1), RIP3 (Liu et al., 2022). It is characterized by mitochondrial changes and plasma membrane permeability, resulting in the release of cytoplasmic contents into extracellular space and inflammation (Beretta and Zaffaroni, 2022). Preclinical and clinical evidence show that it is the outstanding pro-inflammatory characteristics of necroptosis that contribute to the correlation between necroptosis and cancer pathophysiology (Pasparakis and Vandenabeele, 2015). Necroptosis is regulated by molecular mechanism. Targeting necroptosis has shown substantial potential in tumor treatment with small molecules may have the advantage of bypassing the mechanism of apoptosis resistance (Wu et al., 2022). There is growing evidence that necrosis plays a key role in the development and progression of a wide range of diseases, including neurodegenerative diseases, ischemic cardiovascular disease and cancer metastases (Gong et al., 2019; Chen et al., 2022b). In addition, necrosis has a dual role in promoting and inhibiting tumor growth in a variety of tumor types (Seifert et al., 2016; Strilic et al., 2016; Qin et al., 2019). Therefore, from this point of view, the key molecular insights on necroptosis provide a prospect for targeted therapy. The key molecules of necroptosis have been poorly studied in OV.

In the current study, based on the cluster analysis of transcriptional profiles of necroptosis-related genes, we identified the necroptosis subtypes of OV, and described the clinical and molecular characteristics, immune characteristic and association with immunotherapy response. A necroptosis-related risk score (NRRS) model was developed by Weighted Gene Co-Expression Network Analysis (WGCNA) and least absolute shrinkage and selection operator (LASSO) Cox regression analysis, and a clinical decision tree model and nomogram were established to improve the risk stratification of survival in OV patients.

Materials and methods

Extraction and preprocessing of OV cohort data

The RNA sequencing (RNA-seq), somatic mutation, copy number alterations (CNAs) data and clinical follow-up information of OV were found and downloaded in The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) database. In International Cancer Genome Consortium (ICGC, https://dcc.icgc.org/), the samples with detailed RNA-seq and clinical survival data were also included in the analysis. Another OV cohorts (GSE26193, GSE30161, GSE63885, GSE9891) were collected from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database. The clinical features were showed in Table 1.

Consensus clustering analysis was performed on OV samples by obtaining necroptosis related genes

A study by Xin et al. (2022) gave 74 necroptosis-related genes (NRG). Univariate COX regression analysis was carried out to screen NRGs related to prognosis. R package “ConsensusClusterPlus” root conducted the unsupervised hierarchical clustering of OV according to expression of prognosis-related NRGs. Euclidean distance and “pam” were utilized to compute the similarity distance between samples, with 500 iterations and 80% resampling rate, ranging from

TABLE 1 The clinical features of datasets

	TCGA	ICGC	GSE26193	GSE30161	GSE63885	GSE9891
	(N=373)	(N=93)	(N=107)	(N=58)	(N=70)	(N=276)
OS						
0	143 (38.3%)	19 (20.4%)	31 (29.0%)	22 (37.9%)	4 (5.7%)	163 (59.1%)
1	230 (61.7%)	74 (79.6%)	76 (71.0%)	36 (62.1%)	66 (94.3%)	113 (40.9%)
Age						
Mean (SD)	59.6 (11.4)			62.6 (10.6)		59.6 (10.5)
Median [Min, Max]	59.0 [30.0, 87.0]			62.0 [38.0, 85.0]		59.0 [22.0, 80.0]
Stage						
I	1 (0.3%)		21 (19.6%)			24 (8.7%)
II	21 (5.6%)		10 (9.3%)		1 (1.4%)	17 (6.2%)
III	291 (78.0%)	79 (84.9%)	59 (55.1%)	53 (91.4%)	59 (84.3%)	212 (76.8%)
IV	57 (15.3%)	14 (15.1%)	17 (15.9%)	5 (8.6%)	10 (14.3%)	22 (8.0%)
Missing	3 (0.8%)					1 (0.4%)
Grade						
G1	1 (0.3%)					
G2	42 (11.3%)			25 (43.1%)	8 (11.4%)	
G3	319 (85.5%)			33 (56.9%)	44 (62.9%)	
G4	1 (0.3%)				18 (25.7%)	
Missing	10 (2.7%)					

2 to 10. The final optimal clustering number, was defined by the cumulative distribution function (CDF) curve and delta area, showed high consistency within the cluster, low variation coefficient without significant change in the area under the CDF curve.

Detection of tumor mutation

The "maftools" package (Mayakonda et al., 2018) was employed to analyze and visualize the single nucleotide variation (SNV) data processed by mutect2 in TCGA. Firstly, the genes with mutation frequency >3 in the sample were screened, and the statistical differences of high frequency mutation genes between subgroups were analyzed by fisher test, and the mutations of 20 genes with the highest mutation rate in different subgroups were shown by waterfall map. For number of segments and tumor mutation burden (TMB), fraction altered, homologous recombination defects, between subgroups, Wilcoxon test was used to compare.

Assessment of tumor immune microenvironment

The proportion of immune cells in tumor microenvironment (TME) was estimated by marker genes expressions-based microenvironment cell population (MCP) counter (Becht et al., 2016) and single sample gene set enrichment analysis (ssGSEA), and the results were expressed as enrichment scores. ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumours using Expression data) (Yoshihara et al., 2013) was used to calculate the stromal score and immune score and ESTIMATE score of the sample to quantify the overall level of TME matrix and infiltrating immune components.

Prediction of immunotherapy response

Immune checkpoint expression from HisgAtals and TIDE score from tumor immune dysfunction and exclusion (TIDE, http://tide.dfci.harvard.edu) (Jiang et al., 2018) were used to evaluate the immune checkpoint inhibitors treatment response between different OV subgroups. Different TIDE scores represent different sensitivities to immunotherapy, and low TIDE score is considered to be responsive to immunotherapy.

Weighted gene co-expression network analysis

To identify the key modules that are highly related to the OV subtypes defined by necroptosis, R package WGCNA (Langfelder and Horvath, 2008) was used to convert gene expression data into gene co-expression networks. The samples were clustered based on the Pearson correlation value between each gene pair and average linkage, and the best β was selected by using the "pickSoftThreshold" function to satisfy the scale-free distribution, and the correlation coefficient was more than 0.85. Adjacency matrix was created for correlation strength description among the nodes, and was further transformed into topological overlap matrix (TOM). Next, hierarchical clustering tree was constructed by dynamic hybrid cutting technology to identify modules (parameters: height = 0.25, deepSplit = 2, minModuleSize = 80). After merging similar modules, the modules with strong correlation with OV subtypes defined by necroptosis were identified.

Construction of necroptosis-related risk signature

The genes identified in the module were analyzed by univariate Cox regression, and the genes related to prognosis were included in the R packet "glmnet" for (LASSO Cox regression analysis. Then, the genes further screened by multivariate Cox regression analysis were used to construct risk models: NRRS = the sum of the multivariate LASSO regression coefficient of each gene × the normalized expression value of each gene transformed by log2 and z-score. To analyze the prediction effect of NRRS model on overall survival (OS), Kaplan-Meier survival curve and time-dependent receiver operating characteristic (tdROC) analyses were used.

Enrichment analysis

The candidate gene set was obtained from the hallmark database, and the $\log_2$FC value of each gene was input into GSEA software for gene set enrichment analysis (GSEA) to explore the biological pathway of sample enrichment. $P < 0.05$ was considered to be significantly enriched after adjusting for Enrichment Score (ES). False discovery rate (FDR) < 0.05was defined as the cutoff value. The upregulation pathway was defined based on normalized enrichment scores (NES) > 0, and the downregulation pathway was defined based on NES <0.

Construction of decision tree and nomogram

We used rpart package to build a decision tree based on age, stage, and grade and NRRS. Through the R package "rms," a nomogram was generated. To evaluate the consistency between actual survival and the predicted results, calibration curves were plotted. The net benefit and clinical usefulness of the nomogram and NRRS model were determined by decision curve analysis (DCA) and tdROC.

Statistical analysis

Statistical analysis was carried out by R 4.0.2 (https://www.r-project.org) to analyze the data and generate the results. R package "survminer" were performed to generate Kaplan-Meier curve, and "timeROC" were conducted to generate tdROC curve. The Wilcoxon rank-sum test were applied to compare continuous variables in two groups. Take $p < 0.05$ was the standard of statistical significance.

Results

Identification of necroptosis subtypes for OV

First of all, the sample expression data of four OV cohorts obtained from GEO were merged, and the deviation caused by batch effect was eliminated through the remove Batch Effect function of Limma package. Univariate Cox regression analysis of NRGs was carried out in the merged data, and 15 NRGs related to OV survival were identified. The merged OV samples were unsupervised consensus clustering. After comprehensive consideration of CDF curve and the delta area, KF2 was taken as the final number of clusters (Figures 1A,B). Consensus clustering was conducted in TCGA dataset (Supplementary Figure S1). Therefore, the two necroptosis subtypes that produce OV were cluster 1 (C1) and cluster 2 (C2). C1 was related to the better OV survival outcome of the GEO merger cohort (Figure 1C). And in the GSE with four cohorts, patients with survival accounted for more than 50% of C1, and patients with death accounted for a large proportion of C2 (Figure 1D). For OV samples in TCGA datasets, C1 was also associated with longer OS (Figure 1E). No statistical difference between the two subtypes in the proportion of patients with different survival states was found (Figure 1F). Although there was no significant difference in the proportion of age, stage, grade distribution between the two subtypes, it was obvious that there was a higher proportion of patients with age ≤60, stage Ⅳ and G3 in C2 (Figures 1G–I).

Characterization of the genetic variation for two necroptosis subtypes

The mutation data downloaded from TCGA were analyzed in two necroptosis subtypes. 2,614 genes with mutation frequency >3 were first screened out. A total of 54 genes with significantly different mutation rates were identified by

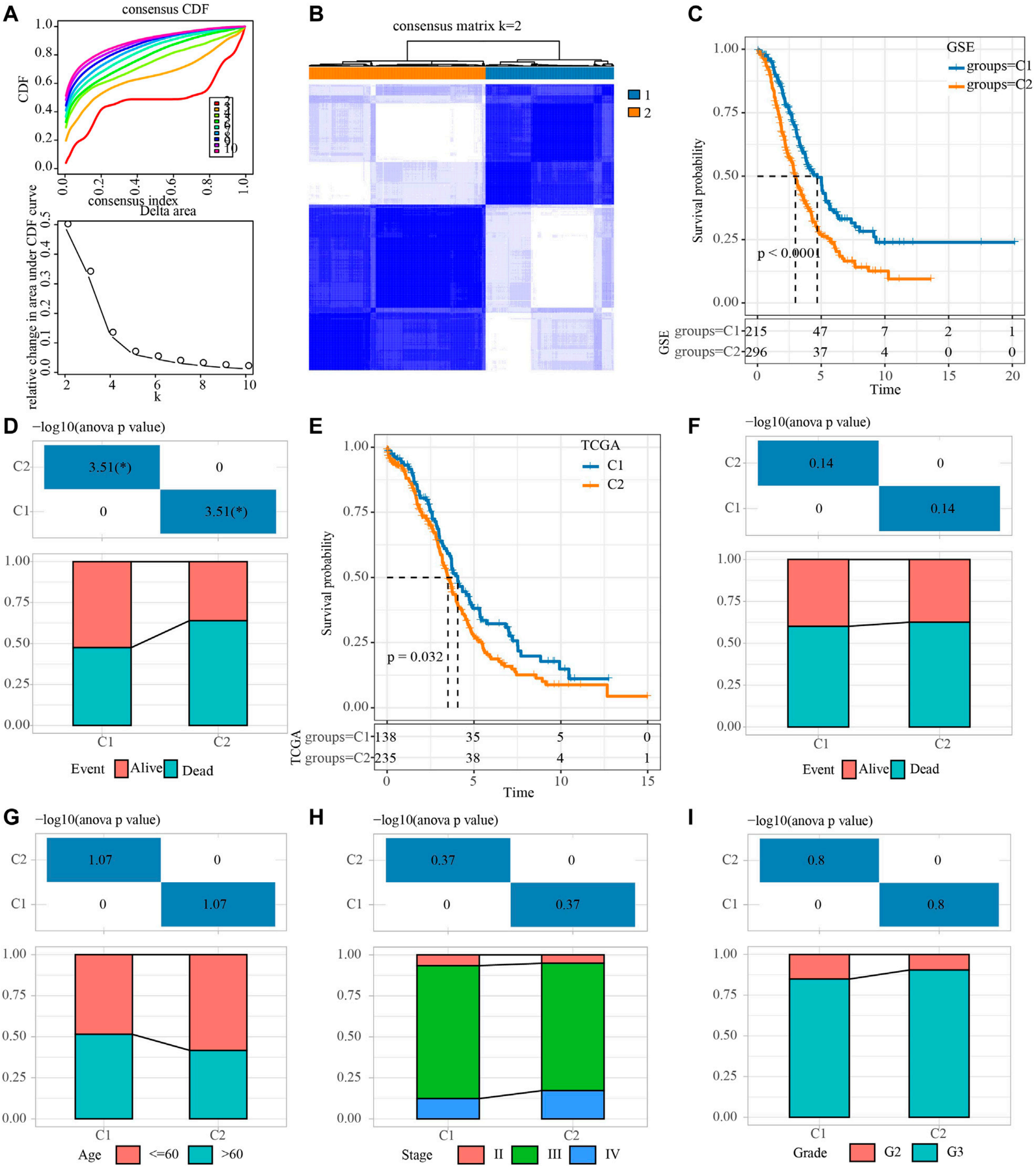


FIGURE 1

Identification of necroptosis subtypes for OV. **(A)** Consensus clustering heatmap for two subgroups. **(B)** CDF curve and the delta area of the clustering result. **(C)** Kaplan-Meier curve for LUAD patients in GSE dataset that merged four OV cohorts. **(D)** The distribution proportion of samples with different survival states in the two necroptosis subtypes of the GSE dataset that merged four OV cohorts. **(E)** Survival curve for LUAD patients in TCGA dataset. **(F)** Analysis of different survival states of two necroptosis subtypes in TCGA dataset. **(G–I)**: The characteristics of age, stage and grade of two necroptosis subtypes in TCGA dataset.

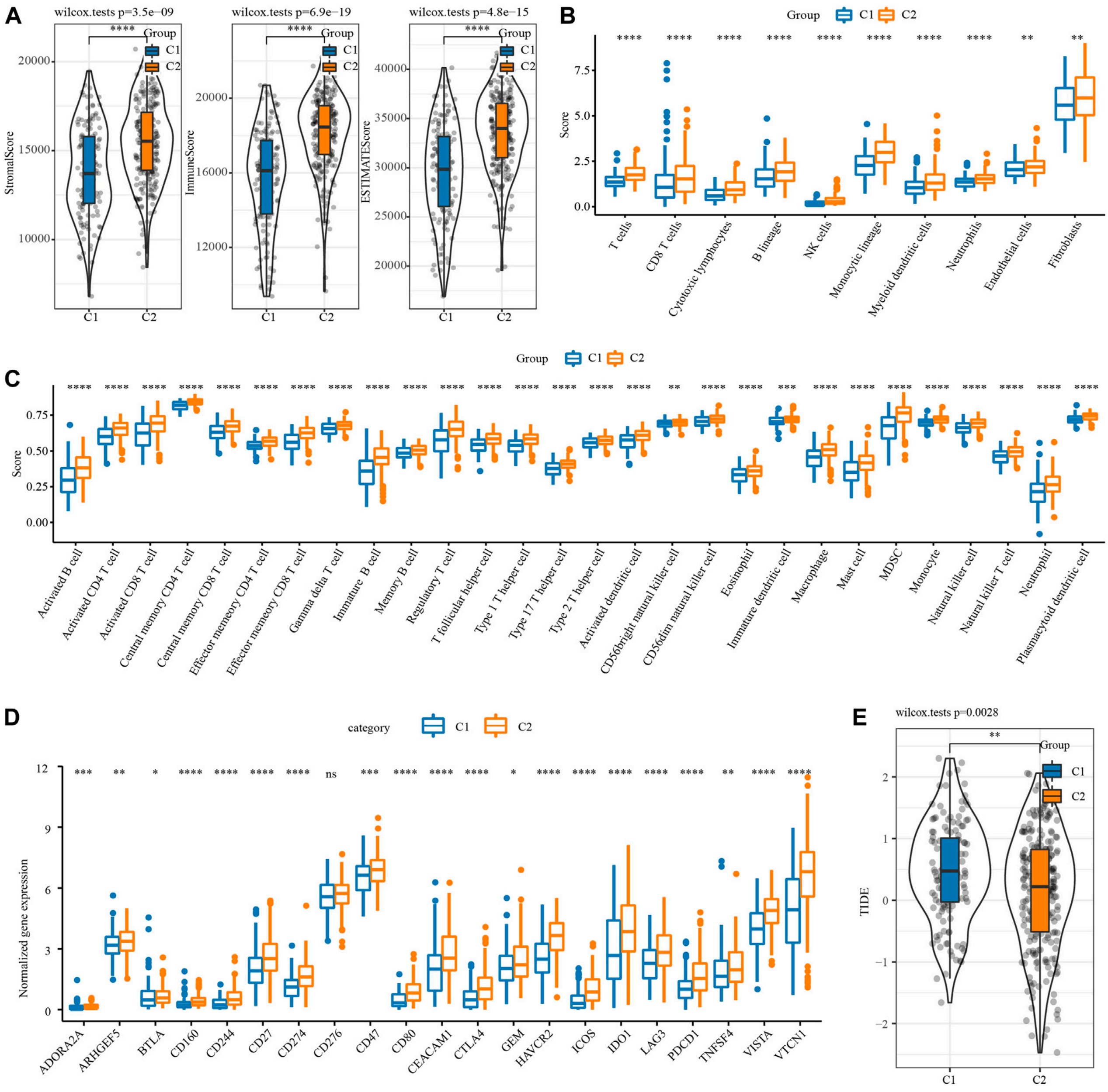


FIGURE 2

Immune microenvironment analysis and immunotherapy response prediction of necroptosis subtypes of OV. **(A)** The stromal score and immune score and tumor purity of two necroptosis subtypes in TCGA. **(B)** The infiltration scores of 10 immune cells calculated by MCP-Counter in the two necroptosis subtypes of OV. **(C)** The enrichment scores of 28 tumor-associated immune cells evaluated by ssGSEA in the two necroptosis subtypes of OV. **(D)** The box chart shows the association between the two necroptosis subtypes of OV and the expression of immune checkpoints. **(E)** TIDE score of two necroptosis subtypes of OV in TCGA. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Fisher test between the two necroptosis subtypes. The first 20 genes with the most significant difference in mutation rate between the two necroptosis subtypes were shown in Supplementary Figure S2A. The overall SNV rate in C2 was higher than that in C1, and the mutation rate of CSMD3 in C2 was the highest, followed by MST1R, PRKDC, PLEKHG1, and SMG1. However, the mutation rate of SPTAN1, SRCAP, FAT1, ROBO1, UBR5 in C1 was significantly higher than that in C2. Tumor mutation burden, homologous recombination defect, fraction altered and number of segments did not show significant differences between the two subgroups (Supplementary Figure S2B).

Necroptosis subtypes of OV showed different immune microenvironment and immunotherapeutic responses

The difference of immune microenvironment between the two necroptosis subtypes was first evaluated by ESTIMATE. C2 showed significantly higher stromal score, immune score and the ESTIMATE score represented the overall microenvironment score relative to C1 (Figure 2A). Then, MCP-counter and ssGSEA were used to analyze the infiltration level of immune cells in the immune microenvironment between the two necroptosis subtypes. The infiltration score of 10 immune cells calculated by MCP-Counter in C2 was significantly higher than that in C1 (Figure 2B). 28 tumor-associated immune cells assessed by ssGSEA showed active enrichment in C2, which was significantly higher compared with C1 (Figure 2C). The results of Figures 2A–C reflected the abundant infiltration of immune cells in C2, and its anti-tumor immune microenvironment might be more active. However, C2 patients had the worst prognosis, which was not consistent with the immune characteristics of this subtype. One possible reason is that the anti-tumor response of C2 was blocked by simultaneously highly expressed immune checkpoints. To verify this conjecture, the expression of 21 immune checkpoint molecules from HisgAtlas database (Liu et al., 2017) was examined. It was found that except for CD276, the expression of 20 immune checkpoints in C2 was significantly up-regulated, such as CD274, CTLA4, GEM, IDO1, LAG3G, PDCD1 and so on (Figure 2D). Considering that the two necroptosis subtypes had different levels of immune checkpoint expression, the response of different necroptosis subtypes to immune checkpoint inhibitor (ICI) was predicted by TIDE algorithm. The TIDE score of C2 was significantly lower in both necroptosis subtypes, suggesting that C2 was more likely to respond to ICB treatment than C1 (Figure 2E).

Identification of necroptosis subtype related gene modules

To construct a co-expression network, WGCNA was used to cluster 373 OV samples from TCGA datasets (Figure 3A). When the lowest soft threshold power was 9, scale-free R2 >0.9, guaranteed a scale-free network (Figure 3B). A clustering tree diagram reflecting the relationship between different modules and clinical features was constructed by using adjacency matrix, and 12 modules were determined (Figures 3C, D). By looking for the correlation between feature genes and external features, we found that the blue module had the strongest correlation with the two necroptosis subtypes, significantly negative correlation with C1 and significant positive correlation with C2 (Figure 3E). The link between each gene and C2 in the blue module was also very high (Figure 3F).

Construction of necroptosis-related risk score model based on hub gene in blue module

To screen the hub genes in the blue module, the genes in the module were analyzed by univariate Cox regression analysis, and 55 genes related to OV survival were obtained. Among them, the higher expression level of 42 was associated with the higher death risk, and the higher expression of 13 was associated with the lower death risk (Figure 4A). LASSO Cox regression penalized the unimportant features in the regularization process, 24 genes were obtained, which need to be further analyzed (Figure 4B). Multivariate Cox regression analysis selected 12 of these genes to calculate the NRRS of the sample (Figure 4C). Among the 12 genes, NACA2, DOCK11, EPB41L3, SCN1B, KRT18, THEMIS2, PLEKHF1 were associated with poor OS of OV, while HMGN3, WAR3, HLA_DOB, FBXO16, PLA2G2D were associated with better OS (Figure 4D). Risk groups were divided based on the median of the sample NRRS in each cohort. The survival analysis was carried out between the high-risk and low-risk packets in each cohort, and the performance of the NRRS model in each queue was evaluated by tdROC curve. Among the 373 samples of TCGA, the survival rate of the high-risk group was significantly lower than that of the low-risk group in the long term and short term. TdROC curve showed that the NRRS model had better long-term predictive ability in the TCGA-OV cohort because its AUC for predicting 5-year OS was 0.75, it was higher than the AUCs for predicting 1-year (0.69) and 3-year (0.73) OS (Figure 4E). The high-risk group of OV samples obtained from ICGC was also associated with a worse prognosis outcome, with AUC of 0.67, 0.71, and 0.7 for 1 -, 3 -, and 5-year OS, respectively (Figure 4F). In the GSE cohort that merged four GEO datasets, the prognosis of the low-risk group was significantly better than that of the high-risk group. The model predicted 1 year AUC = 0.63, 3 years AUC = 0.66, and 5 years AUC = 0.63 of OS (Figure 4G).

Single nucleotide variation and biological characteristics of necroptosis-related risk score model

We further explored the SNV and potential biological pathways related to NRRS. SNV existed in both high-risk and low-risk groups, and genetic mutations were more pronounced in the low-risk group than in the high-risk group, including but not limited to FLNB, UBR4, TRPS1, PCNT, SACS (Figure 5A). The TMB and homologous recombination defect characteristics of the low-risk group were significantly higher than those of the high-risk

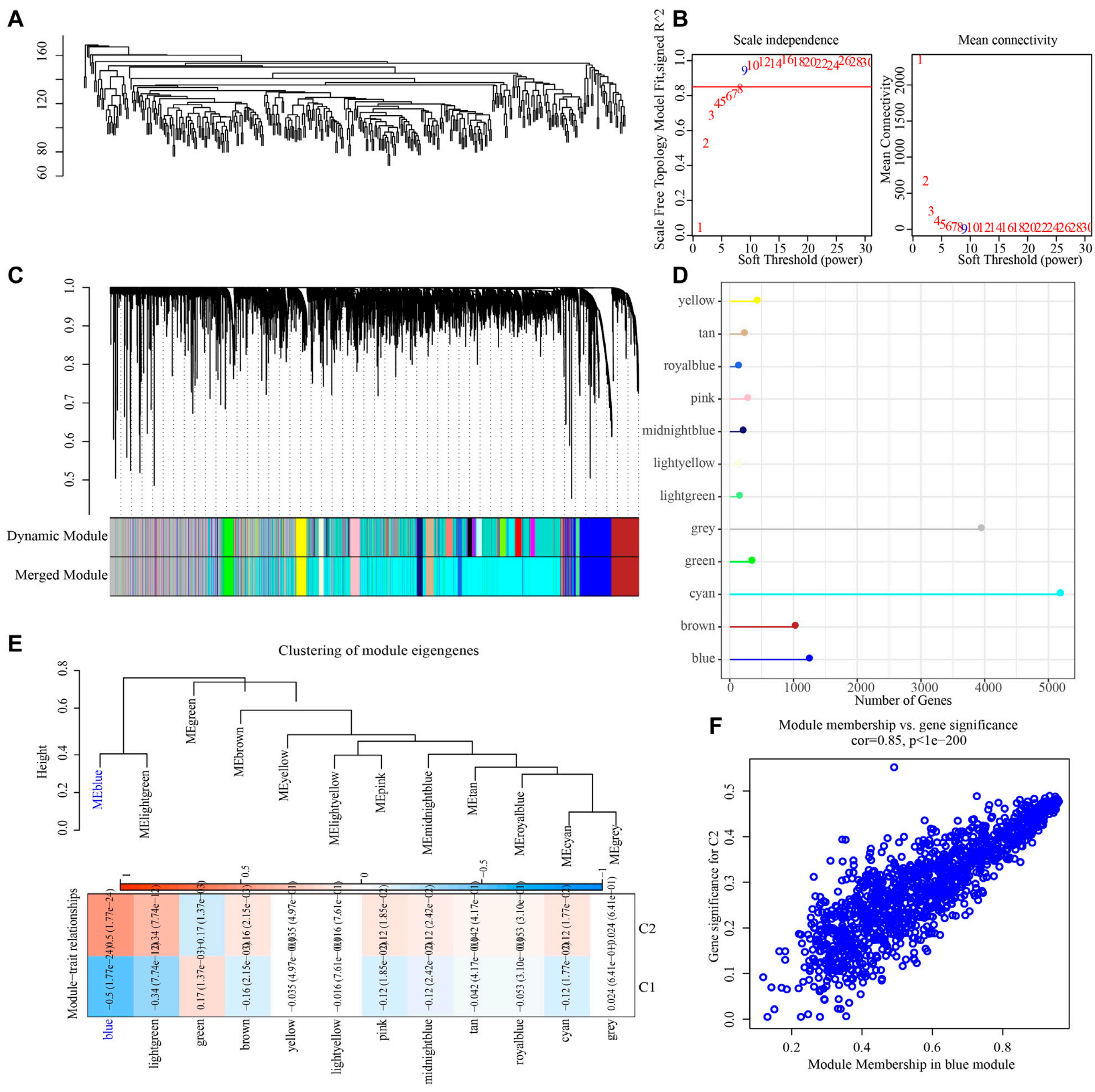


FIGURE 3
Identification of gene modules related to necroptosis subtypes. **(A)** The clustering tree of 373 OV samples in the TCGA dataset. **(B)** Analysis of scale-free exponent and average connectivity of various soft threshold powers. **(C)** Cluster dendrogram of the co-expression network modules. **(D)** The number of genes in each module. **(E)** The heatmap of the relationship between module eigengenes and necroptosis subtypes. **(F)** The association strength between gene significance (GS) and module membership (MM) for the C2 in the blue module.

group (Figure 5B). The correlation between NRRS and tumor biological pathway was analyzed, and the results were shown in Figure 5C. Specifically, epithelial-mesenchymal transition, angiogenesis, coagulation, TGF beta signaling, myogenesis, KRAS signal up, hypoxia and apical junction and apoptosis were all up-regulated in the high NRRS group of the three datasets (Figure 5D).

Construction of a decision tree model and a nomogram to improve the risk stratification of OS for OV patients

To make NRRS more suitable for predicting the prognosis of OV, a decision tree model was constructed using the clinical factors (age, stage, grade) of OV in TCGA and NRRS, and three

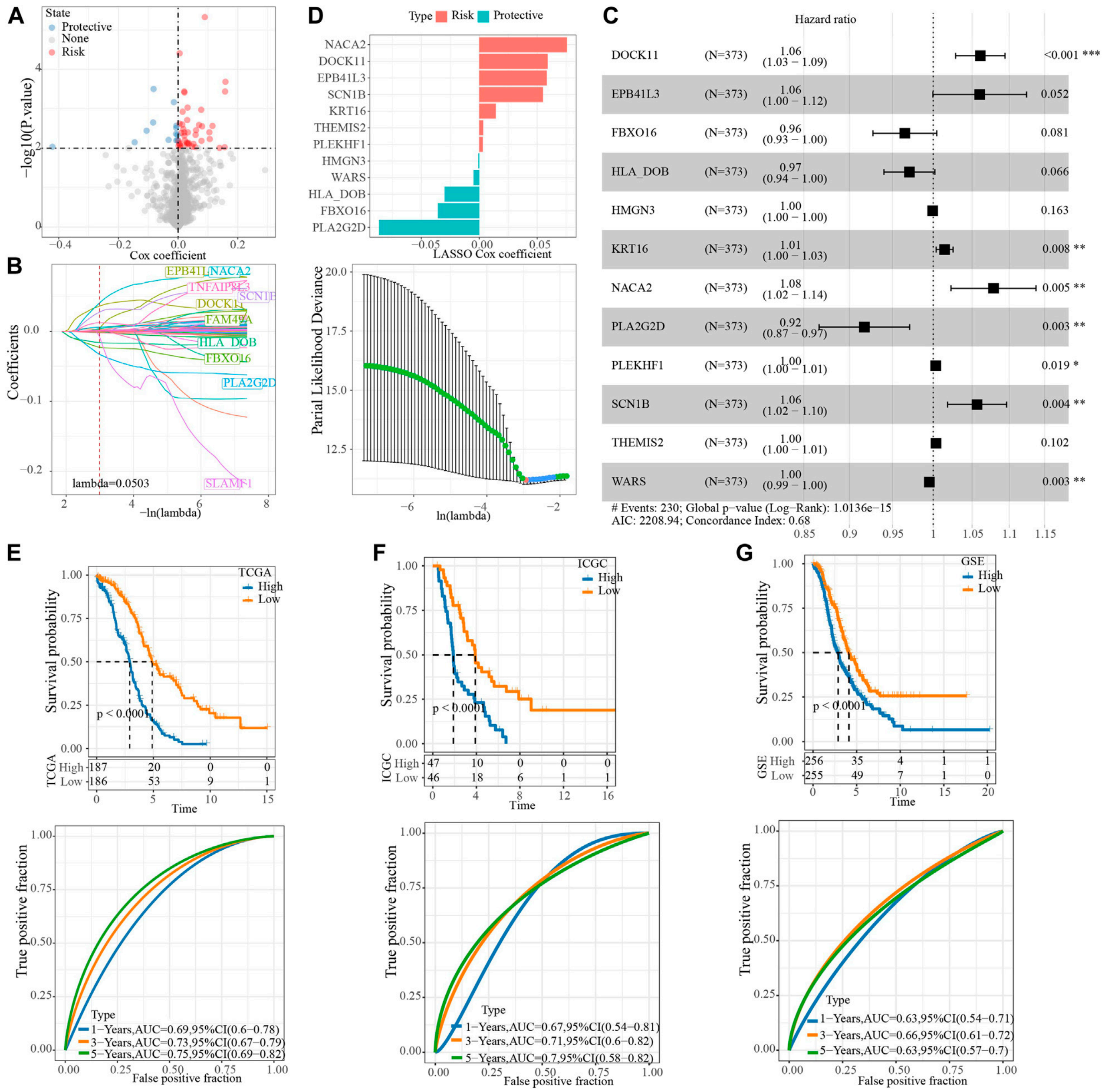


FIGURE 4
Construction of NRRS model based on hub gene in blue module. **(A)** The Cox coefficients of 55 genes related to OV survival. **(B)** LASSO Cox regression penalized the unimportant features in the regularization process. **(C)** The forest map shows the results of multivariate Cox regression analysis for 12 genes. **(D)** LASSO Cox coefficients of 12 genes. **(E)** Kaplan-Meier and tdROC curves of OS predicted by NRRS model in TCGA-OV cohort. **(F)** Kaplan-Meier and tdROC curves of OS predicted by NRRS model in ICGC dataset. **(G)** Kaplan-Meier and tdROC curves of OS predicted by NRRS model in GSE cohort.

different clusters:M1, M2 and M3 were established (Figure 6A). There were significant differences in OS among the three clusters (Figure 6B). All samples in M1 belonged to low NRRS group, while M2 and M3 belonged to high NRRS group (Figure 6C). The proportion of surviving patients in M1 was the highest among the three clusters, followed by M2, and finally M3 (Figure 6D). To construct a nomogram, univariate Cox regression analysis was carried out first, and the age and NRRS fits very well (Figure 6E). Multivariate Cox regression showed that NRRS was an independent prognostic variable for OV (Figure 6F). A nomogram was constructed according to age and NRRS (Figure 6G). The calibration curve showed that the prediction

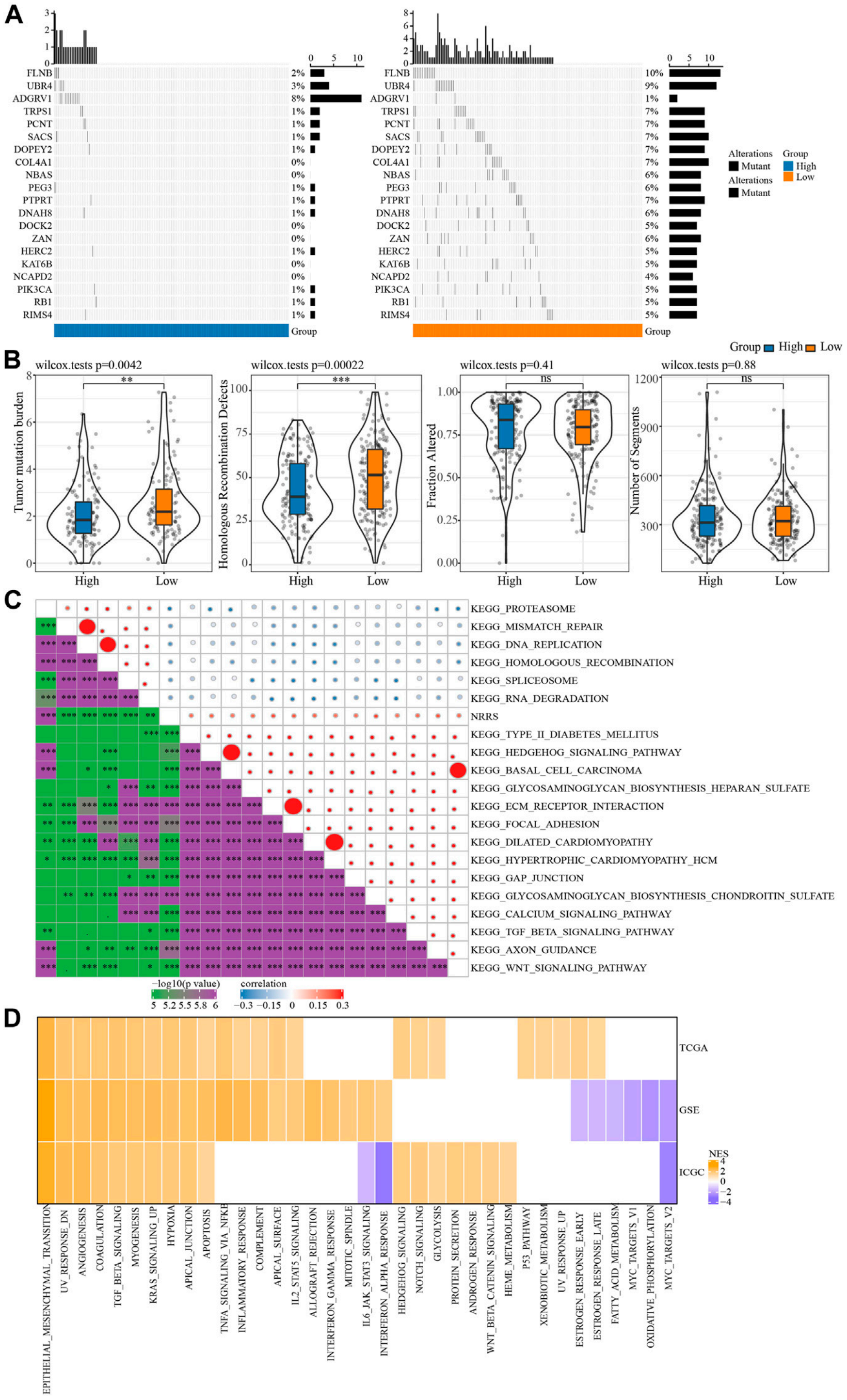


FIGURE 5
SNV and biological characteristics of NRRS model. **(A)** SNV in high-risk and low-risk groups. **(B)** TMB, homologous recombination defect, fraction altered, number of segments of high-risk group and low-risk group. **(C)** The correlation between NRRS and tumor biological pathway. **(D)** The high-risk group compared with the low-risk group in different pathways of NESs. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

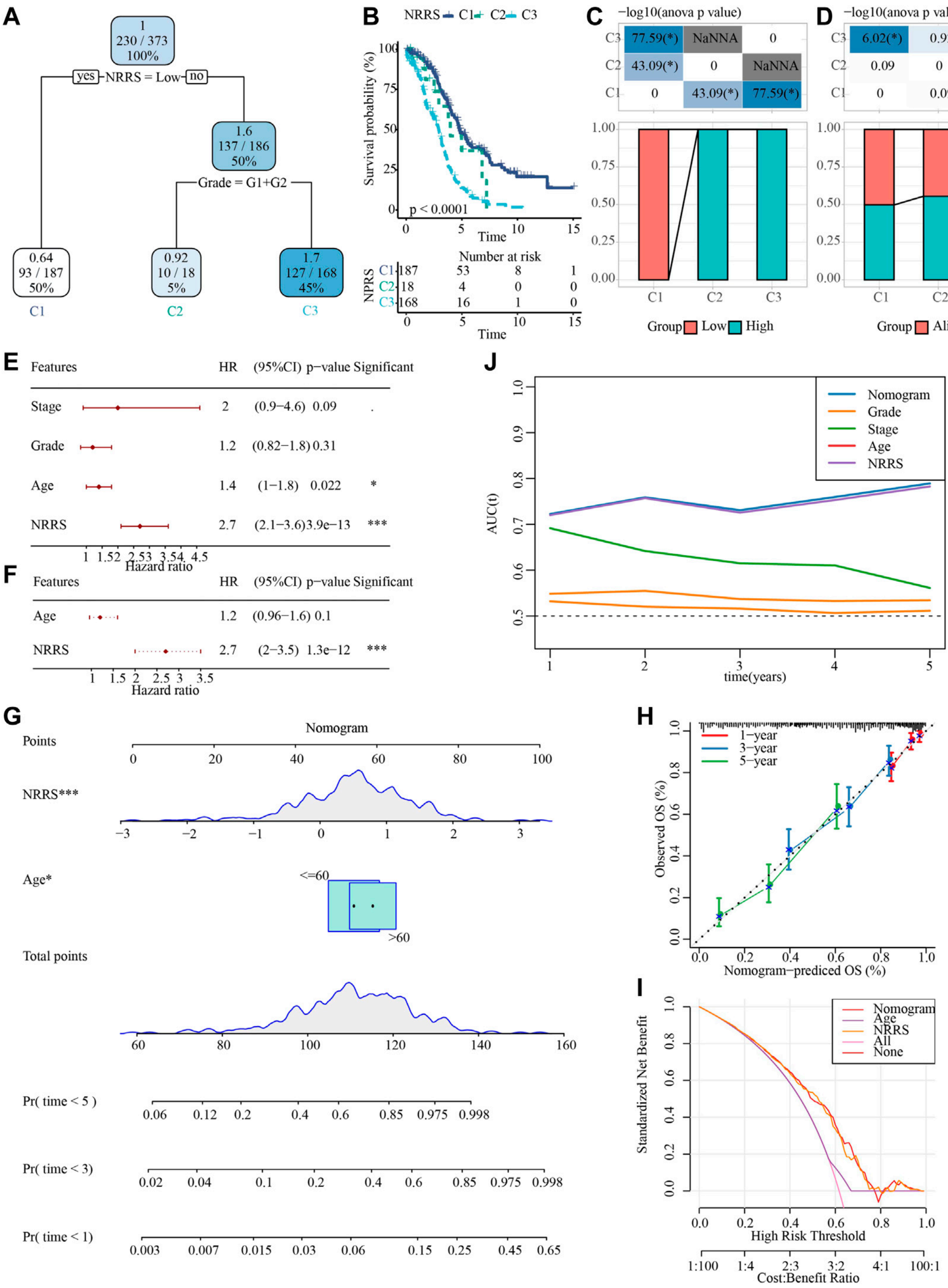


FIGURE 6
Construction of a decision tree model and a nomogram to improve the risk stratification of OS in OV patients. **(A)** Decision tree model based on NRRS and clinical factors (age, stage, grade). **(B)** Survival analysis of three risk subgroups of the decision tree. **(C)** The distribution of NRRS in three risk subgroups of the decision tree. **(D)** The survival status of patients in the three risk subgroups of the decision tree. **(E)** Univariate Cox regression analysis of NRRS and clinical factors of OV. **(F)** Multivariate Cox regression analysis of NRRS and age. **(G)** The nomogram constructed according to age and NRRS. **(H)** The calibration curve evaluates the proximity between the prediction line of the nomogram and the ideal 45-degree calibration line. **(I)** The decision curve shows the net income of NRRS and nomogram. **(J)** The tdROC curve displays the AUCs of NRRS and nomogram. $^{*}p < 0.05$, $^{***}p < 0.001$.

line of nomogram was close to the ideal 45° calibration line, indicating that nomogram had a certain degree of accuracy (Figure 6H). The decision curve showed that NRRS and nomogram have the highest net income (Figure 6I). And tdROC curve displayed that the AUC of NRRS and nomogram was very similar, both above 0.7 (Figure 6J).

Discussion

OV is a complex disease with multiple subtypes, each of which has different histopathology and different responses to treatment. Accurate classification and typing of OV can reliably predict disease progression and provide insight into the potentially targeted molecular mechanisms unique to each subtype (Cook and Vanderhyden, 2019). The study of Seehawer et al. provides a revolutionary insight that necroptotic microenvironment direct the lineage commitment of liver cancer and thus determine cancer subtypes (Seehawer et al., 2018). It is unclear whether necroptosis can affect the subtypes of other cancers. In recent years, several studies have focused on the effects of necroptosis-related genes on cancer typing, prognosis and biological effects (He et al., 2022; Nie et al., 2022; Xin et al., 2022). In this study, we systematically studied the effects of necroptosis on OV typing, prognosis, TMB, tumor microenvironment, immunotherapy response and biological pathway by bioinformatics analysis, which might be provide new molecular insights for necroptosis in OV.

First of all, we classified OV into two necroptosis subtypes according to 15 OV prognosis-related NRGs out of 74 NRGs. Compared with C1, C2 with a worse prognosis. The possible reason was that C2 showed a high immunosuppressive microenvironment. In OV, a large number of immunosuppressive cells, including tumor-associated macrophages, regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs) and Tumor associated dendritic cells, act as accomplices to coordinate highly complex immunosuppressive networks, inhibit anti-tumor immunity and help tumor cells escape immune attacks (Cai and Jin, 2017). Besides, in the early stage of tumor, the adaptive immune resistance mechanism may occur in patients with high CD8T cell density and predict a poor prognosis of the tumor (Peske et al., 2015). In the immunosuppressive environment, it is rich in immunosuppressive molecules such as IDO, PD-1, PD-L1, VISTA, LAG3, etc (Dempke et al., 2017; Drakes and Stiff, 2018). In C2, the levels of above immune cells and immunosuppressive molecules were higher than those in C1, which explained why the prognosis of C2 iwas worse than that of C1.

Then, the blue module most related to necroptosis subtype was identified by constructing a co-expression network, and the hub gene of the module was identified by LASSO Cox regression analysis. A NRRS model containing 12 genes (NACA2, DOCK11, EPB41L3, SCN1B, KRT18, THEMIS2, PLEKHF1, HMGN3, WAR3, HLA_DOB, FBXO16, PLA2G2D) was constructed. Among all 12 prognostic related genes, 8 genes (EPB41L3, SCN1B, KRT18, THEMIS2, PLEKHF1, HLA_DOB, FBXO16, PLA2G2D) (Dafou et al., 2010; Trisdale et al., 2016; Li et al., 2020; Brummelhuis et al., 2021; Ji et al., 2021; Zheng et al., 2021; Huang et al., 2022; Zhao et al., 2022) have been reported to be involved in tumorigenesis of OC or to be important predictors of overall survival. This implies that our bioinformatics analyses using cohorts have prognostic value. The remaining 4 genes have not previously been found to be associated with the prognosis of ovarian cancer and may serve as new potential biomarkers for the disease.

NRRS had many far-reaching clinical significances. First, it was related to the genomic stability of tumors. The NRRS low-risk group showed higher levels of SNV, TMB and homologous recombination defect. Second, it was related to the biological process of tumor. Specifically, compared with the NRRS low-risk group, epithelial-mesenchymal transition, angiogenesis, coagulation, TGF beta signaling, myogenesis, KRAS signal up, hypoxia and apical junction and apoptosis pathways were significantly up-regulated in the high-risk group. Third, NRRS was an independent prognostic variable of OV, and it was more accurate than other clinical parameters in predicting the prognosis of OV. And the decision tree and nomogram combined with NRRS and other clinical factors improved the risk stratification of OS in patients with OV.

However, some limitations of this study must be recognized. This study was purely from the bioinformatics analysis of the public database, the sample size of each cohort was relatively small, clinical information was prone to deviation, large-scale, multicenter, prospective studies are needed to further confirm our results. And the impact of the model needs biological experiments and clinical data to support it. In addition, the specific molecular mechanism of the model in OV remains to be further explored.

To sum up, OV was divided into two necroptosis subtypes in this study. There were significant differences in OS, immune cell infiltration, immune checkpoint expression and applicability to immunotherapy between patients with different subtypes. Moreover, a NRRS model was constructed to identify high-risk patients with OV, and combined with clinical factors to build a decision tree and nomogram to optimize the risk stratification of OS. Our research may provide molecular insights into the effects of necroptosis in cancer.

Data availability statement

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

Author contributions

All authors contributed to this present work: [CJ] designed the study and revised the manuscript, [YH] acquired the data. [YW] drafted the manuscript. All authors read and approved the 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1043870/full#supplementary-material

References

Achimas-Cadariu, P., Kubelac, P., Irimie, A., Berindan-Neagoe, I., and Ruhli, F. (2022). Evolutionary perspectives, heterogeneity and ovarian cancer: A complicated tale from past to present. *J. Ovarian Res.* 15 (1), 67. doi:10.1186/s13048-022-01004-1

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Beretta, G. L., and Zaffaroni, N. (2022). Necroptosis and prostate cancer: Molecular mechanisms and therapeutic potential. *Cells* 11 (7), 1221. doi:10.3390/cells11071221

Brummelhuis, I. S., Fiascone, S. J., Hasselblatt, K. T., Frendl, G., and Elias, K. M. (2021). Voltage-gated sodium channels as potential biomarkers and therapeutic targets for epithelial ovarian cancer. *Cancers (Basel)* 13 (21), 5437. doi:10.3390/cancers13215437

Cai, D. L., and Jin, L. P. (2017). Immune cell population in ovarian tumor microenvironment. *J. Cancer* 8 (15), 2915–2923. doi:10.7150/jca.20314

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, J., Wei, Z., Fu, K., Duan, Y., Zhang, M., Li, K., et al. (2022). Non-apoptotic cell death in ovarian cancer: Treatment, resistance and prognosis. *Biomed. Pharmacother.* 150, 112929. doi:10.1016/j.biopha.2022.112929

Cook, D. P., and Vanderhyden, B. C. (2019). Ovarian cancer and the evolution of subtype classifications using transcriptional profiling†. *Biol. Reprod.* 101 (3), 645–658. doi:10.1093/biolre/ioz099

Dafou, D., Grun, B., Sinclair, J., Lawrenson, K., Benjamin, E. C., Hogdall, E., et al. (2010). Microcell-mediated chromosome transfer identifies EPB41L3 as a functional suppressor of epithelial ovarian cancers. *Neoplasia (New York, NY)* 12 (7), 579–589. doi:10.1593/neo.10340

Dempke, W. C. M., Fenchel, K., Uciechowski, P., and Dale, S. P. (2017). Second- and third-generation drugs for immuno-oncology treatment-The more the better? *Eur. J. Cancer* 74, 55–72. doi:10.1016/j.ejca.2017.01.001

Drakes, M. L., and Stiff, P. J. (2018). Regulation of ovarian cancer prognosis by immune cells in the tumor microenvironment. *Cancers (Basel)* 10 (9), 302. doi:10.3390/cancers10090302

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

Guan, L. Y., and Lu, Y. (2018). New developments in molecular targeted therapy of ovarian cancer. *Discov. Med.* 26 (144), 219–229.

He, R., Zhang, M., He, L., Huang, J., Man, C., Wang, X., et al. (2022). Integrated analysis of necroptosis-related genes for prognosis, immune microenvironment infiltration, and drug sensitivity in colon cancer. *Front. Med.* 9, 845271. doi:10.3389/fmed.2022.845271

Hou, J., Zhang, Y., Zhu, Y., Zhou, B., Ren, C., Liang, S., et al. (2019). α-Pinene induces apoptotic cell death via caspase activation in human ovarian cancer cells. *Med. Sci. Monit.* 25, 6631–6638. doi:10.12659/MSM.916419

Huang, W. C., Yen, J. H., Sung, Y. W., Tung, S. L., Chen, P. M., Chu, P. Y., et al. (2022). Novel function of THEMIS2 in the enhancement of cancer stemness and chemoresistance by releasing PTP1B from MET. *Oncogene* 41 (7), 997–1010. doi:10.1038/s41388-021-02136-2

Ji, M., Zhao, Z., Li, Y., Xu, P., Shi, J., Li, Z., et al. (2021). FBXO16-mediated hnRNPL ubiquitination and degradation plays a tumor suppressor role in ovarian cancer. *Cell Death Dis.* 12 (8), 758. doi:10.1038/s41419-021-04040-9

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

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

Li, Y., Li, H., Yang, B., Wei, J., Zhen, C., and Feng, L. (2020). Clinical significance of PI3 and HLA-DOB as potential prognostic predicators for ovarian cancer. *Transl. Cancer Res.* 9 (2), 466–476. doi:10.21037/tcr.2019.11.30

Liu, J., Hong, M., Li, Y., Chen, D., Wu, Y., and Hu, Y. (2022). Programmed cell death tunes tumor immunity. *Front. Immunol.* 13, 847345. doi:10.3389/fimmu.2022.847345

Liu, Y., He, M., Wang, D., Diao, L., Liu, J., Tang, L., et al. (2017). HisgAtlas 1.0: A human immunosuppression gene database. *Database.* 2017, bax094. doi:10.1093/database/bax094

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

Nie, S., Huili, Y., He, Y., Hu, J., Kang, S., and Cao, F. (2022). Identification of bladder cancer subtypes based on necroptosis-related genes, construction of a prognostic model. *Front. Surg.* 9, 860857. doi:10.3389/fsurg.2022.860857

Pasparakis, M., and Vandenabeele, P. (2015). Necroptosis and its role in inflammation. *Nature* 517 (7534), 311–320. doi:10.1038/nature14191

Peske, J. D., Woods, A. B., and Engelhard, V. H. (2015). Control of CD8 T-cell infiltration into tumors by vasculature and microenvironment. *Adv. Cancer Res.* 128, 263–307. doi:10.1016/bs.acr.2015.05.001

Qin, X., Ma, D., Tan, Y. X., Wang, H. Y., and Cai, Z. (2019). The role of necroptosis in cancer: A double-edged sword? *Biochim. Biophys. Acta. Rev. Cancer* 1871 (2), 259–266. doi:10.1016/j.bbcan.2019.01.006

Rakina, M., Kazakova, A., Villert, A., Kolomiets, L., and Larionova, I. (2022). Spheroid Formation and peritoneal metastasis in ovarian cancer: The role of stromal and immune components. *Int. J. Mol. Sci.* 23 (11), 6215. doi:10.3390/ijms23116215

Revythis, A., Limbu, A., Mikropoulos, C., Ghose, A., Sanchez, E., Sheriff, M., et al. (2022). Recent insights into PARP and immuno-checkpoint inhibitors in epithelial ovarian cancer. *Int. J. Environ. Res. Public Health* 19 (14), 8577. doi:10.3390/ijerph19148577

Seehawer, M., Heinzmann, F., D'Artista, L., Harbig, J., Roux, P. F., Hoenicke, L., et al. (2018). Necroptosis microenvironment directs lineage commitment in liver cancer. *Nature* 562 (7725), 69–75. doi:10.1038/s41586-018-0519-y

Seifert, L., Werba, G., Tiwari, S., Giao Ly, N. N., Alothman, S., Alqunaibit, D., et al. (2016). The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression. *Nature* 532 (7598), 245–249. doi:10.1038/nature17403

Strilic, B., Yang, L., Albarrán-Juárez, J., Wachsmuth, L., Han, K., Müller, U. C., et al. (2016). Tumour-cell-induced endothelial cell necroptosis via death receptor 6 promotes metastasis. *Nature* 536 (7615), 215–218. doi:10.1038/nature19076

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

Trisdale, S. K., Schwab, N. M., Hou, X., Davis, J. S., and Townson, D. H. (2016). Molecular manipulation of keratin 8/18 intermediate filaments: Modulators of FAS-mediated death signaling in human ovarian granulosa tumor cells. *J. Ovarian Res.* 9, 8. doi:10.1186/s13048-016-0217-z

van Stein, R. M., Aalbers, A. G. J., Sonke, G. S., and van Driel, W. J. (2021). Hyperthermic intraperitoneal chemotherapy for ovarian and colorectal cancer: A review. *JAMA Oncol.* 7 (8), 1231–1238. doi:10.1001/jamaoncol.2021.0580

Whelan, E., Kalliala, I., Semertzidou, A., Raglan, O., Bowden, S., Kechagias, K., et al. (2022). Risk factors for ovarian cancer: An umbrella review of the literature. *Cancers* 14 (11), 2708. doi:10.3390/cancers14112708

Wood, G. E., and Ledermann, J. A. (2022). Adjuvant and post-surgical treatment in high-grade epithelial ovarian cancer. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 78, 64–73. doi:10.1016/j.bpobgyn.2021.09.002

Wu, J., Ye, J., Xie, Q., Liu, B., and Liu, M. (2022). Targeting regulated cell death with pharmacological small molecules: An update on autophagy-dependent cell death, ferroptosis, and necroptosis in cancer. *J. Med. Chem.* 65 (4), 2989–3001. doi:10.1021/acs.jmedchem.1c01572

Xin, S., Mao, J., Duan, C., Wang, J., Lu, Y., Yang, J., et al. (2022). Identification and quantification of necroptosis landscape on therapy and prognosis in kidney renal clear cell carcinoma. *Front. Genet.* 13, 832046. doi:10.3389/fgene.2022.832046

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

Zhang, M., Chen, Z., Wang, Y., Zhao, H., and Du, Y. (2022). The role of cancer-associated fibroblasts in ovarian cancer. *Cancers (Basel)* 14 (11), 2637. doi:10.3390/cancers14112637

Zhao, N., Xing, Y., Hu, Y., and Chang, H. (2022). Exploration of the immunotyping landscape and immune infiltration-related prognostic markers in ovarian cancer patients. *Front. Oncol.* 12, 916251. doi:10.3389/fonc.2022.916251

Zheng, M., Long, J., Chelariu-Raicu, A., Mullikin, H., Vilsmaier, T., Vattai, A., et al. (2021). Identification of a novel tumor microenvironment prognostic signature for advanced-stage serous ovarian cancer. *Cancers (Basel)* 13 (13), 3343. doi:10.3390/cancers13133343

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fgene.2023.1100020

OPEN ACCESS

EDITED BY
Xiang Xue,
University of New Mexico, United States

REVIEWED BY
Mingxin Yu,
China Medical University, China
Tiansheng Cao,
Southern Medical University, China

*CORRESPONDENCE
Min Zhu,
✉ zhumin@fybjsd.org.cn

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

RECEIVED 16 November 2022
ACCEPTED 13 March 2023
PUBLISHED 23 March 2023

CITATION
Lan Y, Jia Q, Feng M, Zhao P and Zhu M (2023), A novel natural killer cell-related signatures to predict prognosis and chemotherapy response of pancreatic cancer patients.
Front. Genet. 14:1100020.
doi: 10.3389/fgene.2023.1100020

A novel natural killer cell-related signatures to predict prognosis and chemotherapy response of pancreatic cancer patients

Yongting Lan[1], Qing Jia[1], Min Feng[1], Peiqing Zhao[1] and Min Zhu[2]*

[1]Department of Gastroenterology, Zibo Central Hospital, Zibo, China, [2]Department of Neonatology, Zibo Maternal and Child Health Hospital, Zibo, China

Background: Natural killer (NK) cells are involved in monitoring and eliminating cancers. The purpose of this study was to develop a NK cell-related genes (NKGs) in pancreatic cancer (PC) and establish a novel prognostic signature for PC patients.

Methods: Omic data were downloaded from The Cancer Genome Atlas Program (TCGA), Gene Expression Omnibus (GEO), International Cancer Genome Consortium (ICGC), and used to generate NKG-based molecular subtypes and construct a prognostic signature of PC. NKGs were downloaded from the ImmPort database. The differences in prognosis, immunotherapy response, and drug sensitivity among subtypes were compared. 12 programmed cell death (PCD) patterns were acquired from previous study. A decision tree and nomogram model were constructed for the prognostic prediction of PC.

Results: Thirty-two prognostic NKGs were identified in PC patients, and were used to generate three clusters with distinct characteristics. PCD patterns were more likely to occur at C1 or C3. Four prognostic DEGs, including MET, EMP1, MYEOV, and NGFR, were found among the clusters and applied to construct a risk signature in TCGA dataset, which was successfully validated in PACA-CA and GSE57495 cohorts. The four gene expressions were negatively correlated with methylation level. PC patients were divided into high and low risk groups, which exerts significantly different prognosis, clinicopathological features, immune infiltration, immunotherapy response and drug sensitivity. Age, N stage, and the risk signature were identified as independent factors of PC prognosis. Low group was more easily to happened on PCD. A decision tree and nomogram model were successfully built for the prognosis prediction of PC patients. ROC curves and DCA curves demonstrated the favorable and robust predictive capability of the nomogram model.

Abbreviation: NK, Natural killer; NKGs, NK cell-related genes; PC, pancreatic cancer; TME, tumor microenvironment; TCGA, The Cancer Genome Atlas Program; GEO, Gene Expression Omnibus; PAM, partition around medoids; DEGs, differently expressed genes; FDR, false discovery rate; FC, fold change; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; MSigDB, Molecular Signatures Database; IC_{50}, half-maximal inhibitory concentration; DCA, decision curve analysis; TMB, tumor mutation burden; ICG, immune checkpoint genes; TNFSF10, tumor necrosis factor ligand superfamily member 10; TNF, tumor necrosis factor; PIK3CB, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform; HCK, hematopoietic cell kinase; LCK, src-family kinases p56; TCR, T-cell antigen receptor; GEP, gene expression profile; CYT, cytolytic activity score.

Conclusion: We characterized NKGs-derived molecular subtypes of PC patients, and established favorable prognostic models for the prediction of PC prognosis, which may serve as a potential tool for prognosis prediction and making personalized treatment in PC.

KEYWORDS

natural killer cells, pancreatic cancer, consensus clustering, nomogram, methylation, programmed cell death, prognosis

1 Introduction

Pancreatic cancer (PC) as a lethal malignancy shows a high mortality worldwide, causing over 331000 deaths per year globally (Rawla et al., 2019). Although advances in the treatment of PC, patients who received surgical resection have a five-year survival rate ranging from 10% to 25% (Siegel et al., 2020). PC was usually diagnosed at a late stage due to the impalpable symptoms at the early stage, and approximately 80%–85% of PC was unresectable or metastatic at the time of diagnosis (Okasha et al., 2017). Currently, chemotherapy is the main treatment for PC but remains an unsatisfactory prognosis, and more effective and precise therapies are required (Mizrahi et al., 2020).

Immunotherapy has been recently developed to help improve the prognosis of various cancer types, such as renal cell carcinoma (Cho et al., 2017), non-small cell lung cancer (Hellmann et al., 2018), hematologic malignancies (Nelson and Paulos, 2015), and melanoma (Ribas and Wolchok, 2018). The principle of tumor immunotherapy is to fight against tumors through the activation of immune system, during which restarting and maintaining tumor-immune cycle plays a crucial role. Therapeutic targeting of immune checkpoints with immune checkpoint inhibitors has revolutionized cancer treatment (Komatsubara and Carvajal, 2017; Pulluri et al., 2017; Considine and Hurwitz, 2019). It was reported that checkpoint blockade in combination with GVAX has the potential for clinical benefit for patients with PC (Le et al., 2013). T-cell immunity is associated with the exceptional outcome of the few long-term survivors. A study identified unique neoantigens as T-cell targets in PC patients, which might be used to guide the application of immunotherapies (Balachandran et al., 2017). Pembrolizumab is a PD-1 inhibitor and has been approved for tumor patients with deficient mismatch repair or high microsatellite instability, including PC (Boyiadzis et al., 2018). However, the efficacy was restricted to a rare population due to the complex, highly immunosuppressive tumor microenvironment of PC (O'Reilly et al., 2019).

The tumor immune microenvironment (TME) contains tumor cells, immune cells, cytokines, etc., and its heterogeneity can potentially impact the patient's response to immunotherapy. Natural killer (NK) cells are a subset of innate immune cells and play a crucial role as effector cells against tumors. NK cell can directly kill malignant even at a relatively low ratio in the early presence of tumors (Huntington et al., 2007) or promotes adaptive T-cell immunological responses to limit cancer cell aggressiveness (López-Soto et al., 2017). The activation of NK cells is controlled by the integration of signals from cytokine receptors and a range of germline-encoded inhibitory and activating receptors (Moretta et al., 2006; Lanier, 2008). Studies found that NK cell activity was significantly negatively correlated with the risk of malignancy (Imai et al., 2000), and patients with a higher NK cell infiltration into cancers had better outcomes (Coca et al., 1997; Ishigami et al., 2000; Cursons et al., 2019). Cutting-edge immunotherapy targeting NK cells exerts great potential in the treatment of cancer and become an attractive alternative to T cell immunotherapies (Guillerey et al., 2016; Souza-Fonseca-Guimaraes et al., 2019). Accumulating evidence described the molecular characteristics of NK cells in various cancers (Sun et al., 2021a; Sun et al., 2021b), but a comprehensive molecular characterization of NK cells in PC remains poorly understood.

In the present study, the PC patients were clustered on the basis of natural killer cell-related genes (NKGs), and further comparison of the clinicopathological, mutational, immunological and pathway characteristics among subtypes was conducted. In addition, we identified prognostic differentially expressed genes (DEGs) among subgroups and constructed a risk signature for prognosis prediction. The decision tree and nomogram model were built using clinicopathological features and the risk signature to assist in prognostic prediction and personalized treatment of patients with PC.

2 Materials and methods

2.1 Data collection and preprocessing

Transcriptome files and clinicopathological data of patients with PC were obtained from the Cancer Genome Atlas Program (TCGA) (https://tcga-data.nci.nih.gov/tcga/), Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/), and the International Cancer Genome Consortium (ICGC) (https://www.icgc.org) databases. After removal of patients without complete clinical information and outcome status, as well as follow-up of fewer than 30 days, 176 PC patients from the TCGA pancreatic adenocarcinoma (TCGA-PAAD) cohort were retained as a training set. Ensembl was converted into gene symbol, and median value was kept when a genes had multiple gene symbols. The validation set contains 63 samples from the GSE57495 cohort and 215 patients of the PACA-CA cohort from the ICGC database. When multiple gene symbols appear or multiple probes appear for a gene, the median is taken as the gene expression value. A total of 134 human NKGs were downloaded from the ImmPort (https://www.immport.org/resource) database.

2.2 Consensus clustering

The prognostic NKGs were identified *via* univariate Cox regression analysis and were used to perform consensus clustering of PC patients. Consensus clustering analysis was conducted using the "ConsensusClusterPlus" R package to determine subgroups of PC patients based on the prognostic NKGs (Wilkerson and Hayes, 2010). The best classification was determined using the partition around medoids (PAM) algorithm and 1-Pearson correlation distance, with 500 bootstraps.

2.3 Risk score

The DEGs among NKGs-derived clusters were screened out using "limma" package according to the false discovery rate (FDR) < 0.05 and |log2 [fold change (FC)]| > log2 (2) (Ritchie et al., 2015). The univariate and the least absolute shrinkage and selection operator (LASSO) Cox regression analysis were adopted to identify and filter prognosis-related NKGs, respectively. Finally, by choosing the optimal penalty parameter lambda correlated with the minimum 10-fold cross-validation, multivariate Cox regression analysis was then implemented to establish the prognostic signature. The formula for the risk signature was as follows: risk score = $\sum \beta i \times Expi$. Where the βi represents the coefficient and Expi represents the normalized expression level of a gene. Two risk groups (high and low) were generated by a threshold of zero, and K–M analysis was conducted to compare overall survival (OS) differences between the high- and low-risk groups. The receiver operating characteristic (ROC) analysis was performed to estimate the predictive accuracy of the risk score.

2.4 Gene set enrichment analysis

GSEA was performed to analyze the differences in specific gene sets using the "GSVA" R package (Hänzelmann et al., 2013). The hallmark gene sets from the Molecular Signatures Database (MSigDB), the inflammation-related gene sets (Liu et al., 2020), and the angiogenesis-related gene set (Masiero et al., 2013) were used to be analyzed. These pathways with the FDR <0.05 was considered to be significant. Functional enrichment analysis included Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) (biological process (BP), cellular component (CC), and molecular function (MF)) analysis was performed on DEGs in clusters using WebGestaltR package (Yu et al., 2012).

2.5 Immune infiltration, chemotherapeutic sensitivity, and immunotherapy response predictions

The relative proportion of immune cells was calculated using the CIBERSORT algorithm (https://cibersort.stanford.edu/), which performs cell type enrichment analysis from gene expression data for 22 immune cells. The "ESTIMATE" R package was applied to estimate and extrapolate the fraction of stromal and immune cells in tumor samples (Yoshihara et al., 2013). The expression levels of the immune checkpoints were compared in different groups. To predict the chemosensitivity of osteosarcoma patients to several common anti-cancer drugs (methotrexate, paclitaxel, cisplatin, and doxorubicin), we adopted the "pRRophetic" R package to infer the half-maximal inhibitory concentration (IC_{50}) values by constructing the ridge regression model based on Genomics of Drug Sensitivity in Cancer (GDSC) (www.cancerrxgene.org/) cell line expression spectrum and gene expression profiles (Geeleher et al., 2014).

2.6 Establishment of a predictive nomogram

The decision tree model was applied to classify subgroups based on clinicopathogicial features and risk scores by using the "rpart" R package (https://cran.r-project.org/web/packages/rpart/index.html). The independent prognostic factors of OS for PC were identified by univariate and multivariate Cox regression analysis. A nomogram integrating the risk signature and independent prognostic clinicopathological factors was constructed in the TCGA cohort by the "rms" R package (https://cran.r-project.org/web/packages/rms/index.html). The calibration curves were utilized to evaluate the prediction accuracy between the predicted 1-, 2- and 3-year OS probabilities and the actual observations. The discriminate ability of the nomogram was assessed by time-dependent ROC curves. The decision curve analysis (DCA) was conducted to test the clinical utility of the nomogram using the "rmda" R package (https://cran.r-project.org/web/packages/rmda/index.html).

2.7 Mutation analysis

Tumor mutation burden (TMB) is was determined as the number of somatic indels and base substitutions per million bases in the coding region of the genome detected. Gene mutation data of PC patients were downloaded from the TCGA database and TMB was calculated using the "maftools" package (Mayakonda et al., 2018) as previously described (Chalmers et al., 2017).

2.8 Programmed cell death (PCD) analysis

12 PCD patterns were acquired from previous (Zou et al., 2022). Altogether, 580 apoptosis genes, 52 pyroptosis genes, 87 ferroptosis genes, 367 autophagy genes, 14 cuproptosis genes, 9 parthanatos genes, 15 entotic cell death genes, 101 necroptosis genes, 8 netotic cell death genes, 7 alkaliptosis genes, 220 lysosome-dependent cell death genes, and 5 oxeiptosis genes were collected. Based on the expression data of above gene sets, ssGSEA analysis was conducted on tumor samples using the R package GSVA.

2.9 Statistical analysis

The R software (v3.6.3) was used for statistical analyses. Wilcoxon test compared differences between two groups. Survival differences were compared using K–M curves with a Log-rank test. The Cox proportional hazard model was performed to estimate the β regression coefficient, hazard

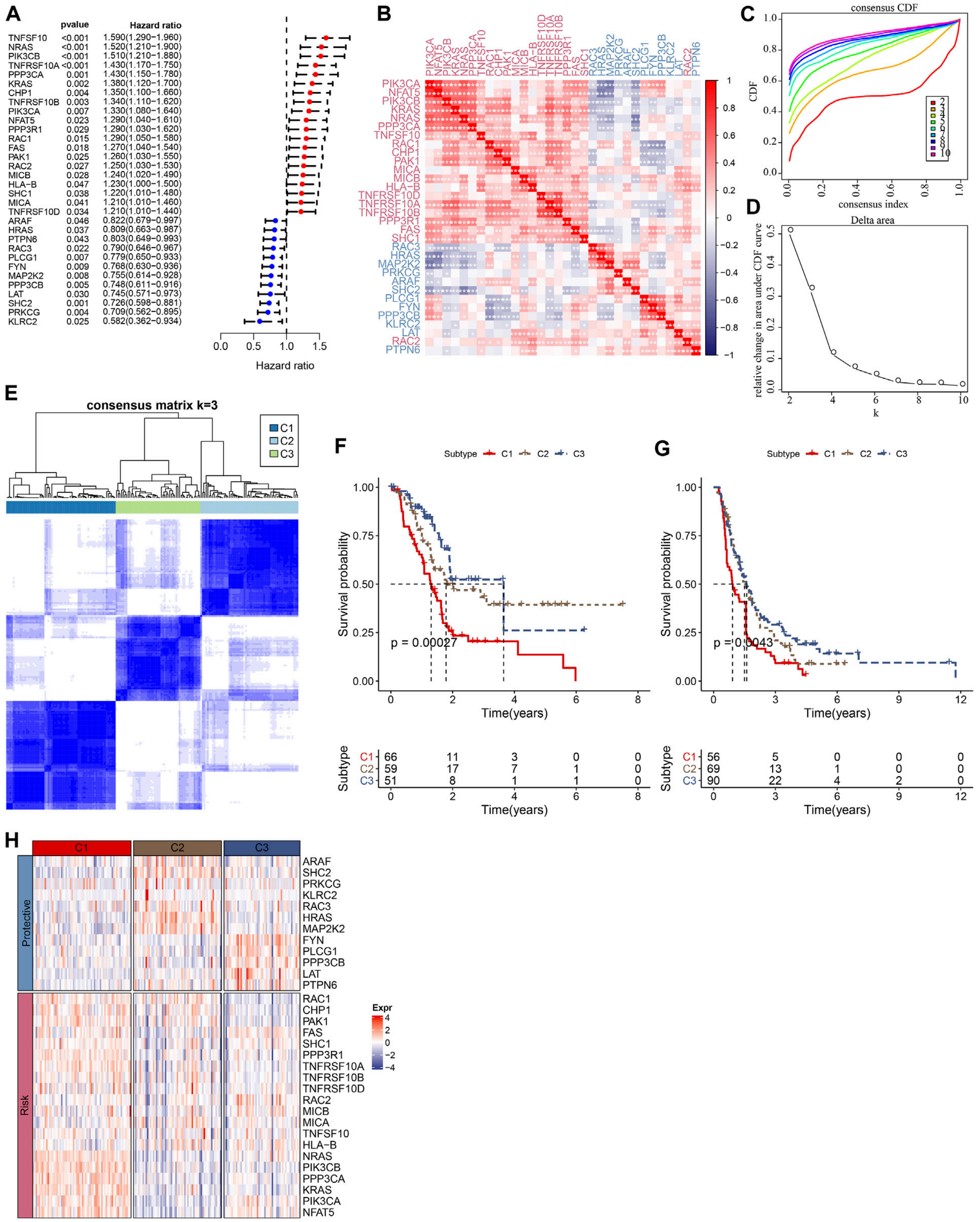


FIGURE 1

Consensus clustering of PC patients based on NKG signature. **(A)** Forest plot of prognosis-related NKGs in the TCGA-PAAD cohort; **(B)** The correlations among 32 prognosis-related NKGs in the TCGA-PAAD cohort; **(C)** Consensus cumulative distribution function (CDF) diagram when different k values, **(D)** Delta area plot for the relative change in the area under CDF curve for k compared to k-1, **(E)** Consensus matrix when the number of groups (k) = 3. In the consensus matrix, white meant that samples were impossibly clustered together, and dark blue meant that samples were always clustered together. Both rows and columns of the matrix represented samples, **(F)** and **(G)** represented the survival analysis of the clusters in the TCGA-PAAD cohort and PACA-CA cohort, respectively. **(H)** The heatmap of expression of 32 prognosis-related NKGs in the TCGA-PAAD cohort.

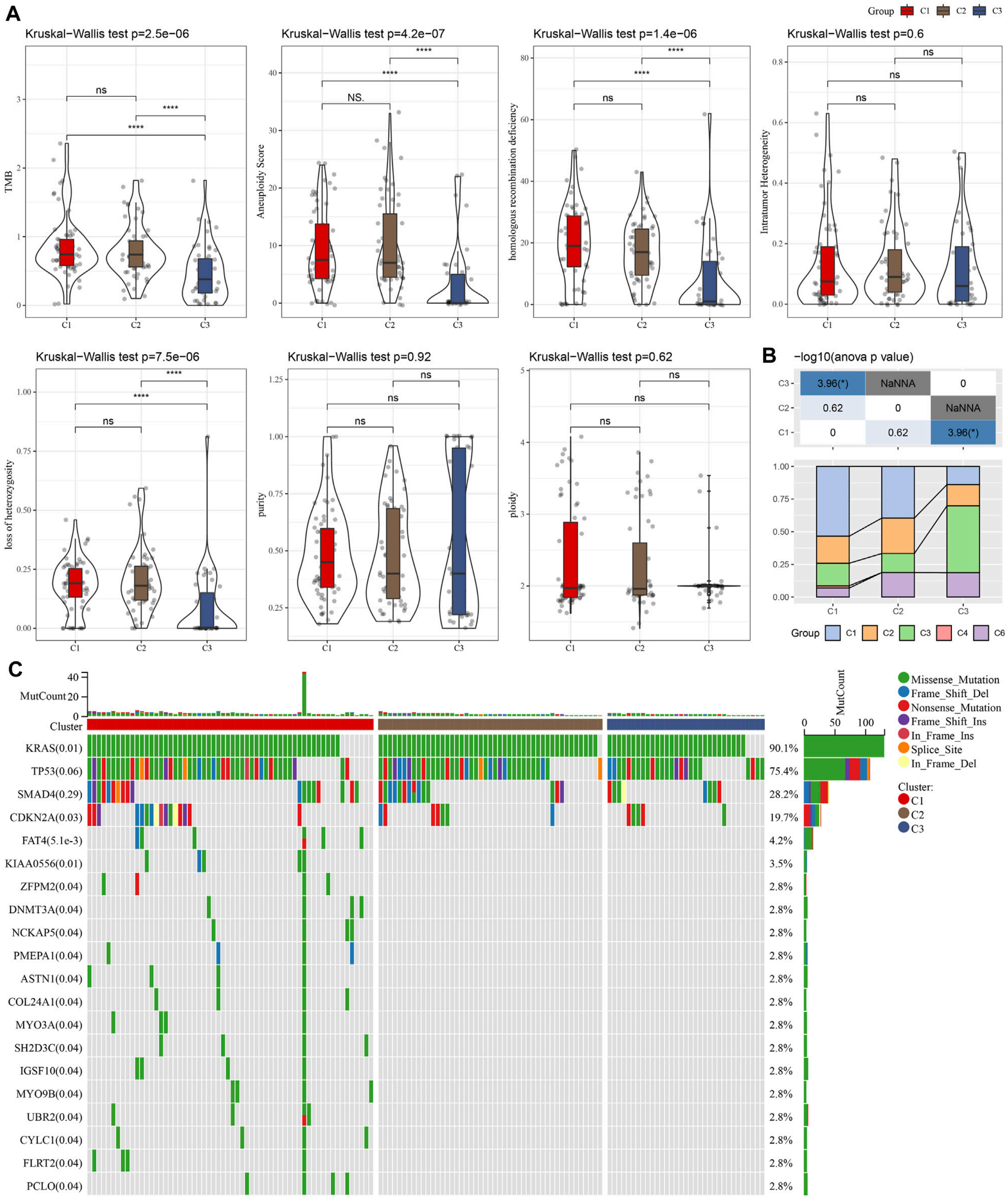


FIGURE 2
The comparison between our defined cluster with other existing subtypes. **(A)** The comparison of our defined clusters with the subtypes derived from the pan-cancer study. **(B)** The comparison of our defined clusters with immune signature-derived subtypes. **(C)** The comparison of somatic mutations in NKGs-derived subtypes in the TCGA cohort. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; and ****$p < 0.0001$.

ratios, *p*-value, and their corresponding 95% confidence interval for each of the selected risk predictors. a nomogram was constructed with the "rms" package in R. The C-index and calibration curve with the bootstrap method were used to evaluate the prediction performance of the nomogram. A *p*-value <0.05 was deemed to be a statistical significance.

3 Results

3.1 Molecular subtypes derived from natural killer cell-related genes

The flowchart is shown in Supplementary Figure S1. To obtain molecular subtypes of PC based on NKG, we first identified 32 NKGs that were significantly associated with the prognosis of PC ($p < 0.05$, Figure 1A). Notably, positive correlations among the expression of the 32 NKGs were observed in Figure 1B. Subsequently, consensus clustering of the 32 NKGs generated three stable clusters (C1, C2, and C3) in the TCGA-PAAD cohort (Figures 1C–E). Survival analysis demonstrated that the C3 cluster had a favorable prognosis whereas the C1 cluster had a poorer prognosis (Figure 1F). The individuals in the PACA-CA cohort were also divided into three clusters, which exerted similar prognosis characteristics as the clusters in the TCGA-PAAD cohort (Figure 1G). Among the 32 NKGs, the risk genes were generally overexpressed in the C1 cluster, and the protective genes were mainly elevated in the C3 clusters (Figure 1H).

3.2 Genomic landscapes among molecular subtypes

We compared defined three clusters with the molecular subtypes derived from a pan-cancer study and immune signatures (Thorsson et al., 2018). As shown in Figure 2A, the C1 cluster presented with a higher TMB, aneuploidy, homologous recombination defects, and loss of heterozygosity. Meanwhile, a significantly higher proportion of immune signature-derived C3 subtype in our defined C3 subtype was observed (Figure 2B). The immune signature-derived C3 subtype was characterized by the overexpression of TH17 and Th1 genes, a low to moderate proliferation rate of tumor cells, and lower levels of aneuploidy and overall somatic copy number alterations. Meanwhile, the immune signature-derived C3 subtype showed a better prognosis than other subtypes, which is consistent with our defined C3 cluster showing the best prognosis, as shown in Figure 1F. The gene mutations in each cluster were compared and the top 20 genes with a lower p-value were illustrated in Figure 2C. Most mutations were present in KRAS, TP53, and SMAD4, accounting for 75.3%, 28.2%, and 19.7%, respectively. It was noticed that the C1 cluster with a poor prognosis had more gene mutations.

3.3 Pathway characteristics among molecular subtypes

GSEA was performed to elucidate the pathway features in each cluster by using the Hallmark candidate gene sets. As shown in Figure 3A, the C1 cluster was significantly enriched in 38 pathways in the TCGA cohort. Generally, the activated pathways mainly included cell cycle-related pathways, such as E2F_TARGETS, G2M_CHECKPOINT, MYC_TARGETS_V1, whereas the inhibited pathways primarily contained INFLAMMATORY_RESPONSE, COMPLEMENT, and INTERFERON_GAMMA_RESPONSE. Similar results were also observed in the PACA-CA cohort. In addition, we compared the pathway characteristics between clusters (Figures 3B–D). It revealed that PC patients with the 3 subtype had activated immune pathways, such as cell cycle-related pathways, indicating that the 32 NKGs might play vital roles in the regulation of cell cycle and TME.

3.4 Immune signatures between molecular subtypes and differences in immunotherapy/chemotherapy/PCD

Furthermore, we assessed the relative abundance of 22 immune cells in the TCGA-PAAD and PACA-CA cohorts using the CIBERSORT algorithm. As shown in Figures 4A, C, significant differences among three clusters were observed for several immune cell types, such as CD8+T cells and activated CD4$^+$ memory T cells. Meanwhile, we observed a significantly higher immune score in the C3 cluster than in other clusters (Figures 4B, D), indicating that the C3 cluster had a higher immune infiltration. In addition, we investigated the 7 inflammation-related metagenes clusters in the three molecular subtypes. As a result, 6 of the 7 metagenes clusters were significantly differently expressed among subtypes, except for interferon (Figure 4E). Overall, the C1 cluster presented with a higher inflammation activity than other clusters. Meanwhile, we also observed a higher enrichment score of LCK and MHC-II, and STAT1 in the C1 cluster than the other two clusters in the PACA-CA cohort (Figure 4F). The ssGSEA analysis of 12 PCD patterns indicated that 10 PCD patterns had obviously differences among 3 subtypes, and in general, C1 or C3 subtype had higher ssGSEA scores (Figure 4G).

3.5 Immunotherapy response and drug sensitivity among clusters

Immunotherapy achieved favorable therapeutic effects in various cancers and immune checkpoint genes (ICG) play vital roles in these processes. Therefore, we evaluated the expression of ICGs among clusters and found an elevated expression of PD-1, PD-L1, and CTLA4 in the C3 cluster, as shown in Figure 5A. Meanwhile, we assessed the capability of clusters in predicting immunotherapy response using the T cell inflamed GEP score and observed a higher score in the C3 cluster than in other clusters (Figure 5B). INF-γ is a cytokine that plays a key role in immune regulation and anticancer immunity (Zhang et al., 2017), therefore, we calculated the ssGSEA score of the GOBP_RESPONSE_TO_INTERFERON_GAMMA gene set and found a significantly higher score of INF-γ response in the C3 cluster (Figure 5C). In addition, we also observed a higher CYT score in the C3 cluster than in other clusters (Figure 5D), which was used to reflect cytotoxic effects. Moreover, our data showed that the C1 cluster was more sensitive to cisplatin, gemcitabine, and erlotinib.

3.6 Establishment of a risk signature

A total of 294 DEGs among clusters were identified, as shown in Figures 6A–C. Enrichment analysis on the DEGs was

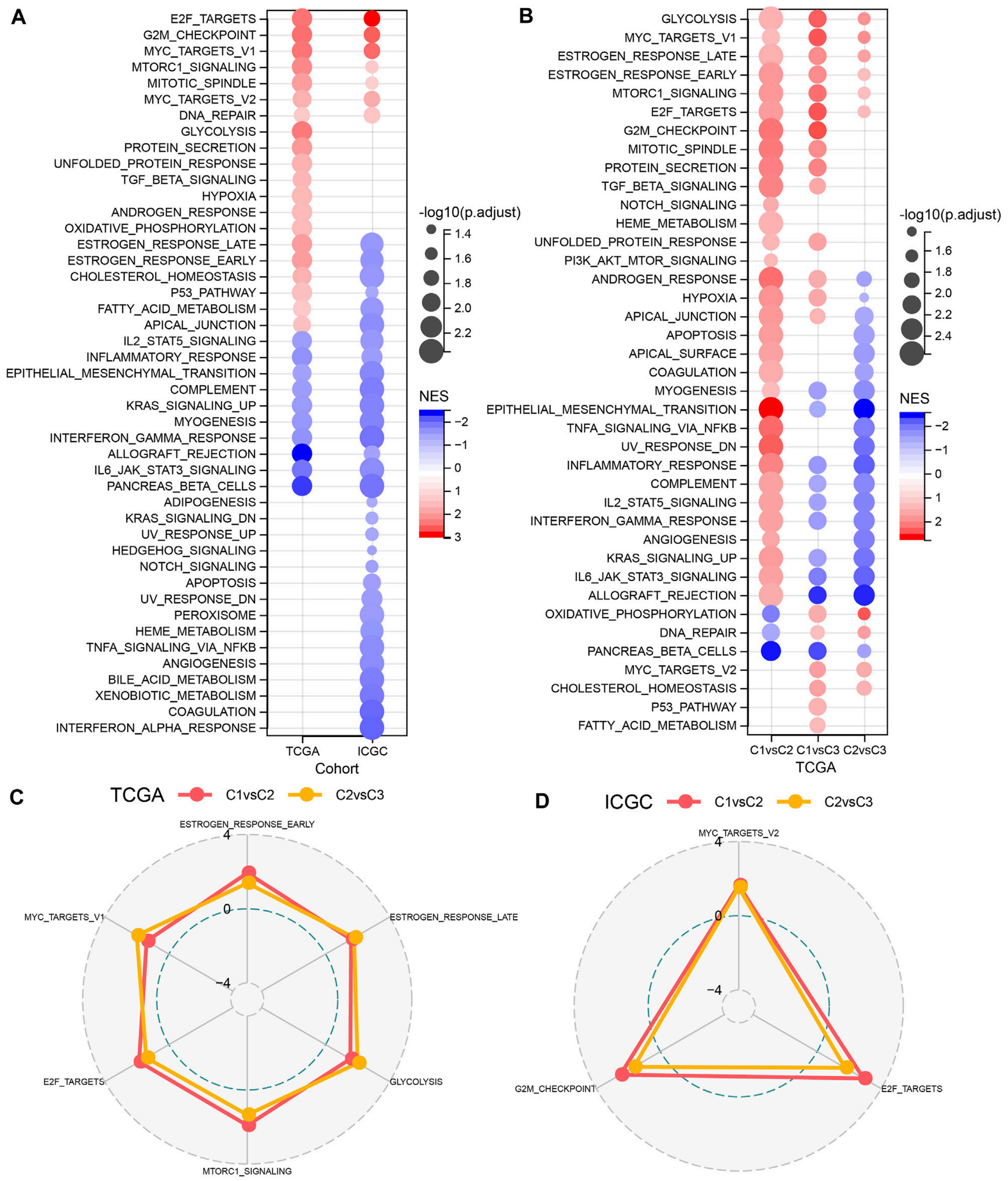


FIGURE 3
The comparison of pathways between molecular subtypes. **(A)** Bubble chart of GSEA analysis results of the TCGA cohort and the ICGC cohort. **(B)** Bubble chart of the GSEA analysis results of C1 vs. C3, C1 vs. C2, and C2 vs. C3 in the TCGA cohorts; **(C)** The radar chart of the C1vsC2, C2vsC3 coherent activation pathway in the TCGA cohort; **(D)** The radar plot of the C1vsC2, C2vsC3 coherent activation pathway in the ICGC cohort.

performed and the results showed that the C3 cluster contained DEGs that were significantly associated with immune-related pathways (Figure 6D). Univariate COX analysis showed that 122 of the 293 DEGs were significantly associated with the prognosis of PC ($p < 0.01$), including 84 risk genes and 38 protective genes (Figure 7A). Subsequently, lasso COX

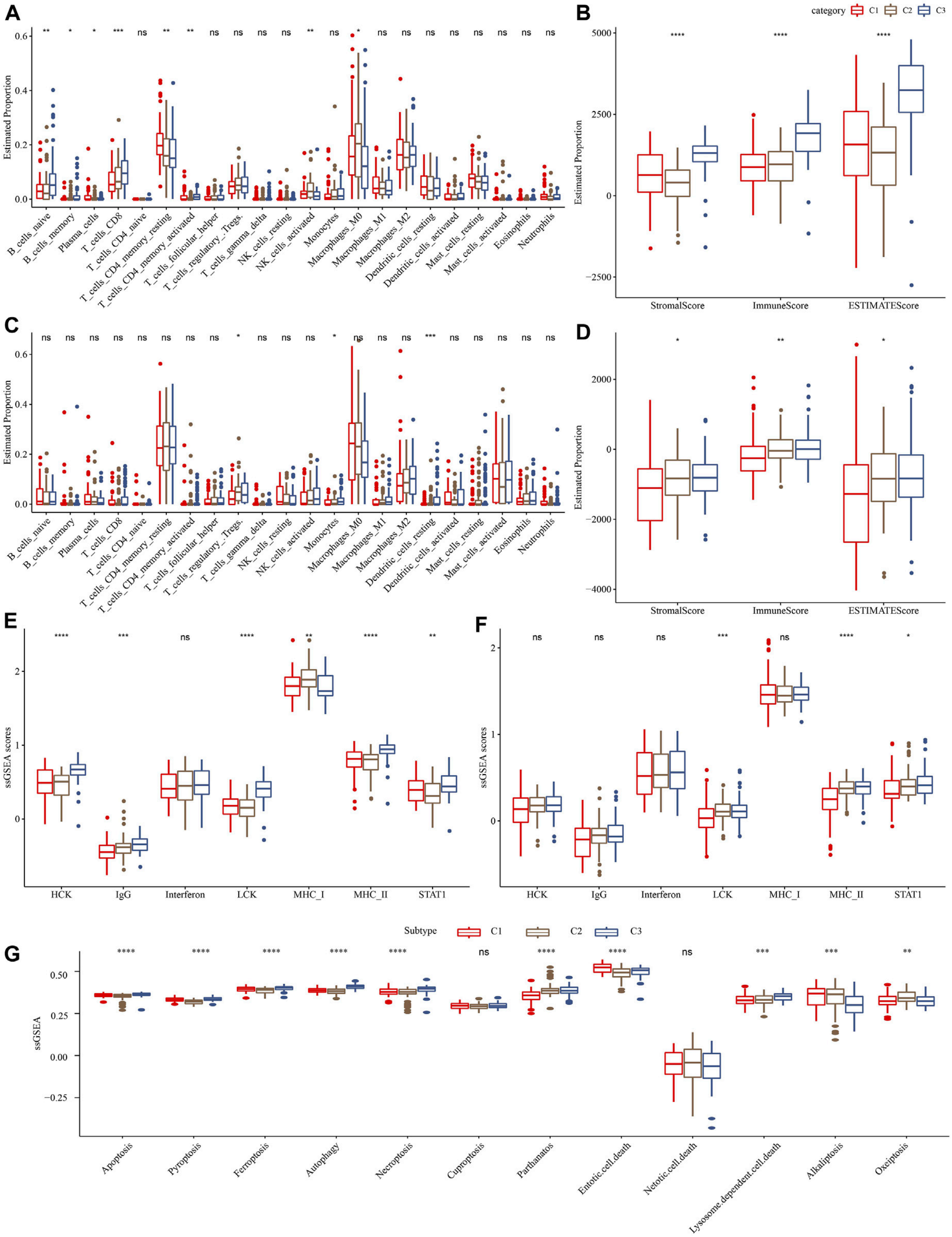


FIGURE 4
Comparison of immune infiltration and inflammation activity among three molecular subtypes. **(A)** and **(C)** represented the differences in the relative abundance of 22 immune cells among different molecular subtypes in the TCGA-PAAD and PACA-CA cohorts, respectively. **(B)** and **(D)** represented the comparison of the ESTIMETE results among clusters in the TCGA-PAAD and PACA-CA cohorts, respectively. **(E)** and **(F)** represented the differences in the inflammation activity among clusters in the TCGA-PAAD and PACA-CA cohorts, respectively. **(G)** The ssGSEA score differences of 12 programmed cell death patterns among 3 molecular subtypes.

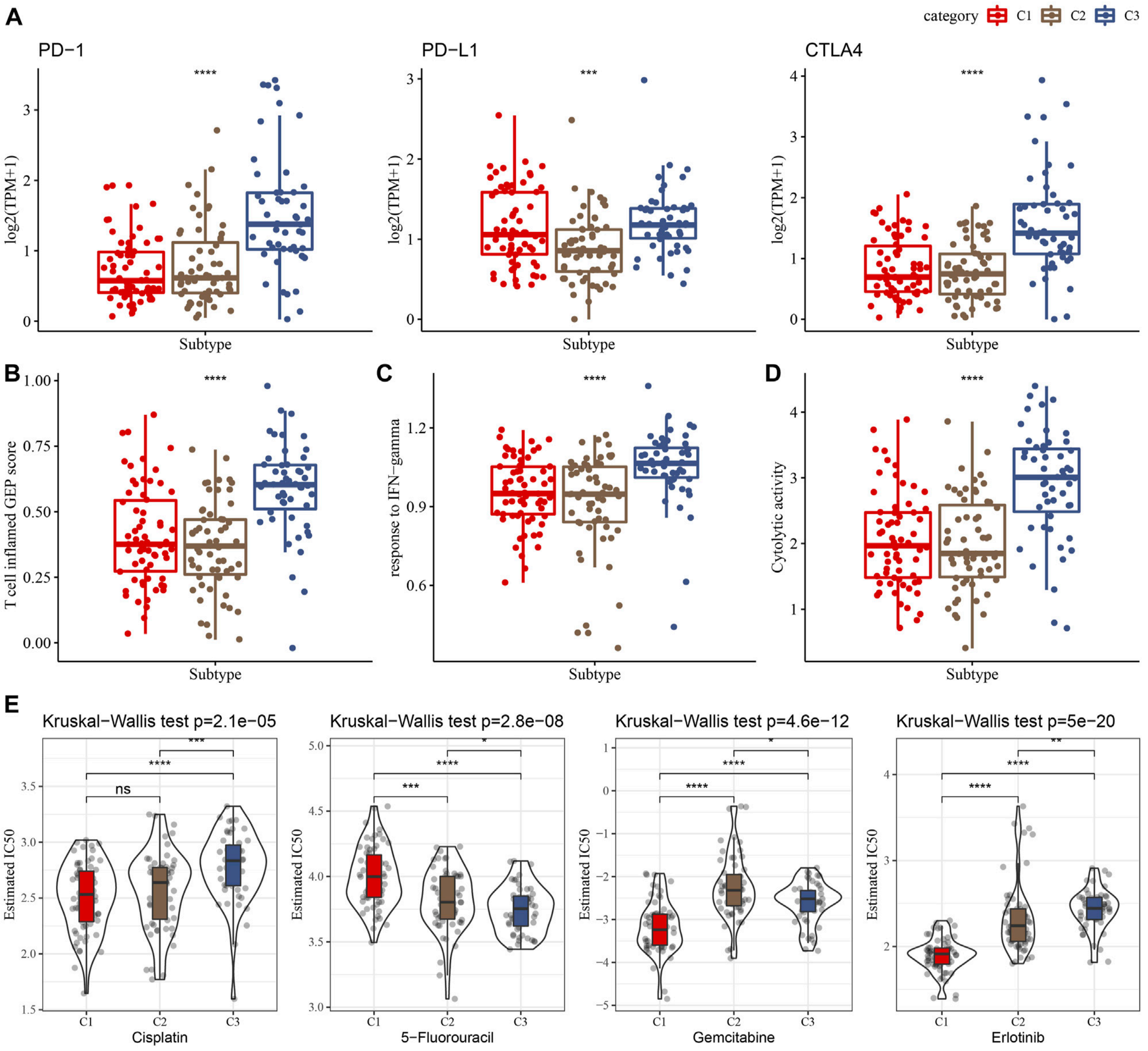


FIGURE 5
Differences in the immunotherapy response and drug sensitivity among clusters in the TCGA cohort. **(A)** Comparison of the ICGs among clusters. **(B–D)** Showing comparisons of the T cell inflamed GEP score, response to IFN-γ, and Cytolytic activity among clusters, respectively. **(E)** The box plots of the estimated IC_{50} for cisplatin, 5-Fluorouracil, gemcitabine and erlotinib in the three clusters.

regression was adopted to compress the gene number and found 9 candidate genes when lambda = 0.0666 (Figures 7B, C). Finally, four genes were identified after stepwise multivariate regression analysis on the 9 candidate genes and were used to construct a prognosis model (Figure 7D), RiskScore = +0.306*MET+0.299*EMP1-0.225*NGFR+0.182*MYEOV. The four gene expressions were negatively correlated with methylation level (Supplementary Figure S2). The risk score was calculated for each patient in the TCGA cohort and was used to divided the patient into the high and low group (Figure 8A). ROC analysis demonstrated a favorable predictive capability in forecasting the 1-, 3-, and 5-year survival rates (Figure 8B). Survival analysis showed a significantly difference in prognosis between the high and low groups (Figure 8C). In addition, we evaluated the robustness of the prognosis model in the PACA-CA and GSE57495 cohorts, which had similar results as the TCGA cohort (Figures 8D–G).

3.7 Differences in clinicopathological features and clusters between the high and low groups

The correlations between risk score and clinicopathological characteristics were analyzed in the TCGA and PACA-CA cohorts, and the results found significant associations between

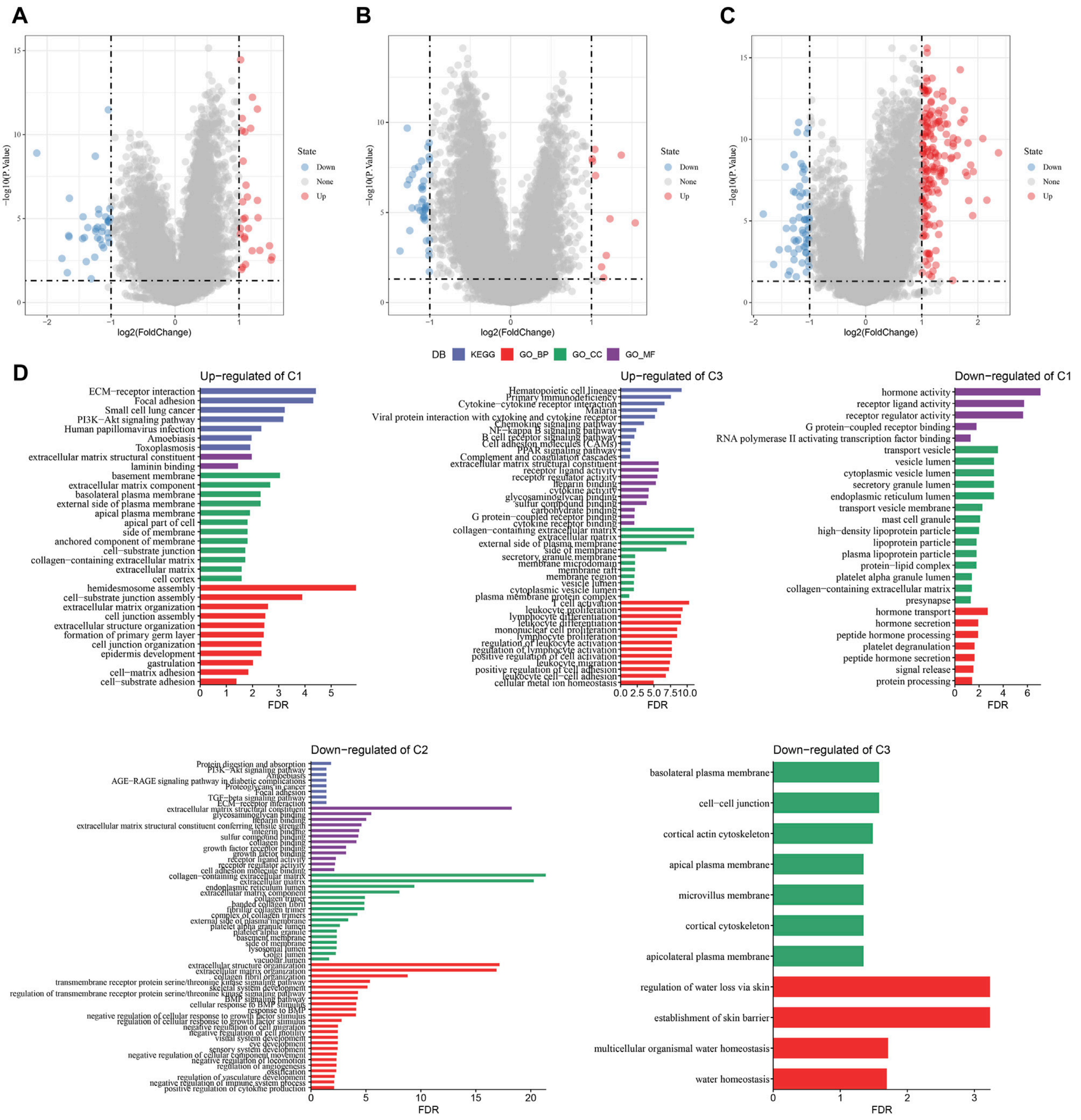


FIGURE 6
The identification of DEGs in each cluster. **(A–C)** Volcano plot of DEGs in the TCGA-PAAD cohort; **(D)** Functional enrichment analysis of DEGs of each cluster.

risk score and grade, but not stage, age, and gender (Figures 9A, C). Meanwhile, the risk score was significantly different among the three clusters, which manifested by a higher risk score in the C1 cluster and a lower risk score in the C3 cluster (Figures 9B, D). In addition, K-M curves showed that the risk score exhibited a favorable capability in the prognostic prediction of PC in sub-populations with specific clinicopathological features (Figures 9E, F).

3.8 Immune infiltration and pathway characteristics in different risk groups

As shown in Figure 10A, we observed a significantly difference in the relative abundance of four immune cells, including naive B cells, CD8 T cells, monocytes, and M0 macrophages, between the high and low groups in the

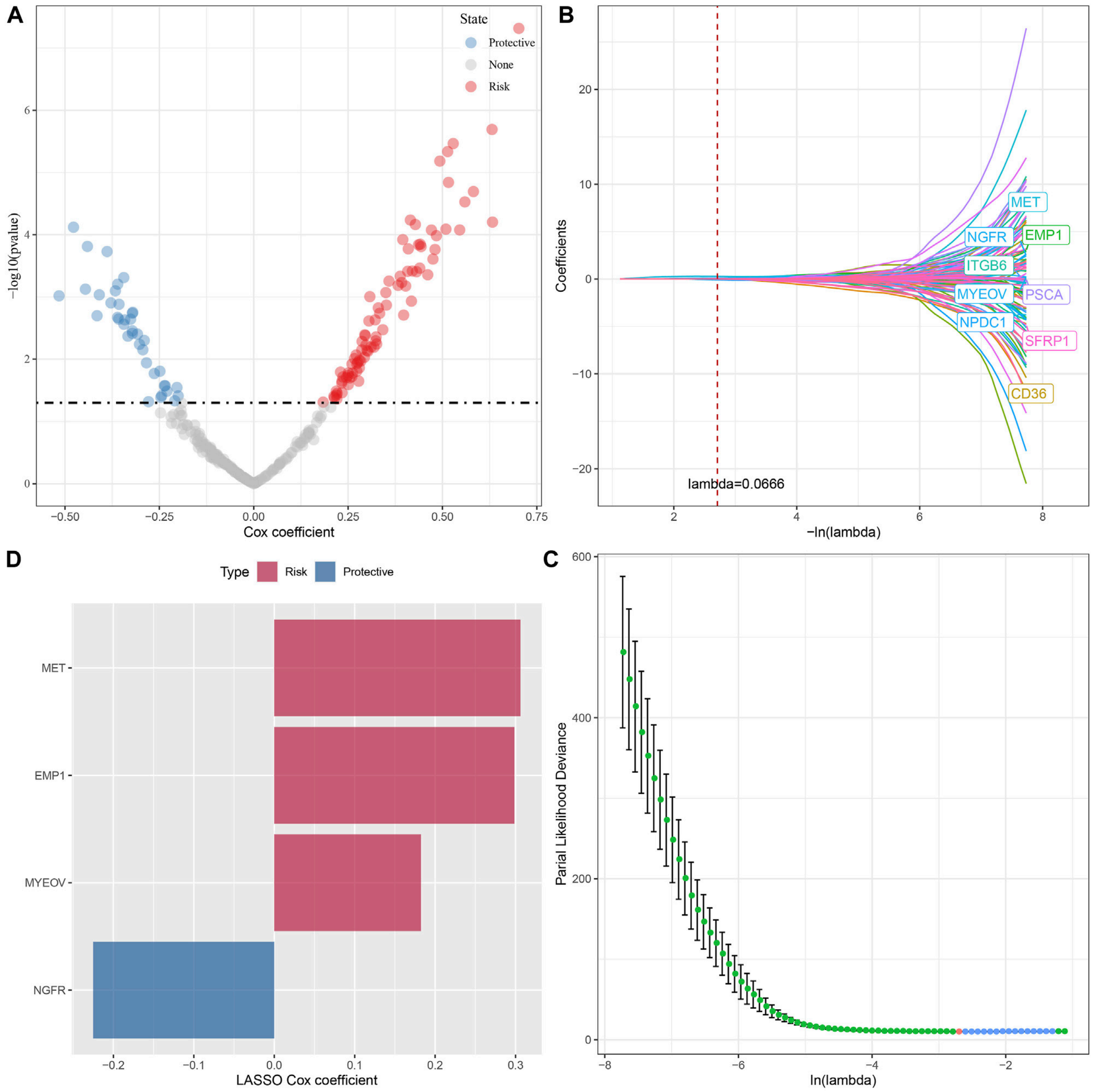


FIGURE 7
The identification of hub genes for the construction of the prognosis model, **(A)** A total of 122 promising candidates were identified from the DEGs; **(B)** The trajectory of each independent variable with lambda; **(C)** Confidence interval under lambda; **(D)** Distribution of LASSO coefficients of the Natural Killer Cell-related prognostic gene signature.

TCGA cohort. The correlations between risk score and immune cells were illustrated in Figure 10B. In addition, a higher immune score was observed in the low group than the high group, indicating a higher immune infiltration in the low group (Figure 10C). The ssGSEA scores on each pathway were calculated for individuals and were compared between two risk groups. The results demonstrated that the High group was significantly associated with cell cycle-related pathways (Figures 10D, E).

3.9 Immunotherapy response, chemotherapy sensitivity and PCD between two risk groups

As shown in Figure 11A, we observed a significantly higher T cell inflamed GEP score in the Low group as compared with those in the High group. Our data also revealed a significantly higher response to IFN-γ and cytolytic activity in the Low group, when compared with the high group (Figures 11B, C). In addition, we found elevated expression

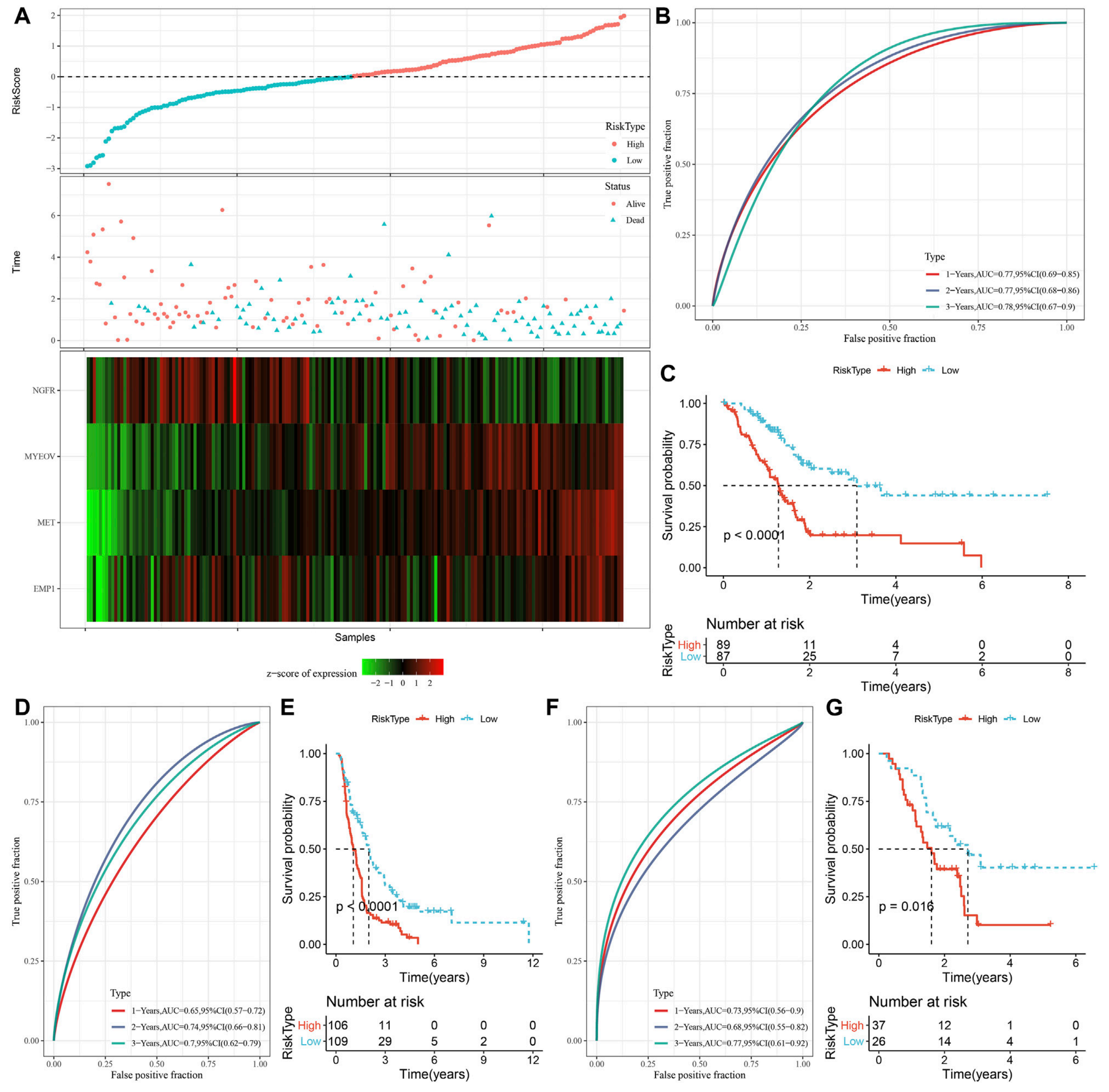


FIGURE 8
Construction and validation of the prognosis model for PC. **(A)** The risk scores of patients in the TCGA cohort. **(B)** The ROC results of the prognostic model in the TCGA cohort. **(C)** The survival analysis results of the prognostic model in the TCGA cohort. **(D, E)** ROC curve and KM survival curve of risk score in PACA-CA cohort; **(F, G)** ROC curve and KM survival curve of risk score in GSE57495 cohort.

of PD-1 and CTLA4, but not PD-L1, in the low group (Figure 11D), suggesting potential differences in immunotherapy response between the two risk groups. Chemotherapy sensitivity in different risk groups was analyzed and found that the patients in the high group were more likely to be sensitive to gemcitabine, cisplatin, and erlotinib, as shown in Figure 11E. In addition, four of 12 PCD patterns had increased ssGSEA score in low group, while 3 PCD had higher ssGSEA score in high group (Figure 11F, left). Furthermore, we analyzed the correlation between RiskScore, four model genes and 12 PCD patterns, and there were different degrees of correlation with each other (Figure 11F, right).

3.10 Improvement of the prognostic model

As shown in Figure 12A, a decision tree was constructed based on the risk score and clinicopathological features and generated four groups (Lowest, Low, Mediate, High) using three parameters (risk score, N stage, age). Survival analysis demonstrated significant differences in prognosis among the four groups (Figure 12B, $p < 0.001$). The correlations between the decision tree-derived groups and risk groups were illustrated in Figures 12C, D. Univariate regression analysis showed that T stage, N stage, age, and risk

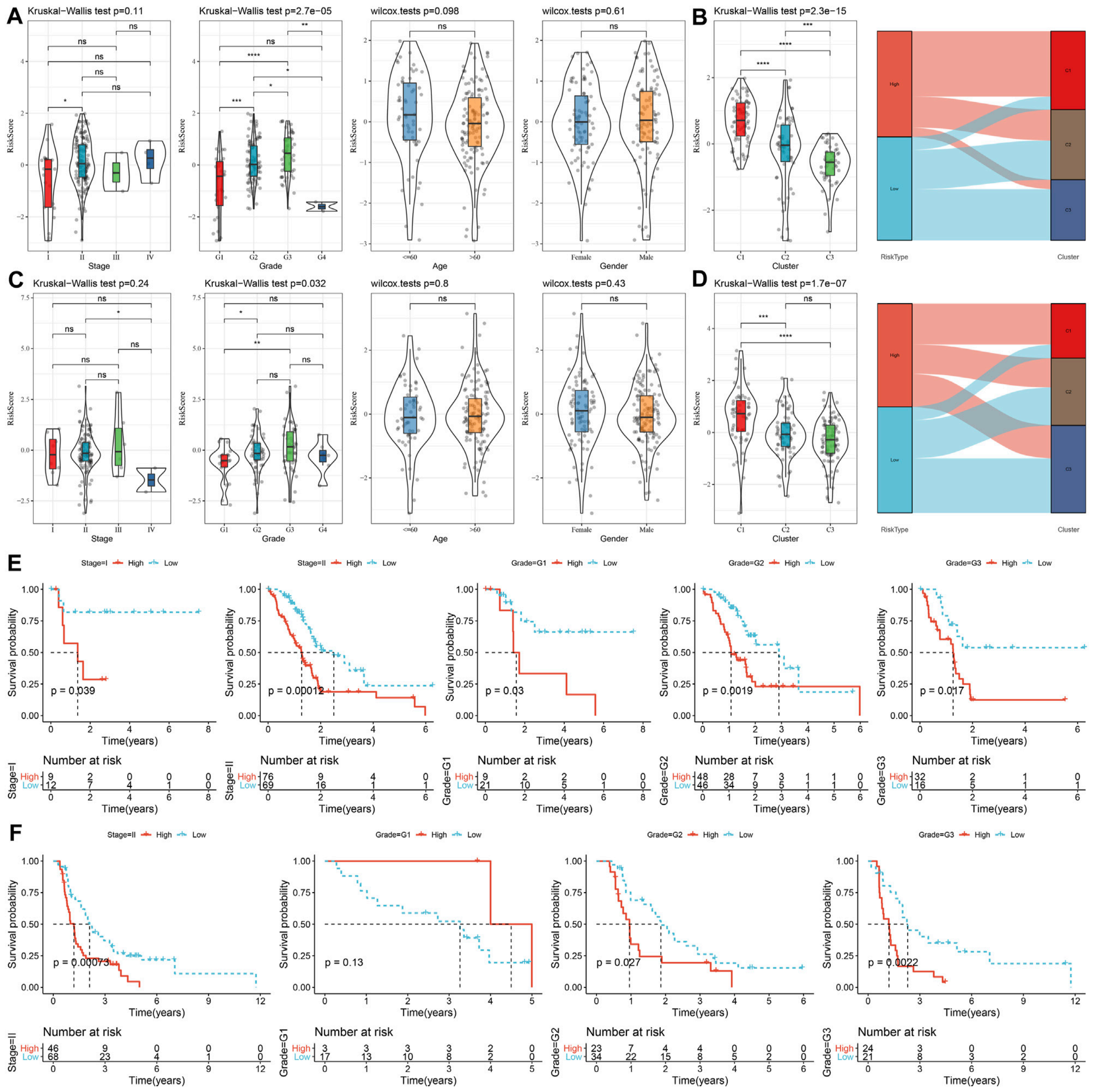


FIGURE 9
The distribution and predictive capability of the risk score in sub-population with distinct clinicopathological features. **(A)** and **(C)** Differences in risk score among different clinicopathological groups in the TCGA-PAAD and PACA-CA cohorts, respectively; **(B)** and **(D)** Difference in risk score among different molecular subtypes in the TCGA-PAAD and PACA-CA cohorts, respectively; **(E)** and **(F)** K-M curve of risk score-derived groups in different clinicopathological groups the TCGA-PAAD and PACA-CA cohorts, respectively.

score was associated with the prognosis of PC, and three of them (N stage, age, and risk score) were identified as independent risk factors *via* multivariate regression analysis (Figures 12E, F). Therefore, a nomogram was built using the three factors (Figure 12G). It was observed that the predicted values were close to the observed values in terms of the 1-, 2, and 3-year OS (Figure 12H), indicating that the nomogram had good prediction performance. In addition, a decision curve was used to evaluate the reliability of the model, and it was observed that the risk signature and nomogram model had a higher standardized net benefit as compared with other clinicopathological features (Figure 12I).

4 Discussion

Tumor immunotherapy has brought hope for cancer treatment, and more and more studies have shown that innate immune cells, including NK cells, have unique advantages in anti-

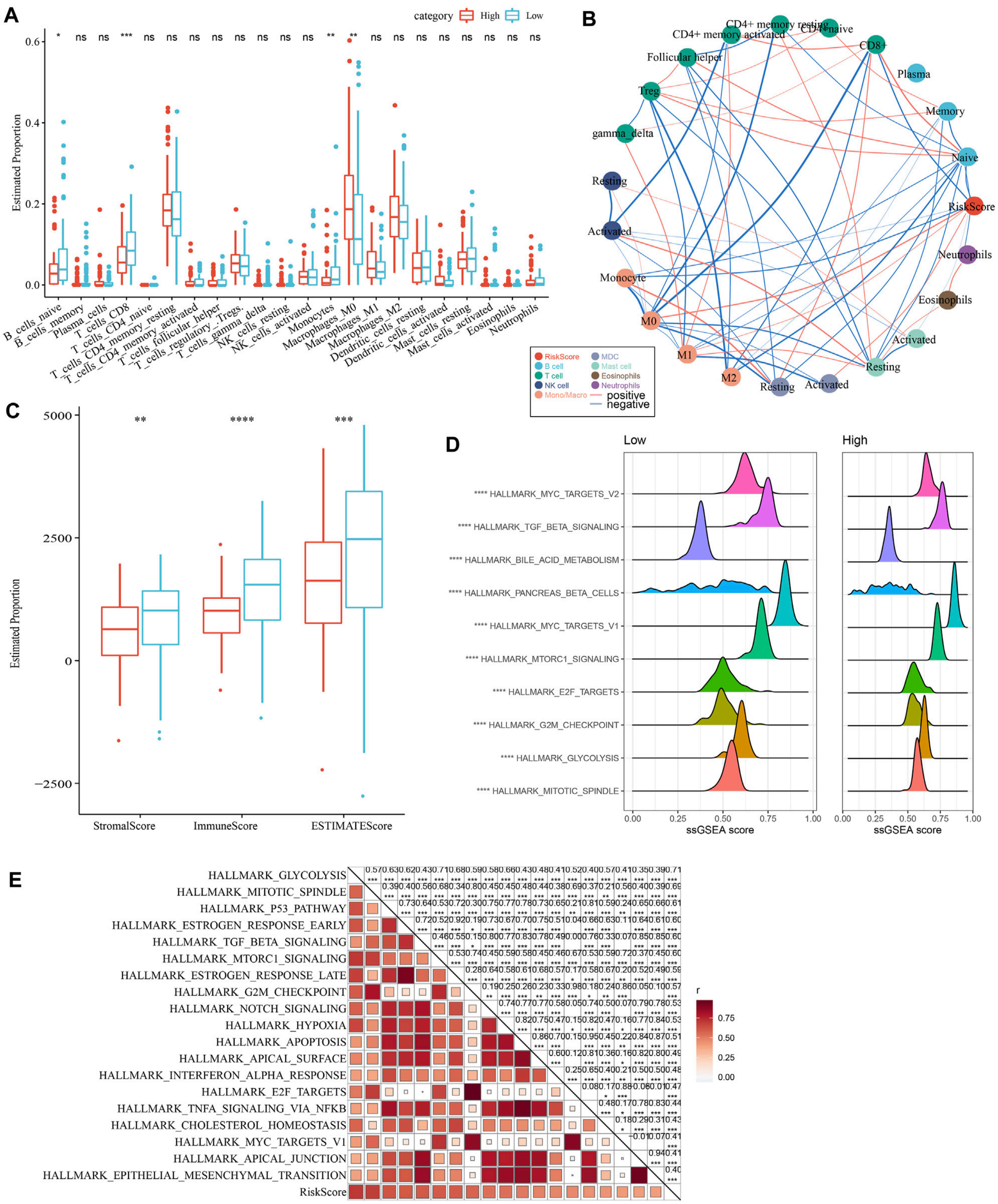


FIGURE 10
Comparison of immune infiltration and pathways between two risk groups. **(A)** Comparison of the proportion of immune cells in the TCGA cohort; **(B)** Correlation analysis between 22 immune cells and risk score in the TCGA cohort; **(C)** Comparison of the results of ESTIMATE between two risk groups; **(D)** The top 10 pathway with the most significant difference between the two risk groups. **(E)** Correlation analysis between pathways and risk score.

tumor immunotherapy. However, most of the current research focuses on adaptive immune cells, and the role of innate immune cells has not been paid enough attention. Studies have shown that the abundance of tumor infiltrating NK cells is closely related to the prognosis of patients with various solid tumors (Villegas et al., 2002; Cursons et al., 2019; Meng et al.,

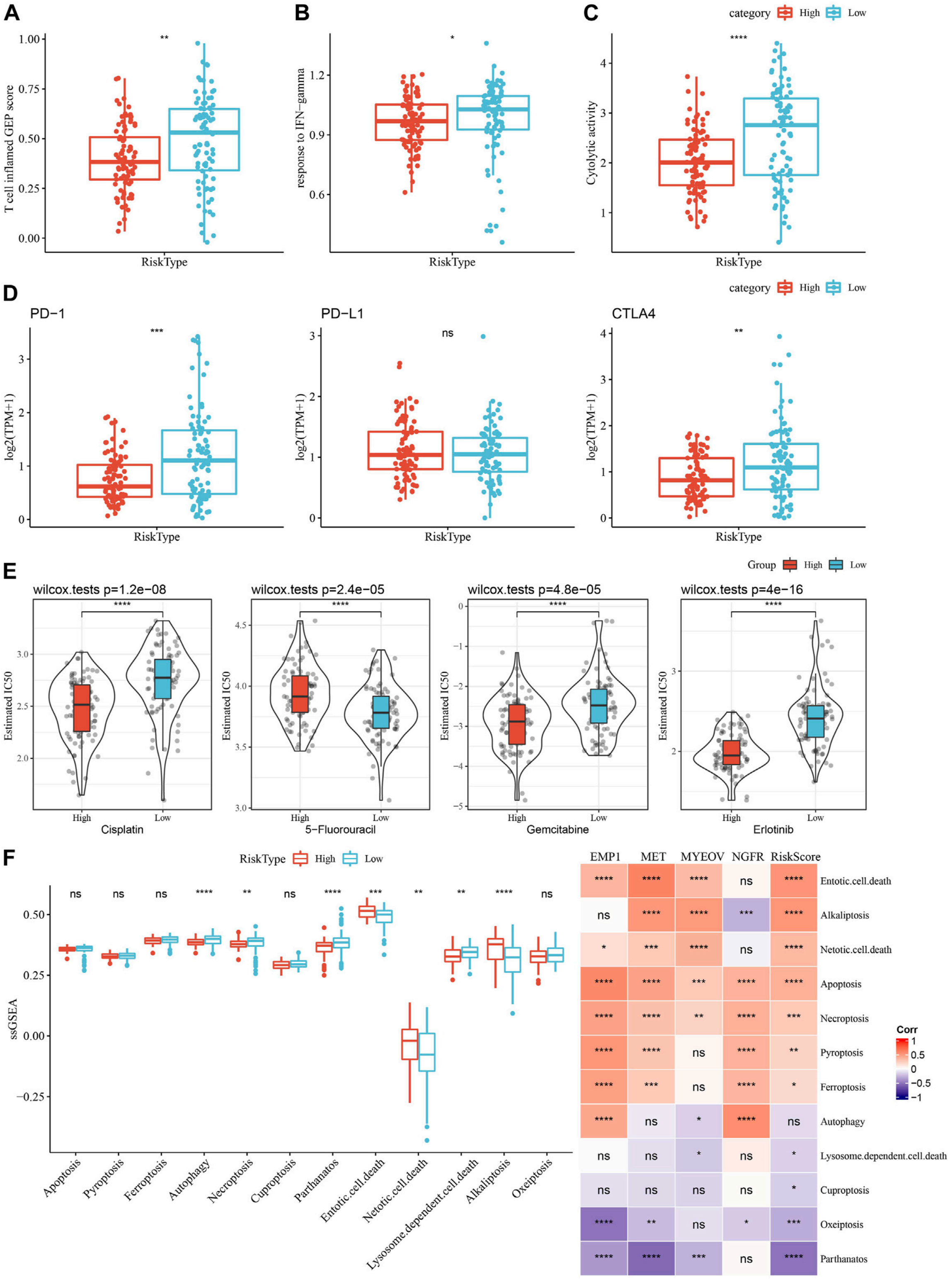


FIGURE 11
Comparison of the immunotherapy response and chemotherapy sensitivity between two risk groups in TCGA-PAAD cohort. **(A–C)** Represented the comparison of the T cell inflamed GEP score, response to IFN-γscore and cytolytic activity between the two risk groups, respectively. **(D)** Differences of expression ICGs between different groups; **(E)** The box plots of the estimated IC_{50} for cisplatin, 5-Fluorouracil, gemcitabine and erlotinib between the risk groups. **(F)** Left, the ssGSEA score differences of 12 programmed cell death patterns between high- and low-group. Right, the correlation analysis between 12 programmed cell death patterns and RiskScore.

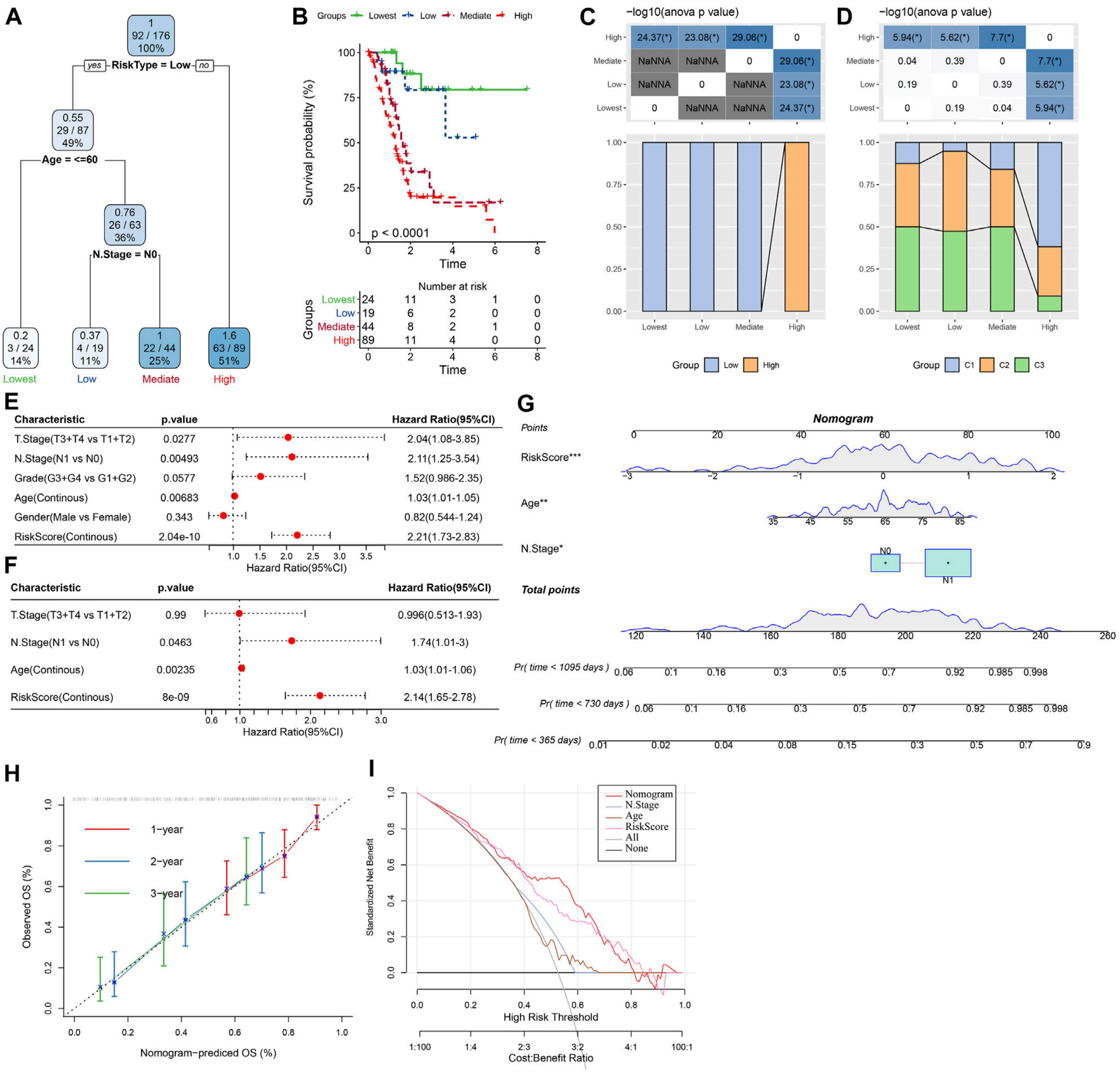


FIGURE 12
Construction of prognostic models of PC using the risk score and clinicopathological features. **(A)** A decision tree model generated four risk groups based on age, N stage and the risk score. **(B)** Survival analysis of the decision tree-derived risk groups showing distinct prognoses among the groups. **(C, D)** showed the correlations between risk score-derived groups and decision tree-derived groups. **(E)** and **(F)** Univariate and multivariate Cox analysis of risk score and clinicopathological characteristics. **(G)** The nomogram model consists of age, N stage, and the risk score; **(H)** 1-, 2-, and 3-year calibration curves of the established nomogram; **(I)** The decision curves showing the capacity for survival prediction.

2019). The prognostic model based on NKG has the potential ability to predict prognosis and immunotherapy response (Cursons et al., 2019). Meanwhile, a novel human NK cell-based immunotherapy was developed and showed efficacy in human metastatic PC models (Teng et al., 2022). Inspired by these findings, we attempted to investigate the molecular subtypes of PC based on prognosis-related NKGs using transcriptomic data in this study. Distinct differences in prognosis, immunotherapy response, and drug sensitivity among subtypes were observed, indicating the crucial role of NK cells in the progression and treatment of PC. Functional enrichment analysis showed that NKGs involved in activated immune pathways, such as cell cycle-related pathways, indicating that the those NKGs might play vital roles in the regulation of cell cycle and TME. Furthermore, we developed a novel prognostic prediction signature based on DEGs that were found among NKGs-derived molecular subtypes of PC, which exerts a favorable capability of prognostic prediction.

Herein, we identified 32 prognosis-related NKGs in PC, including 12 protective genes and 20 risk genes, and the expression of most of these genes was significantly correlated. A number of studies had proposed potential roles of these NKGs in PC. For instance, the tumor necrosis factor ligand superfamily member 10 (TNFSF10), also known as TRAIL, encodes a cytokine that belongs to the tumor necrosis factor (TNF) ligand family, it preferentially induces apoptosis in transformed and tumor cells and was proposed as a prognostic indicator of PC (Wang et al., 2021; Wang et al., 2022). As a well-known driver gene, KRAS frequently mutated in PC patients (Waters and Der, 2018), our data revealed that KRAS was the most mutated gene in the TCGA-PAAD cohort. KRAS gene mutations has been reported to be involved in the invasion and metastasis of tumor cells, as well as chemoresistance (Mueller et al., 2018; Buscail et al., 2020). It was found that TMB was associated with the sensitivity of immunotherapy response and was more effective than ICG expression in screening patients suitable for immunotherapy (Choucair et al., 2020). This finding may result from the enrichment of immune cells due to the elevated production of "non-self" neoantigen under high TMB (Schumacher and Schreiber, 2015). In addition, it was observed that the phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform (PIK3CB) was involved in metastasis of PC cells (Qu et al., 2021). Therefore, further investigation on these prognostic NKGs and their mutations might provide clues for the development of novel treatment of PC.

Three stable clusters with distinct differences in prognosis were generated based on the prognostic NKGs, and GSEA results found significant differences in cell cycle pathways and immunity-related pathways among clusters. Therefore, the inferior prognosis of patients in the C1 cluster may be partly attributed to the disturbance of cell cycle regulation, which is closely related to tumor proliferation and progression (Tang et al., 2020). Meanwhile, these data indicated that the prognostic NKGs used for molecular typing play important roles in the cell cycle process and tumor immune microenvironment. For example, Rac1 plays an important role in regulating cell function, and its activation affects cell morphology (Etienne-Manneville and Hall, 2002), cell cycle and gene expression (Yoshida et al., 2010), survival and apoptosis (Liang et al., 2021). Tyrosine kinase FYN was reported to be associated with mediating mitogenic signals and involved in regulating cell cycle and proliferation (Zheng et al., 2017). Besides, we observed significant differences in immune cells infiltration among NKG-derived clusters. The C1 cluster was characterized as so-called "cold tumor" since it presented with a lower immune cell infiltration. The tumor-infiltrating immune cells participated in tumor development and influence prognosis (Barnes and Amir, 2017), and anti-tumor activity of "cold tumor" is decreased because low immune cell infiltration could increase tumor cell escape from immune surveillance and contribute to tumor progression (Bonaventura et al., 2019). These finding may partly contribute to the significant reduction in survival of the C1 and C2 clusters. Meanwhile, a lower stromal score was observed in the C1 and C2 clusters, which was suggested to be associated with a poor OS of osteosarcoma (Alves et al., 2019).

Since GSEA revealed significant inhibition of inflammatory response among clusters, we further evaluated the relationships between NKG-derived clusters and inflammatory activities by analyzing inflammatory-related metagenes. Notably, significant differences in hematopoietic cell kinase (HCK), IgG, MHC-II, src-family kinases p56 (LCK), MHC-I, and were observed among clusters. HCK plays a pivotal role in innate immunity and was overexpressed in various cancers. It could regulate the phagocytosis of neutrophils and macrophages (Roseweir et al., 2019), as well as immune cell infiltration in the TME (Cheng et al., 2022). LCK is critical for proximal T-cell antigen receptor (TCR) signal transduction and is involved in the earliest steps of TCR-mediated T-cell activation (Salmond et al., 2009). MHC-I and MHC-II are two pivotal molecules presenting with the function of antigenpresentation, and their loss of expression would make tumor cells escape T-cell killing (Garrido and Aptsiauri, 2019). Therefore, a lower level of these inflammatory-related metagenes may partly account for the immunosuppressive microenvironment in the C1 and C2 clusters.

Discrepancy between inflammatory activities and immune cell infiltration among clusters prompted us to explore the immunotherapy response. It was suggested that ICG expression partly contribute to the success of immune checkpoint blockade therapy. Herein, we revealed significant differences in ICG expression among clusters, indicating potential differences in the response to immunotherapy among clusters. In addition, a T cell-inflamed gene expression profile (GEP) was found to be effective in predicting response to anti-PD-1-directed therapy (Ayers et al., 2017). Our data showed that the C3 cluster had a significantly higher T cell-inflamed GEP score, indicating that PC patients in the C3 cluster might be more sensitive to anti-PD-1 therapy. Cytokine IFN-γ plays a key role in anticancer immunity and immune regulation, and the C3 cluster presented with a higher elevated expression of the gene set that responds to IFN-γ. Moreover, the cytolytic activity score (CYT) has been considered as a useful tool to evaluate anti-tumor immunity. It has been revealed that high CYT was associated with better prognosis of colorectal cancer, which could be explained by increased immunity and cytolytic activity of T cells and M1 macrophages (Narayanan et al., 2018). In this study, elevated cytolytic activity was observed in the C3 cluster. Moreover, our data also revealed the differences in chemotherapeutic drug sensitivity among clusters. The clusters derived from NKG have significant differences in immunotherapy and chemotherapy responses, which has potential value to guide individualized treatment strategies.

On the basis of NKG-derived clusters, we established a novel prognostic signature using the DEGs found among clusters. This prognostic signature has satisfactory prognostic performance and has shown good predictive power in immunotherapy response and chemotherapeutic drug sensitivity. Despite the promising findings obtained, several limitations in this study should be acknowledged. First, due to the high heterogeneity of the tumor immune microenvironment, the prognosis-predicting ability of NKG-derived molecular subtypes and subsequent prognostic models was limited. Second, analysis of NK cell characteristics based on single cell sequencing will help to further understand its role in PC. Finally, further study is required to investigate the underlying mechanism of the genes in the risk signature and PC patients' outcomes.

5 Conclusion

In conclusion, we established three molecular clusters of PC using 32 prognosis-related NKGs and revealed differences in clinicopathological and genomic features, pathways, immunotherapy response, and drug sensitivity among clusters. Furthermore, a prognostic signature with robust prognosis-predicting ability was built and validated. The NKG-derived molecular clusters and prognostic signature might serve as a useful tool for assisting in the decision of individualized treatment and the selection of suitable individuals for chemotherapy.

Data availability statement

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

Author contributions

YL: Writing; YL, QJ, and MF collecting data; PZ analyzed the data; MZ supervised and submitted the paper.

Funding

The present study was supported by the National Natural Science Foundation of China (81972002); Natural Science Foundation of Shandong Province (ZR2022MC174).

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.2023.1100020/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Working flow chart

SUPPLEMENTARY FIGURE S2
The correlation of expression and Methylation level in 4 genes

References

Alves, P. M., de Arruda, J. A. A., Arantes, D. A. C., Costa, S. F. S., Souza, L. L., Pontes, H. A. R., et al. (2019). Evaluation of tumor-infiltrating lymphocytes in osteosarcomas of the jaws: A multicenter study. *Virchows Arch.* 474 (2), 201–207. doi:10.1007/s00428-018-2499-6

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

Balachandran, V. P., Łuksza, M., Zhao, J. N., Makarov, V., Moral, J. A., Remark, R., et al. (2017). Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* 551 (7681), 512–516. doi:10.1038/nature24462

Barnes, T. A., and Amir, E. (2017). HYPE or HOPE: The prognostic value of infiltrating immune cells in cancer. *Br. J. Cancer* 117 (4), 451–460. doi:10.1038/bjc.2017.220

Bonaventura, P., Shekarian, T., Alcazer, V., Valladeau-Guilemond, J., Valsesia-Wittmann, S., Amigorena, S., et al. (2019). Cold tumors: A therapeutic challenge for immunotherapy. *Front. Immunol.* 10, 168. doi:10.3389/fimmu.2019.00168

Boyiadzis, M. M., Kirkwood, J. M., Marshall, J. L., Pritchard, C. C., Azad, N. S., and Gulley, J. L. (2018). Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. *J. Immunother. Cancer* 6 (1), 35. doi:10.1186/s40425-018-0342-x

Buscail, L., Bournet, B., and Cordelier, P. (2020). Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* 17 (3), 153–168. doi:10.1038/s41575-019-0245-4

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 9 (1), 34. doi:10.1186/s13073-017-0424-2

Cheng, F., Li, Q., Wang, J., Wang, L., Li, W., and Zeng, F. (2022). HCK is a potential prognostic biomarker that correlates with immune cell infiltration in acute myeloid leukemia. *Dis. Markers* 2022, 3199589. doi:10.1155/2022/3199589

Cho, Y. H., Kim, M. S., Chung, H. S., and Hwang, E. C. (2017). Novel immunotherapy in metastatic renal cell carcinoma. *Investig. Clin. Urol.* 58 (4), 220–227. doi:10.4111/icu.2017.58.4.220

Choucair, K., Morand, S., Stanbery, L., Edelman, G., Dworkin, L., and Nemunaitis, J. (2020). Tmb: A promising immune-response biomarker, and potential spearhead in advancing targeted therapy trials. *Cancer Gene Ther.* 27 (12), 841–853. doi:10.1038/s41417-020-0174-y

Coca, S., Perez-Piqueras, J., Martinez, D., Colmenarejo, A., Saez, M. A., Vallejo, C., et al. (1997). The prognostic significance of intratumoral natural killer cells in patients with colorectal carcinoma. *Cancer* 79 (12), 2320–2328. doi:10.1002/(sici)1097-0142(19970615)79:12<2320::aid-cncr5>3.0.co;2-p

Considine, B., and Hurwitz, M. E. (2019). Current status and future directions of immunotherapy in renal cell carcinoma. *Curr. Oncol. Rep.* 21 (4), 34. doi:10.1007/s11912-019-0779-1

Cursons, J., Souza-Fonseca-Guimaraes, F., Foroutan, M., Anderson, A., Hollande, F., Hediyeh-Zadeh, S., et al. (2019). A gene signature predicting natural killer cell infiltration and improved survival in melanoma patients. *Cancer Immunol. Res.* 7 (7), 1162–1174. doi:10.1158/2326-6066.CIR-18-0500

Etienne-Manneville, S., and Hall, A. (2002). Rho GTPases in cell biology. *Nature* 420 (6916), 629–635. doi:10.1038/nature01148

Garrido, F., and Aptsiauri, N. (2019). Cancer immune escape: MHC expression in primary tumours versus metastases. *Immunology* 158 (4), 255–266. doi:10.1111/imm.13114

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

Guillerey, C., Huntington, N. D., and Smyth, M. J. (2016). Targeting natural killer cells in cancer immunotherapy. *Nat. Immunol.* 17 (9), 1025–1036. doi:10.1038/ni.3518

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

Hellmann, M. D., Ciuleanu, T. E., Pluzanski, A., Lee, J. S., Otterson, G. A., Audigier-Valette, C., et al. (2018). Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N. Engl. J. Med.* 378 (22), 2093–2104. doi:10.1056/NEJMoa1801946

Huntington, N. D., Vosshenrich, C. A., and Di Santo, J. P. (2007). Developmental pathways that generate natural-killer-cell diversity in mice and humans. *Nat. Rev. Immunol.* 7 (9), 703–714. doi:10.1038/nri2154

Imai, K., Matsuyama, S., Miyake, S., Suga, K., and Nakachi, K. (2000). Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: An 11-year follow-up study of a general population. *Lancet* 356 (9244), 1795–1799. doi:10.1016/S0140-6736(00)03231-1

Ishigami, S., Natsugoe, S., Tokuda, K., Nakajo, A., Xiangming, C., Iwashige, H., et al. (2000). Clinical impact of intratumoral natural killer cell and dendritic cell infiltration in gastric cancer. *Cancer Lett.* 159 (1), 103–108. doi:10.1016/s0304-3835(00)00542-5

Komatsubara, K. M., and Carvajal, R. D. (2017). Immunotherapy for the treatment of uveal melanoma: Current status and emerging therapies. *Curr. Oncol. Rep.* 19 (7), 45. doi:10.1007/s11912-017-0606-5

Lanier, L. L. (2008). Evolutionary struggles between NK cells and viruses. *Nat. Rev. Immunol.* 8 (4), 259–268. doi:10.1038/nri2276

Le, D. T., Lutz, E., Uram, J. N., Sugar, E. A., Onners, B., Solt, S., et al. (2013). Evaluation of ipilimumab in combination with allogeneic pancreatic tumor cells transfected with a GM-CSF gene in previously treated pancreatic cancer. *J. Immunother.* 36 (7), 382–389. doi:10.1097/CJI.0b013e31829fb7a2

Liang, J., Oyang, L., Rao, S., Han, Y., Luo, X., Yi, P., et al. (2021). Rac1, A potential target for tumor therapy. *Front. Oncol.* 11, 674426. doi:10.3389/fonc.2021.674426

Liu, Q., Cheng, R., Kong, X., Wang, Z., Fang, Y., and Wang, J. (2020). Molecular and clinical characterization of PD-1 in breast cancer using large-scale transcriptome data. *Front. Immunol.* 11, 558757. doi:10.3389/fimmu.2020.558757

López-Soto, A., Gonzalez, S., Smyth, M. J., and Galluzzi, L. (2017). Control of metastasis by NK cells. *Cancer Cell.* 32 (2), 135–154. doi:10.1016/j.ccell.2017.06.009

Masiero, M., Simoes, F. C., Han, H. D., Snell, C., Peterkin, T., Bridges, E., et al. (2013). A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis. *Cancer Cell.* 24 (2), 229–241. doi:10.1016/j.ccr.2013.06.004

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

Meng, N., Glorieux, C., Zhang, Y., Liang, L., Zeng, P., Lu, W., et al. (2019). Oncogenic K-ras induces mitochondrial OPA3 expression to promote energy metabolism in pancreatic cancer cells. *Cancers (Basel)* 12 (1), 65. doi:10.3390/cancers12010065

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

Moretta, L., Bottino, C., Pende, D., Castriconi, R., Mingari, M. C., and Moretta, A. (2006). Surface NK receptors and their ligands on tumor cells. *Semin. Immunol.* 18 (3), 151–158. doi:10.1016/j.smim.2006.03.002

Mueller, S., Engleitner, T., Maresch, R., Zukowska, M., Lange, S., Kaltenbacher, T., et al. (2018). Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. *Nature* 554 (7690), 62–68. doi:10.1038/nature25459

Narayanan, S., Kawaguchi, T., Yan, L., Peng, X., Qi, Q., and Takabe, K. (2018). Cytolytic activity score to assess anticancer immunity in colorectal cancer. *Ann. Surg. Oncol.* 25 (8), 2323–2331. doi:10.1245/s10434-018-6506-6

Nelson, M. H., and Paulos, C. M. (2015). Novel immunotherapies for hematologic malignancies. *Immunol. Rev.* 263 (1), 90–105. doi:10.1111/imr.12245

Okasha, H., Elkholy, S., El-Sayed, R., Wifi, M. N., El-Nady, M., El-Nabawi, W., et al. (2017). Real time endoscopic ultrasound elastography and strain ratio in the diagnosis of solid pancreatic lesions. *World J. Gastroenterol.* 23 (32), 5962–5968. doi:10.3748/wjg.v23.i32.5962

O'Reilly, E. M., Oh, D. Y., Dhani, N., Renouf, D. J., Lee, M. A., Sun, W., et al. (2019). Durvalumab with or without tremelimumab for patients with metastatic pancreatic ductal adenocarcinoma: A phase 2 randomized clinical trial. *JAMA Oncol.* 5 (10), 1431–1438. doi:10.1001/jamaoncol.2019.1588

Pulluri, B., Kumar, A., Shaheen, M., Jeter, J., and Sundararajan, S. (2017). Tumor microenvironment changes leading to resistance of immune checkpoint inhibitors in metastatic melanoma and strategies to overcome resistance. *Pharmacol. Res.* 123, 95–102. doi:10.1016/j.phrs.2017.07.006

Qu, J., Zheng, B., Ohuchida, K., Feng, H., Chong, S. J. F., Zhang, X., et al. (2021). PIK3CB is involved in metastasis through the regulation of cell adhesion to collagen I in pancreatic cancer. *J. Adv. Res.* 33, 127–140. doi:10.1016/j.jare.2021.02.002

Rawla, P., Sunkara, T., and Gaduputi, V. (2019). Epidemiology of pancreatic cancer: Global trends, etiology and risk factors. *World J. Oncol.* 10 (1), 10–27. doi:10.14740/wjon1166

Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359 (6382), 1350–1355. doi:10.1126/science.aar4060

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

Roseweir, A. K., Powell, A. G. M. T., Horstman, S. L., Inthagard, J., Park, J. H., McMillan, D. C., et al. (2019). Src family kinases, HCK and FGR, associate with local inflammation and tumour progression in colorectal cancer. *Cell. Signal* 56, 15–22. doi:10.1016/j.cellsig.2019.01.007

Salmond, R. J., Filby, A., Qureshi, I., Caserta, S., and Zamoyska, R. (2009). T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance. *Immunol. Rev.* 228 (1), 9–22. doi:10.1111/j.1600-065X.2008.00745.x

Schumacher, T. N., and Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science* 348 (6230), 69–74. doi:10.1126/science.aaa4971

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* 70 (3), 145–164. doi:10.3322/caac.21601

Souza-Fonseca-Guimaraes, F., Cursons, J., and Huntington, N. D. (2019). The emergence of natural killer cells as a major target in cancer immunotherapy. *Trends Immunol.* 40 (2), 142–158. doi:10.1016/j.it.2018.12.003

Sun, Y., Sedgwick, A. J., Palarasah, Y., Mangiola, S., and Barrow, A. D. (2021). A transcriptional signature of PDGF-DD activated natural killer cells predicts more favorable prognosis in low-grade glioma. *Front. Immunol.* 12, 668391. doi:10.3389/fimmu.2021.668391

Sun, Y., Sedgwick, A. J., Khan, M. A. A. K., Palarasah, Y., Mangiola, S., and Barrow, A. D. (2021). A transcriptional signature of IL-2 expanded natural killer cells predicts more favorable prognosis in bladder cancer. *Front. Immunol.* 12, 724107. doi:10.3389/fimmu.2021.724107

Tang, D., Liu, H., Zhao, Y., Qian, D., Luo, S., Patz, E. F., Jr, et al. (2020). Genetic variants of BIRC3 and NRG1 in the NLRP3 inflammasome pathway are associated with non-small cell lung cancer survival. *Am. J. Cancer Res.* 10 (8), 2582–2595.

Teng, K. Y., Mansour, A. G., Zhu, Z., Li, Z., Tian, L., Ma, S., et al. (2022). Off-the-Shelf prostate stem cell antigen-directed chimeric antigen receptor natural killer cell therapy to treat pancreatic cancer. *Gastroenterology* 162 (4), 1319–1333. doi:10.1053/j.gastro.2021.12.281

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48 (4), 812–830.e14. doi:10.1016/j.immuni.2018.03.023

Villegas, F. R., Coca, S., Villarrubia, V. G., Jimenez, R., Chillon, M. J., Jareno, J., et al. (2002). Prognostic significance of tumor infiltrating natural killer cells subset CD57 in patients with squamous cell lung cancer. *Lung Cancer* 35 (1), 23–28. doi:10.1016/s0169-5002(01)00292-6

Wang, X., Dou, X., Ren, X., Rong, Z., Sun, L., Deng, Y., et al. (2021). A ductal-cell-related risk model integrating single-cell and bulk sequencing data predicts the prognosis of patients with pancreatic adenocarcinoma. *Front. Genet.* 12, 763636. doi:10.3389/fgene.2021.763636

Wang, X., Ni, M., and Han, D. (2022). Identification of a novel risk model: A five-gene prognostic signature for pancreatic cancer. *Evid. Based Complement. Altern. Med.* 2022, 3660110. doi:10.1155/2022/3660110

Waters, A. M., and Der, C. J. (2018). Kras: The critical driver and therapeutic target for pancreatic cancer. *Cold Spring Harb. Perspect. Med.* 8 (9), a031435. doi:10.1101/cshperspect.a031435

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

Yoshida, T., Zhang, Y., Rivera Rosado, L. A., Chen, J., Khan, T., Moon, S. Y., et al. (2010). Blockade of Rac1 activity induces G1 cell cycle arrest or apoptosis in breast cancer cells through downregulation of cyclin D1, survivin, and X-linked inhibitor of apoptosis protein. *Mol. Cancer Ther.* 9 (6), 1657–1668. doi:10.1158/1535-7163.MCT-09-0906

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

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

Zhang, X., Zeng, Y., Qu, Q., Zhu, J., Liu, Z., Ning, W., et al. (2017). PD-L1 induced by IFN-γ from tumor-associated macrophages via the JAK/STAT3 and PI3K/AKT signaling pathways promoted progression of lung cancer. *Int. J. Clin. Oncol.* 22 (6), 1026–1033. doi:10.1007/s10147-017-1161-7

Zheng, J., Li, H., Xu, D., and Zhu, H. (2017). Upregulation of tyrosine kinase FYN in human thyroid carcinoma: Role in modulating tumor cell proliferation, invasion, and migration. *Cancer Biother Radiopharm.* 32 (9), 320–326. doi:10.1089/cbr.2017.2218

Zou, Y., Xie, J., Zheng, S., Liu, W., Tang, Y., Tian, W., et al. (2022). Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery. *Int. J. Surg.* 107, 106936. doi:10.1016/j.ijsu.2022.106936

Frontiers in Genetics

TYPE Original Research
PUBLISHED 27 March 2023
DOI 10.3389/fgene.2023.1106359

Transcriptional landscape of myasthenia gravis revealed by weighted gene coexpression network analysis

Demin Zhang*†, Liqin Luo, Feng Lu, Bo Li and Xiaoyun Lai

Department of Neurology, The 923rd Hospital of the Joint Logistics Support Force of the People's Liberation Army, Nanning, China

OPEN ACCESS

EDITED BY
Xiang Xue,
University of New Mexico, United States

REVIEWED BY
Sarbjeet Makkar,
Washington University in St. Louis, United States
Mi Jian,
Yantai Yuhuangding Hospital, China
Ke Mo,
YuanDong International Academy Of Life Sciences, China

*CORRESPONDENCE
Demin Zhang,
samczx@yeah.net

†These authors contributed equally to this work

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

RECEIVED 23 November 2022
ACCEPTED 14 March 2023
PUBLISHED 27 March 2023

CITATION
Zhang D, Luo L, Lu F, Li B and Lai X (2023), Transcriptional landscape of myasthenia gravis revealed by weighted gene coexpression network analysis. *Front. Genet.* 14:1106359.
doi: 10.3389/fgene.2023.1106359

Background: As one of the most common autoimmune diseases, myasthenia gravis (MG) severely affects the quality of life of patients. Therefore, exploring the role of dysregulated genes between MG and healthy controls in the diagnosis of MG is beneficial to reveal new and promising diagnostic biomarkers and clinical therapeutic targets.

Methods: The GSE85452 dataset was downloaded from the Gene Expression Omnibus (GEO) database and differential gene expression analysis was performed on MG and healthy control samples to identify differentially expressed genes (DEGs). The functions and pathways involved in DEGs were also explored by functional enrichment analysis. Significantly associated modular genes were identified by weighted gene co-expression network analysis (WGCNA), and MG dysregulated gene co-expression modular-based diagnostic models were constructed by gene set variance analysis (GSVA) and least absolute shrinkage and selection operator (LASSO). In addition, the effect of model genes on tumor immune infiltrating cells was assessed by CIBERSORT. Finally, the upstream regulators of MG dysregulated gene co-expression module were obtained by Pivot analysis.

Results: The green module with high diagnostic performance was identified by GSVA and WGCNA. The LASSO model obtained NAPB, C5orf25 and ERICH1 genes had excellent diagnostic performance for MG. Immune cell infiltration results showed a significant negative correlation between green module scores and infiltration abundance of Macrophages M2 cells.

Conclusion: In this study, a diagnostic model based on the co-expression module of MG dysregulated genes was constructed, which has good diagnostic performance and contributes to the diagnosis of MG.

KEYWORDS

myasthenia gravis, biomarkers, WGCNA, infiltrated immune cells, LASSO

Abbreviations: MG, myasthenia gravis; GEO, gene expression omnibus; DEGs, Differentially expressed genes; WGCNA, Weighted gene co-expression network analysis; GSVA, Gene set variance analysis; LASSO, Least absolute shrinkage and selection operator; AChR, acetylcholine receptor; MuSK, muscle-specific kinase; PBMCs, peripheral blood mononuclear cells; GO, Gene Ontology; KEGG, kyoto encyclopedia of genes and genomes; GSEA, gene set enrichment analysis; MsigDB, molecular signature database; ROC, recipient operating characteristic; AUC, area under the curve; RBPs, RNA binding proteins; PCA, principal component analysis; NAPB, N-ethylmaleimide-sensitive accessory protein beta; MS, multiple sclerosis; NMO, neuromyelitis optica; GBS, guillain-barré syndrome.

Introduction

As an autoimmune disease, myasthenia gravis (MG) manifests primarily as fluctuating muscle weakness caused by autoantibodies and cell-mediated disruption of acetylcholine receptors (Nations et al., 1999). It is characterized by dysfunctional transmission of the neuromuscular junction, resulting in muscle weakness (Meriggioli and Sanders, 2009). MG reduces the quality of life of patients and can be life-threatening in severe cases (Phillips, 2004; Andersen et al., 2014; Bettini et al., 2017). The prevalence of MG is estimated to be 0.3–2.8/100,000, with a global prevalence of 700,000, and the current mortality rate of MG is 5%–9% (Alshekhlee et al., 2009; Carr et al., 2010; Deenen et al., 2015). In recent years, many advances have been made in the treatment of MG, and more evidence-based medical evidence has been accumulated, which has significantly improved the prognosis of the vast majority of patients and enabled the effective control of a small number of refractory MG cases (Batocchi et al., 2000; Vincent and Drachman, 2002; Rowin et al., 2004; Grob et al., 2008; Muscle Study, 2008). However, the clinical manifestations of MG are highly heterogeneous (Rodolico et al., 2002; Grob et al., 2008). Identifying potential biomarkers of MG will help in the diagnosis and treatment of MG.

Currently, serological tests for autoantibodies are commonly used for the diagnosis and disease classification of MG patients (Gilhus and Verschuuren, 2015). About 85% of MG patients have antibodies against the muscle acetylcholine receptor (AChR) (Higuchi et al., 2011). In addition, antibodies against muscle-specific kinase (MuSK) were found in about 6% of patients (Berrih-Aknin et al., 2014), and antibodies against LRP4 were found in about 2% of MG patients (Berrih-Aknin et al., 2014). The pathogenicity of all these autoantibodies has been demonstrated by animal studies (Mori et al., 2012; Shen et al., 2013). However, the pathogenicity of these disease biomarkers is usually uncertain. There is still a need to identify new biomarkers to complement existing diagnostic tools.

Therefore, in this study, we identified dysregulated genes in MG patients and performed a weighted gene co-expression network analysis (WGCNA) on these genes. In addition, we further developed a clinical diagnostic model based on dysregulated genes and revealed the relationship between this clinical diagnostic model and the multi-omics landscape of immunological features and global regulatory networks.

Materials and methods

Data resources

In this study, the MG dataset GSE85452 (Mamrut et al., 2017) was downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The GSE85452 dataset is based on the GPL10558 platform and contains the mRNA expression profiles of 13 MG and 12 healthy control PBMCs.

Differential gene expression analysis

To identify differentially expressed genes (DEGs) between control and MG samples, differential gene expression analysis was performed using the Bioinforcloud application DEbylimma, which was developed based on the limma package (Ritchie et al., 2015). Among the differences, those associated with $p < 0.01$ and $|\mathrm{logFC}| > 0$ were considered significant. Subsequently, heat maps were drawn using the Bioinforcloud application PlotHeatmap to further demonstrate the expression of DEGs between samples.

Weighted gene co-expression network analysis

The weighted gene co-expression network analysis (WGCNA) application in Bioinforcloud was based on the WGCNA package in the R language (Langfelder and Horvath, 2008) being used to perform WGCNA on DEGs. Candidate powers (Nations et al., 1999; Batocchi et al., 2000; Rodolico et al., 2002; Vincent and Drachman, 2002; Phillips, 2004; Rowin et al., 2004; Grob et al., 2008; Muscle Study, 2008; Alshekhlee et al., 2009; Meriggioli and Sanders, 2009; Carr et al., 2010; Higuchi et al., 2011; Mori et al., 2012; Shen et al., 2013; Andersen et al., 2014; Berrih-Aknin et al., 2014; Deenen et al., 2015; Gilhus and Verschuuren, 2015; Bettini et al., 2017; Mamrut et al., 2017) were used to test the average connectivity of the different modules and their independence. Powers were selected if the degree of independence was >0.85. The samples were clustered by using the hclust function of the WGCNA package and checked for outliers. Subsequently, a heat map of module-phenotype correlations was constructed to find module-phenotype correlations and their significance. A high correlation means that the genes of the corresponding module also tend to be highly correlated with the disease state.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of MG dysregulated gene co-expression module genes using the Bioinforcloud application RunMutiGroupclusterProfiler. The application was developed based on the clusterProfiler package in the R language (Yu et al., 2012), and the enriched functions or pathways were considered significant when $p<$ was 0.05.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using the Bioinforcloud application RunGSEA to further explore the potential biological properties of MG dysregulated gene co-expression modules. The application uses the Molecular Signature Database (MsigDB) (Liberzon et al., 2015) of c2. cp.kegg.v7.0. symbols.gmt as the reference gene set, and was developed based on the clusterProfiler package in the R language (Yu et al., 2012), the enrichment results at $p < 0.05$ were considered significant.

Gene set variation analysis

Gene set variation analysis (GSVA) of modular genes using RunGSVA. The application is based on the GSVA package (Hanzelmann et al., 2013) for calculating GSVA scores of MG dysregulated gene co-expression module genes in different samples. Subsequently, heat maps were drawn using the Bioinforcloud application PlotHeatmap to further demonstrate the expression of GSVA scores across samples.

Assessment of diagnostic efficacy

Evaluation of diagnostic efficacy of potential markers using the Bioinforcloud application PlotROC. The application is based on the pROC package in R (Robin et al., 2011) and the results were plotted as receptor operating characteristic (ROC) curves. In this study, the potential of GSVA scores of MG dysregulated gene co-expression module genes as a diagnostic marker for MG was evaluated using this application. In the case of area under the curve (AUC) > 0.5, the closer the AUC is to 1, the better the diagnosis.

Construction of minimum absolute shrinkage and selection operator models

The least absolute shrinkage and selection operator (LASSO) has a strong predictive value and low correlation and is suitable for selecting the best features for high-dimensional data. LASSO regression analysis was performed using the Bioinforcloud application RunLASSO, which was developed based on the glmnet software package (Friedman et al., 2010) and extracted the expression profiles of MG dysregulated genes and co-expression module functional genes with diagnostic efficacy to construct the LASSO model. The expression values of the selected genes were weighted using the regression coefficients of the LASSO analysis to create a model index for each sample with the following equation: Index = ExpGene1*Coef1 + ExpGene2*Coef2 + ExpGene3*Coef3 +.

“Coef” is the regression coefficient of the gene, derived from LASSO Cox regression, and “Exp” indicates the expression value of the gene, thus constructing the MG dysregulated gene co-expression module-based Lasso model.

Immune cell infiltration analysis

In this study, immune cell infiltration analysis was performed using the Bioinforcloud application RunCIBERSORT to assess the abundance of immune cell infiltration in MG as well as between control samples. The application is based on the CIBERSORT tool (Chen et al., 2018). It was developed to enable the estimation of immune infiltration for large volumes of transcripts and thus assess the relationship between gene expression or other phenotypes and immune cell infiltration. In addition, correlation analysis was performed using the Bioinforcloud application PlotCor to explore the correlation between MG dysregulated gene co-expression module-based models, model genes and the abundance of immune cell infiltration, immune checkpoint genes and tertiary lymphoid structural marker genes.

Identification of upstream regulators

In this study, the differentially expressed RNA binding proteins (RBPs) were screened in combination with the results of differential gene expression analysis, and subsequently, the upstream regulators regulating the gene sets of the MG dysregulated gene co-expression module-based model were identified using the Bioinforcloud application Pivot. The application is based on a hypergeometric approach to implement Pivot analysis to identify RBPs in the regulatory gene set.

Data analysis and statistics

All statistical analyses were performed in the Bioinforcloud platform (http://www.bioinforcloud.org.cn), which was applied by calling the appropriate R package. Comparisons between the two groups were made using Student’s t-test and correlation coefficients were calculated using Spearman analysis. $p < 0.05$ was considered significant.

Results

Dysregulated gene co-expression modules characterize the global regulatory pattern of myasthenia gravis

As shown in Figure 1A, gene expression data from PBMC samples of 13 MG patients and 12 healthy controls were analyzed in this study. The bias of sequencing data due to gene length, sequencing volume and other factors was removed by normalization. Principal component analysis (PCA) scatter plots showed good discrimination between different samples (Figure 2A). DEGs between MG and control were identified by differential gene expression analysis, including 861 upregulated DEGs and 643 downregulated DEGs (Figure 2B, Supplementary Table S1). The heat map showed that these dysregulated genes could significantly distinguish between MG and control samples (Figure 2C).

Subsequently, to further explore the relationship between these dysregulated genes and MG, this study screened these genes for WGCNA. to construct a scale-free network, we set the soft threshold power β to 16, and DEGs with similar expression patterns were co-classified into six co-expression modules. The module ring tree diagram demonstrates the neighbor-joining relationships among the dysregulated gene co-expression modules (Figure 2D). In addition, the module heat map further demonstrates the co-expression of MG dysregulated genes in different modules (Figure 2E). The expression correlations of some significantly dysregulated genes were demonstrated in the module correlation plots (Figure 2F), which may be closely related to MG development.

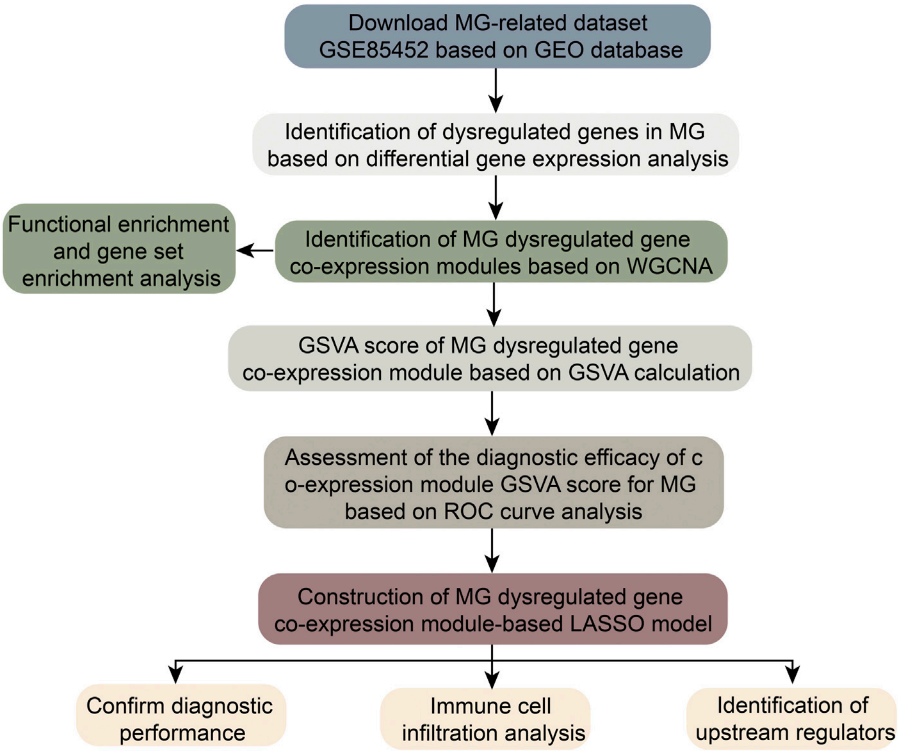


FIGURE 1
Flowchart of this work.

Biological functions and signaling pathways significantly involved in myasthenia gravis dysregulated gene co-expression modules

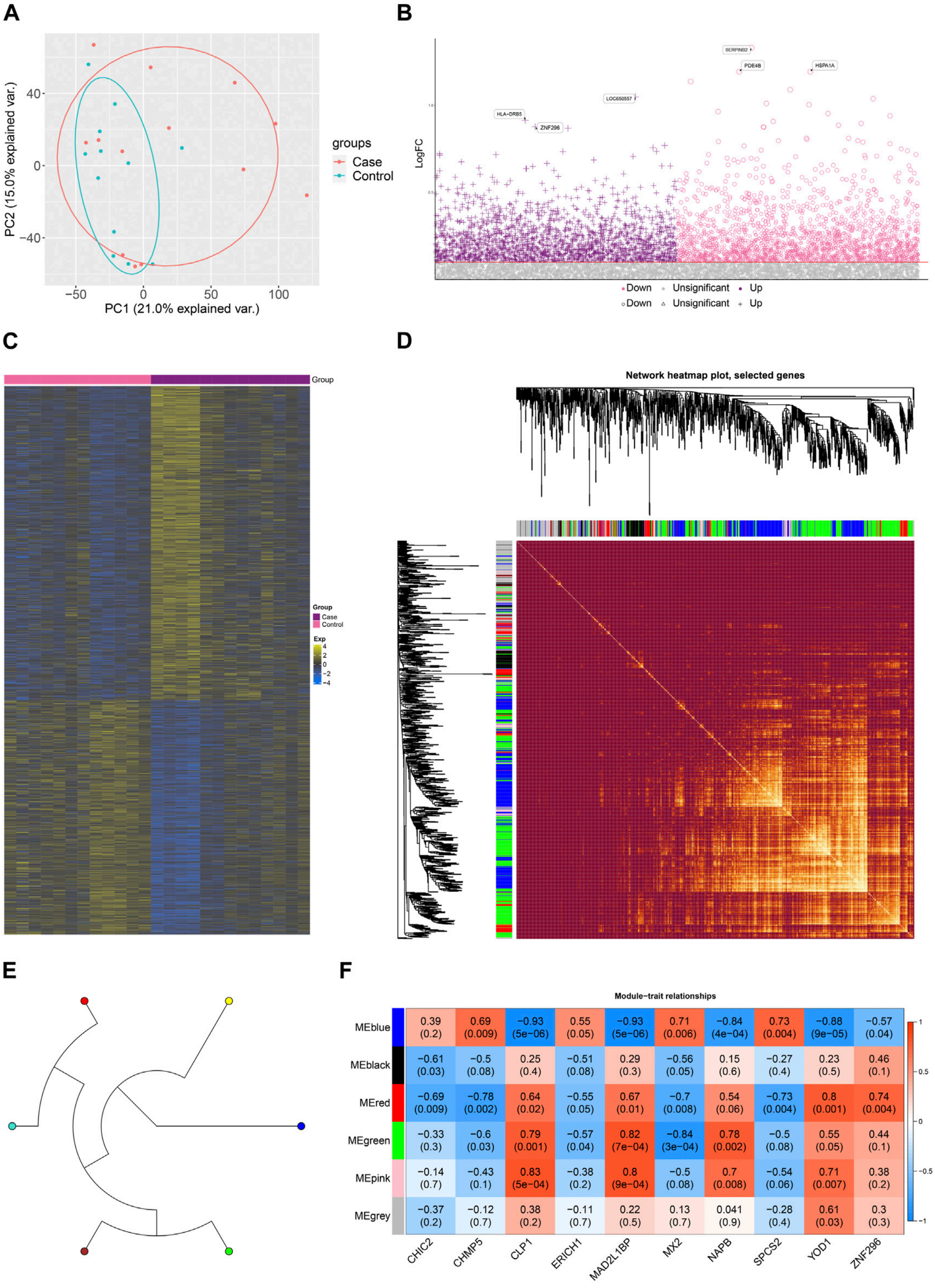
To further investigate the biological functions and signaling pathways significantly involved in MG dysregulated gene co-expression modules, enrichment analysis of these genes was performed. The results showed that these genes are significantly involved in the biological processes of positive regulation of innate immune response, regulation of interferon-beta production and positive regulation of cytokine production and KEGG pathways such as neurodegeneration-multiple diseases pathway, Th1 and Th2 cell differentiation, TGF -beta signaling pathway and mTOR signaling pathway. (Figures 3A, B). In addition, GSEA further confirmed the activation or inhibition of KEGG signaling pathways in different co-expression modules (Figure 3C), suggesting that these pathways may play an important role in the development of MG. In addition to this, we compared the scores of ferroptosis and necroptosis in MG and control samples and found that the scores of necroptosis were higher in the MG group, while the scores of ferroptosis were not significant (Figure 3D).

Myasthenia gravis dysregulated gene co-expression module-based clinical model has significant diagnostic efficacy

The GSVA scores of MG dysregulated gene co-expression modules were calculated based on the GSVA method (Figure 4A), and the diagnostic efficacy of GSVA scores of different modules for MG was identified using ROC analysis. The results showed that the Green module had the best diagnostic efficacy for MG (AUC = 0.584, Supplementary Figure S1). Subsequently, the MG dysregulated gene co-expression module-based clinical model was further constructed using the LASSO method, and three characteristic genes with non-zero regression coefficients were obtained (lambda.min = 0.110, Figures 4B,C). ROC curve analysis showed that the MG dysregulated gene co-expression module-based model showed excellent diagnostic efficacy for MG (AUC = 0.981, Figure 4D), and some of these genes, such as NAPB, showed significantly high expression in MG, while C5orf25 and ERICH1 showed significantly low expression in MG (Figure 4E). In addition, the correlation between NAPB, C5orf25 and ERICH1 genes and ferroptosis and necroptosis was analyzed and the results were shown in Supplementary Table S2, ERICH1 was negatively associated with necroptosis.

Co-expression module reprograms the immune microenvironment of myasthenia gravis

By immune infiltration analysis, this study explored the level of immune cell infiltration in Control and MG samples. Correlation analysis showed a significant negative correlation between the Green module score and the infiltration abundance of Macrophages M2 cells, suggesting that the expression of these module genes may inhibit the infiltration of the corresponding immune cells



	CHIC2	CHMP5	CLP1	ERICH1	MAD2L1BP	MX2	NAPB	SPCS2	YOD1	ZNF296
MEblue	0.39 (0.2)	0.69 (0.009)	−0.93 (5e−06)	0.55 (0.05)	−0.93 (5e−06)	0.71 (0.006)	−0.84 (4e−04)	0.73 (0.004)	−0.88 (9e−05)	−0.57 (0.04)
MEblack	−0.61 (0.03)	−0.5 (0.08)	0.25 (0.4)	−0.51 (0.08)	0.29 (0.3)	−0.56 (0.05)	0.15 (0.6)	−0.27 (0.4)	0.23 (0.5)	0.46 (0.1)
MEred	−0.69 (0.009)	−0.78 (0.002)	0.64 (0.02)	−0.55 (0.05)	0.67 (0.01)	−0.7 (0.008)	0.54 (0.06)	−0.73 (0.004)	0.8 (0.001)	0.74 (0.004)
MEgreen	−0.33 (0.3)	−0.6 (0.03)	0.79 (0.001)	−0.57 (0.04)	0.82 (7e−04)	−0.84 (3e−04)	0.78 (0.002)	−0.5 (0.08)	0.55 (0.05)	0.44 (0.1)
MEpink	−0.14 (0.7)	−0.43 (0.1)	0.83 (5e−04)	−0.38 (0.2)	0.8 (9e−04)	−0.5 (0.08)	0.7 (0.008)	−0.54 (0.06)	0.71 (0.007)	0.38 (0.2)
MEgrey	−0.37 (0.2)	−0.12 (0.7)	0.38 (0.2)	−0.11 (0.7)	0.22 (0.5)	0.13 (0.7)	0.041 (0.9)	−0.28 (0.4)	0.61 (0.03)	0.3 (0.3)

FIGURE 2
Dysregulated gene co-expression modules characterize the global regulatory pattern of myasthenia gravis. **(A)**. Principal component analysis (PCA) plots showing significant differences between disease and control. **(B)** Manhattan plots showing differential expression of Case-Control. **(C)** Heat map showing expression of dysregulated genes in Case-Control groupings. **(D)**. Module ring tree plots showing neighboring relationships between dysregulated gene co-expression modules. **(E)** Module heat map showing gene members of dysregulated gene co-expression modules. **(F)** Module correlation plot showing the expression correlation of gene co-expression modules with Top10 significantly dysregulated genes.

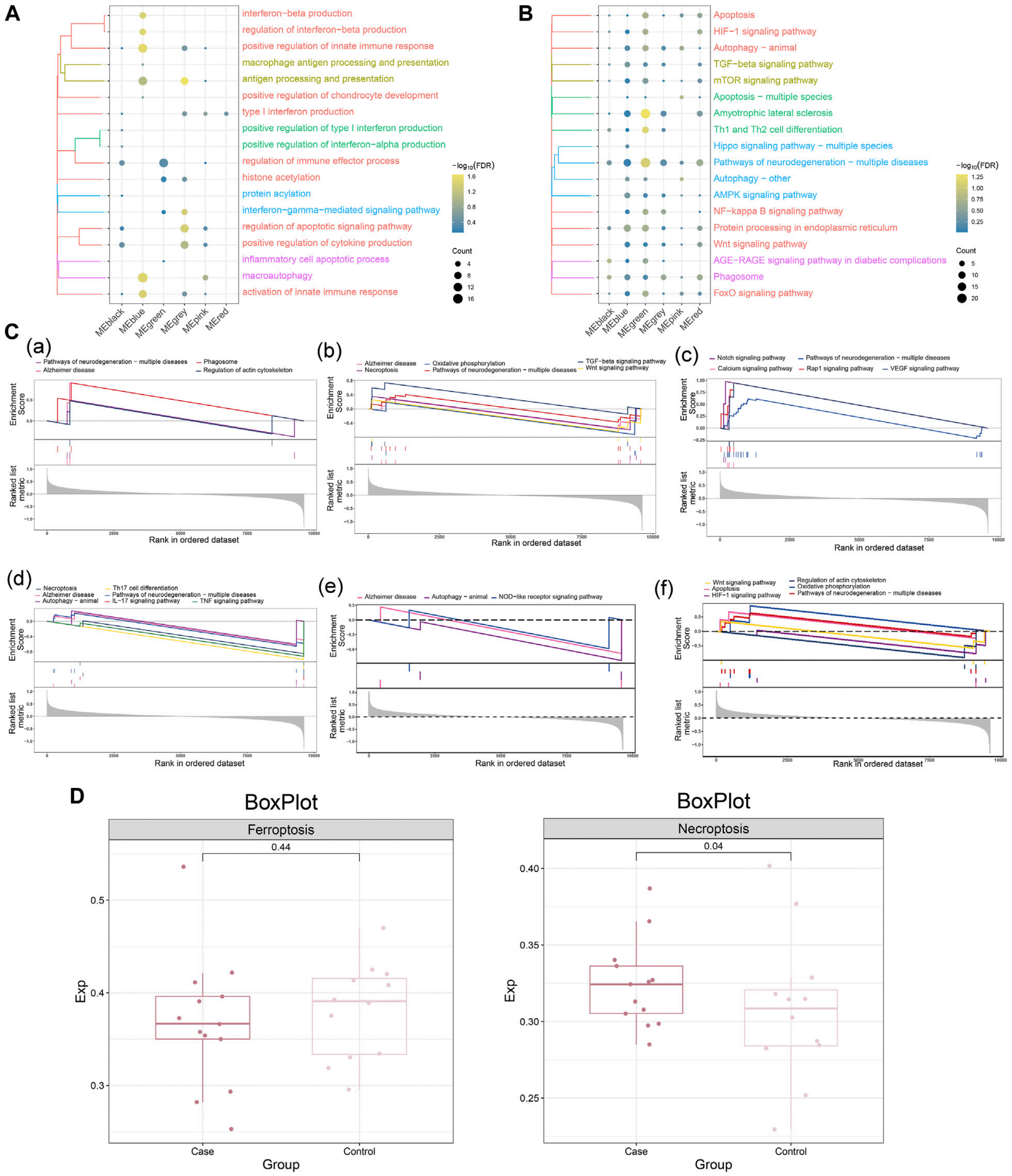


FIGURE 3
Biological functions and signaling pathways significantly involved in myasthenia gravis dysregulated gene co-expression modules. **(A)** Clustered bubble plots showing the biological functions significantly regulated by MG dysregulated gene co-expression modules (listed as different modules). **(B)** Clustered bubble plots showing the signaling pathways significantly regulated by MG dysregulated gene co-expression modules (listed as different modules). **(C)** Comprehensive GSEA diagram showing signaling pathways significantly activated/repressed by MG dysregulated gene co-expression modules, **(a)** black module, **(b)** blue module, **(c)** green module, **(d)** gray module, **(e)** pink module, **(f)** red module. **(D)** Scoring of ferroptosis and necroptosis in MG and control samples.

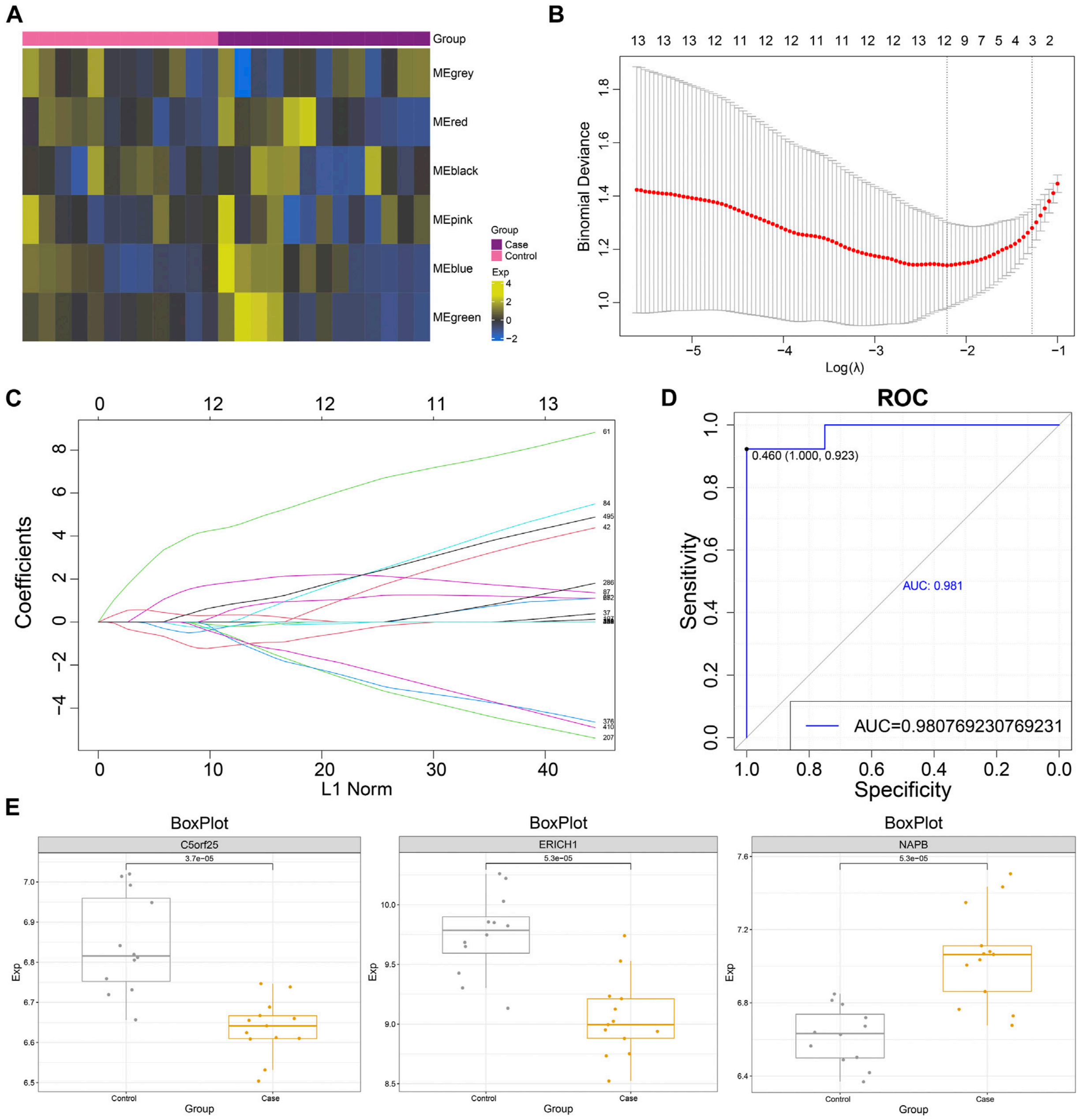


FIGURE 4
The myasthenia gravis dysregulated gene co-expression module-based clinical model has significant diagnostic efficacy. **(A)** Heat map demonstrating significant enrichment of MG dysregulated gene co-expression module gene (GSVA) scores in Case-Control. **(B)** Lambda plot demonstrating model performance of the set of MG dysregulated gene co-expression module functional genes with diagnostic efficacy at different Lambda (single factor significant genes were selected for Lasso modeling). **(C)** Lasso model Plot showing the model confidence of the set of functional genes with diagnostic efficacy of MG dysregulated gene co-expression module at different log (Lambda). **(D)** ROC curve showing the ROC curve of MG dysregulated gene co-expression module-based model. **(E)** Box plot showing the expression level of model genes.

(Figure 5A). Notably, MG dysregulated gene co-expression module scores as well as scored genes showed significant correlation with the abundance of some immune cells, immune checkpoint genes and tertiary lymphoid structural marker genes (Figures 5B, C), suggesting that these module genes may be indirectly involved in reprogramming the MG immune microenvironment by promoting the infiltration of immune cells, or regulating the expression of immune-related genes.

Upstream regulators of dysregulated gene co-expression modules

To construct a global regulatory network for the MG dysregulated gene co-expression module-based model, we further explored the upstream regulators of these genes. The upstream regulators regulating this model gene set, including RBPs such as YTHDF1, U2AF2, TARDBP, STAU1, were identified by Pivot

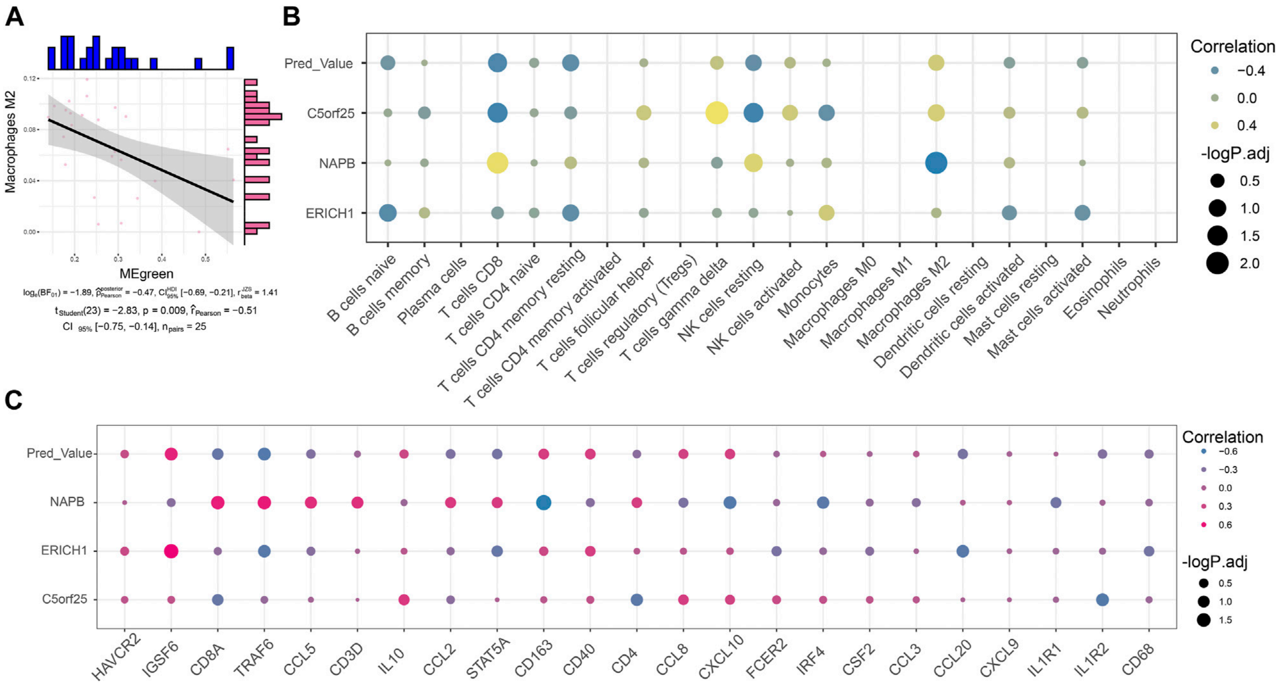


FIGURE 5
Co-expression modules reprogram the immune microenvironment of myasthenia gravis. **(A)** Series of correlation scatter plots showing expression correlation of immune cell infiltration abundance with diagnostic potency of MG dysregulated gene co-expression module scores. **(B)** Bubble plots showing correlation of MG dysregulated gene co-expression module-based Lasso model genes with immune cell infiltration abundance. **(C)** Bubble plots showing correlation of MG dysregulated gene co-expression module-based Lasso model genes correlated with immune checkpoint-associated genes and tertiary lymphoid structural marker genes.

analysis (Figure 6A), where EIF3D, RBM15, STAU1, TARDBP and YTHDF1 showed significant high expression in MG samples (Figure 6B).

Discussion

The clear pathogenesis of MG is still unknown, and due to its heterogeneous and complex pathogenesis, there are no effective treatment options for MG patients (Sanders et al., 2016). The present study is based on the mRNA expression profile of MG from the GEO database. In this study, we identified MG-related DEGs for WGCNA based on the mRNA expression profile of MG in the GEO database, and searched for the most relevant modules to construct a scoring clinical model, which can adequately ensure the interaction between genes. Then, the diagnostic efficacy of the scoring clinical model for MG was determined. In addition, few established immune-related gene profiles were combined with conventional prognostic models to optimize routine clinical practice. They are not very effective as a direct guide to clinical workup. To remedy these shortcomings, we further explored the immune microenvironment of MG based on clinical models of MG scoring. These findings strongly suggest a great potential role for the MG dysregulated gene co-expression module-based model in MG obtained in this study.

WGCNA makes strongly correlated genes strongly correlated after power function treatment, therefore, the construction of WGCNA network helps to identify and screen important modules and key genes associated with specific clinical phenotypes (Langfelder and Horvath, 2008). In this study, WGCNA analysis was performed on RNA-seq datasets downloaded from the GEO database, and DEGs were calculated separately between MG patients and healthy controls, yielding a total of 1504 DEGs as the dataset for subsequent co-expression network analysis to prevent high correlations for genes that were not significantly different. Notably, the co-expression network analysis identified and clustered into six co-expression modules, and the correlation analysis of genes with significantly dysregulated genes was performed for each module, and the Top10 pivotal genes contained in each module were screened for strong interaction with MG, respectively.

Among the different modules, genes were found to be mainly enriched in Pathways of neurodegeneration - multiple diseases, Alzheimer disease, Th1 and Th2 cell differentiation, Regulation of actin cytoskeleton, Oxidative phosphorylation, Necroptosis, TGF-beta signaling pathway, Wnt signaling pathway and other KEGG signaling pathways. Notably, each module was significantly enriched in Pathways of neurodegeneration - multiple diseases. In addition, functional enrichment analysis revealed that module genes were mainly enriched in positive regulation of innate immune response, regulation of interferon-

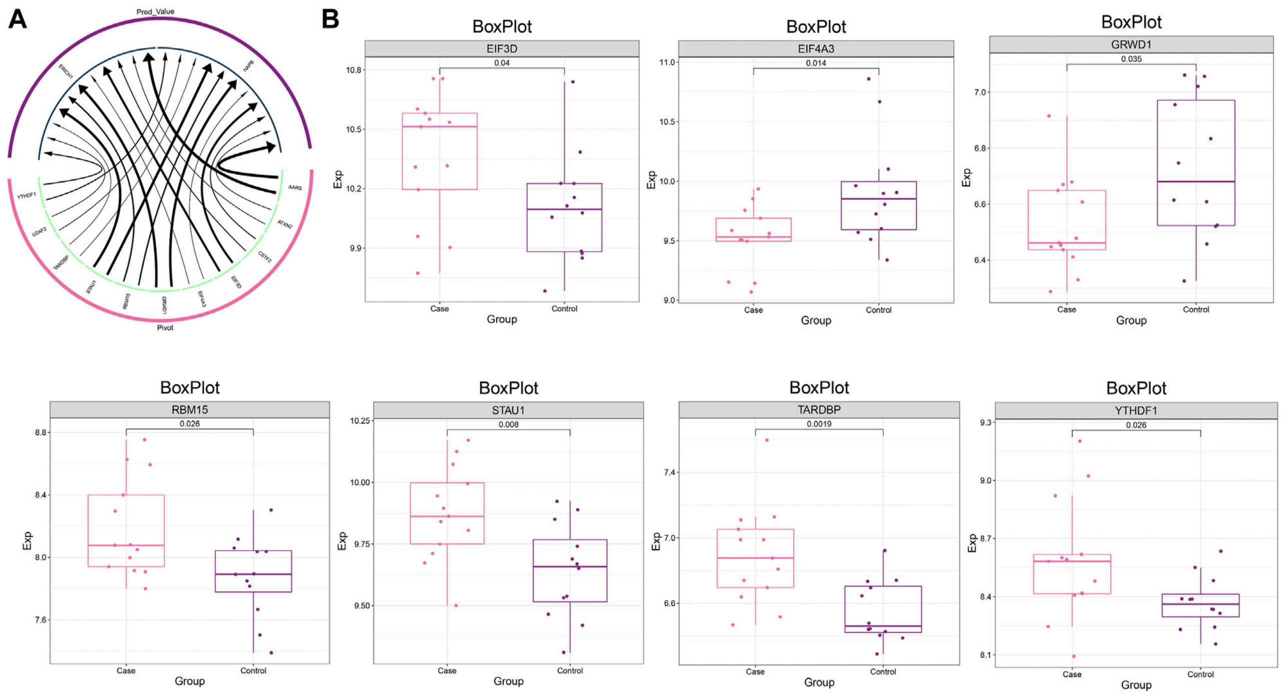


FIGURE 6
Upstream regulators of the dysregulated gene co-expression module. **(A)** Circular network plot demonstrating the regulatory effect of RBP on MG dysregulated gene co-expression module-based Lasso model genes. **(B)** Box plot demonstrating the expression level of RBPs.

beta production, positive regulation of cytokine production, activation of innate immune response This suggests that the main cause of MG development is abnormalities in interferon and immune pathways. In the course of MG, abnormal antigen processing and presentation may contribute to the onset and progression of the disease (Xu et al., 2021). In addition, it has been found that imbalance of various helper T cells (including Th1, Th2, Th17, Th22 and follicular helper T (TFH) cells in MG is associated with immune disorders, suggesting that the balance of Th cells and their cytokines in MG patients is related to the clinical status or severity of MG disease (Wang et al., 2019). It has also been shown that oxidative stress and low antioxidant status play a major role in the pathogenesis of inflammatory and autoimmune diseases, and that MG patients with low antioxidant status have active oxidative processes (Yang et al., 2016; Adamczyk-Sowa et al., 2017). In addition to this, studies have confirmed that AChR-MG may be an acquired interferon disease (Payet et al., 2022). The results of GO and KEGG analysis in this study also suggest that MG dysregulated genes are mainly enriched in interferon and immune-related processes.

To date, there are no studies on the NAPB, C5orf25, and ERICH1 genes in PBMCs in MG. The N-ethylmaleimide-sensitive accessory protein beta (NAPB) gene is associated with brain development as well as brain development in neurological disorders, such as various severe early onset epilepsy (Conroy et al., 2016; Zhao et al., 2021). In addition, NAPB has been shown to act as a pivotal gene in Alzheimer's disease and to be involved in the pathogenesis of Alzheimer's disease (Zhang et al., 2020). CAPN3 has been reported to have multiple muscle cell functions and mutations in this protease cause limb-girdle muscular dystrophy type 2A (Ono et al., 2013). C5orf25 is a novel CAPN3-binding protein that regulates the protease activity of CAPN3 and has the potential to act as a scaffolding protein (Ono et al., 2013). While ERICH1 has been reported to be associated with the risk of multiple sclerosis (MS) (Maltby et al., 2017), MS and MG are two uncommon neurological problems, both of which can affect the nervous system.

Immune cell infiltration analysis showed a significant negative correlation between the infiltration abundance of green module Macrophages M2 cells. Macrophages the cause of the pathogenesis of some human neuroimmune diseases, mainly MS, optic neuromyelitis optica (NMO), MG and Guillain-Barré syndrome (GBS) (Fan et al., 2016).

However, the present study still has limitations, as the sample size of the public database is too small, which may lead to the omission of pivotal genes. In addition, the results of this analysis were obtained exclusively by bioinformatics and failed to experimentally validate the expression of the obtained biomarkers at protein and RNA levels, and further experimental validation is proposed in the future.

Conclusion

The results of this study showed that a gene-based clinical model consisting of NAPB, C5orf25 and ERICH1 showed high diagnostic ability for MG (AUC = 0.981), and this model developed can be used as a diagnostic indicator for MG, which is crucial for subsequent clinical treatment and improvement of disease prognosis.

Data availability statement

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

Author contributions

DZ designed and directed the research. LL and FL contributed to the analysis and interpretation of data and wrote the main manuscript text. BL and XL performed analyzed results. All authors read and approved the final manuscript.

Funding

This work was supported by self-financed scientific research project of the Guangxi Zhuang Autonomous Region Health Department (Z2014364).

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.2023.1106359/full#supplementary-material

References

Adamczyk-Sowa, M., Bieszczad-Bedrejczuk, E., Galiniak, S., Rozmilowska, I., Czyzewski, D., Bartosz, G., et al. (2017). Oxidative modifications of blood serum proteins in myasthenia gravis. *J. Neuroimmunol.* 305, 145–153. doi:10.1016/j.jneuroim.2017.01.019

Alshekhlee, A., Miles, J. D., Katirji, B., Preston, D. C., and Kaminski, H. J. (2009). Incidence and mortality rates of myasthenia gravis and myasthenic crisis in US hospitals. *Neurology* 72 (18), 1548–1554. doi:10.1212/WNL.0b013e3181a41211

Andersen, J. B., Heldal, A. T., Engeland, A., and Gilhus, N. E. (2014). Myasthenia gravis epidemiology in a national cohort; combining multiple disease registries. *Acta Neurol. Scand. Suppl.* 129 (198), 26–31. doi:10.1111/ane.12233

Batocchi, A. P., Evoli, A., Di Schino, C., and Tonali, P. (2000). Therapeutic apheresis in myasthenia gravis. *Ther. Apher.* 4 (4), 275–279. doi:10.1046/j.1526-0968.2000.004004275.x

Berrih-Aknin, S., Frenkian-Cuvelier, M., and Eymard, B. (2014). Diagnostic and clinical classification of autoimmune myasthenia gravis. *J. Autoimmun.* 48-49, 143–148. doi:10.1016/j.jaut.2014.01.003

Bettini, M., Chaves, M., Cristiano, E., Pagotto, V., Perez, L., Giunta, D., et al. (2017). Incidence of autoimmune myasthenia gravis in a Health maintenance organization in buenos aires, Argentina. *Neuroepidemiology* 48 (3-4), 119–123. doi:10.1159/000477733

Carr, A. S., Cardwell, C. R., McCarron, P. O., and McConville, J. (2010). A systematic review of population based epidemiological studies in Myasthenia Gravis. *BMC Neurol.* 10, 46. doi:10.1186/1471-2377-10-46

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

Conroy, J., Allen, N. M., Gorman, K. M., Shahwan, A., Ennis, S., Lynch, S. A., et al. (2016). Napb - a novel SNARE-associated protein for early-onset epileptic encephalopathy. *Clin. Genet.* 89 (2), E1–E3. doi:10.1111/cge.12648

Deenen, J. C., Horlings, C. G., Verschuuren, J. J., Verbeek, A. L., and van Engelen, B. G. (2015). The epidemiology of neuromuscular disorders: A comprehensive overview of the literature. *J. Neuromuscul. Dis.* 2 (1), 73–85. doi:10.3233/jnd-140045

Fan, X., Zhang, H., Cheng, Y., Jiang, X., Zhu, J., and Jin, T. (2016). Double roles of Macrophages in human neuroimmune diseases and their animal models. *Mediat. Inflamm.* 2016, 8489251. doi:10.1155/2016/8489251

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Gilhus, N. E., and Verschuuren, J. J. (2015). Myasthenia gravis: Subgroup classification and therapeutic strategies. *Lancet Neurol.* 14 (10), 1023–1036. doi:10.1016/S1474-4422(15)00145-3

Grob, D., Brunner, N., Namba, T., and Pagala, M. (2008). Lifetime course of myasthenia gravis. *Muscle Nerve* 37 (2), 141–149. doi:10.1002/mus.20950

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

Higuchi, O., Hamuro, J., Motomura, M., and Yamanashi, Y. (2011). Autoantibodies to low-density lipoprotein receptor-related protein 4 in myasthenia gravis. *Ann. Neurol.* 69 (2), 418–422. doi:10.1002/ana.22312

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

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Maltby, V. E., Lea, R. A., Sanders, K. A., White, N., Benton, M. C., Scott, R. J., et al. (2017). Differential methylation at MHC in CD4(+) T cells is associated with multiple sclerosis independently of HLA-DRB1. *Clin. Epigenetics* 9, 71. doi:10.1186/s13148-017-0371-1

Mamrut, S., Avidan, N., Truffault, F., Staun-Ram, E., Sharshar, T., Eymard, B., et al. (2017). Methylome and transcriptome profiling in Myasthenia Gravis monozygotic twins. *J. Autoimmun.* 82, 62–73. doi:10.1016/j.jaut.2017.05.005

Meriggioli, M. N., and Sanders, D. B. (2009). Autoimmune myasthenia gravis: Emerging clinical and biological heterogeneity. *Lancet Neurol.* 8 (5), 475–490. doi:10.1016/S1474-4422(09)70063-8

Mori, S., Kubo, S., Akiyoshi, T., Yamada, S., Miyazaki, T., Hotta, H., et al. (2012). Antibodies against muscle-specific kinase impair both presynaptic and postsynaptic functions in a murine model of myasthenia gravis. *Am. J. Pathol.* 180 (2), 798–810. doi:10.1016/j.ajpath.2011.10.031

Muscle Study, G. (2008). A trial of mycophenolate mofetil with prednisone as initial immunotherapy in myasthenia gravis. *Neurology* 71 (6), 394–399. doi:10.1212/01.wnl.0000312373.67493.7f

Nations, S. P., Wolfe, G. I., Amato, A. A., Jackson, C. E., Bryan, W. W., and Barohn, R. J. (1999). Distal myasthenia gravis. *Neurology* 52 (3), 632–634. doi:10.1212/wnl.52.3.632

Ono, Y., Iemura, S., Novak, S. M., Doi, N., Kitamura, F., Natsume, T., et al. (2013). PLEIAD/SIMC1/C5orf25, a novel autolysis regulator for a skeletal-muscle-specific calpain, CAPN3, scaffolds a CAPN3 substrate, CTBP1. *J. Mol. Biol.* 425 (16), 2955–2972. doi:10.1016/j.jmb.2013.05.009

Payet, C. A., You, A., Fayet, O. M., Dragin, N., Berrih-Aknin, S., and Le Panse, R. (2022). Myasthenia gravis: An acquired interferonopathy? Cells. *Cells* 11 (7), 1218. doi:10.3390/cells11071218

Phillips, L. H. (2004). The epidemiology of myasthenia gravis. *Semin. Neurol.* 24 (1), 17–20. doi:10.1055/s-2004-829593

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

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

Rodolico, C., Toscano, A., Autunno, M., Messina, S., Nicolosi, C., Aguennouz, M., et al. (2002). Limb-girdle myasthenia: Clinical, electrophysiological and morphological features in familial and autoimmune cases. *Neuromuscul. Disord.* 12 (10), 964–969. doi:10.1016/s0960-8966(02)00137-2

Rowin, J., Meriggioli, M. N., Tuzun, E., Leurgans, S., and Christadoss, P. (2004). Etanercept treatment in corticosteroid-dependent myasthenia gravis. *Neurology* 63 (12), 2390–2392. doi:10.1212/01.wnl.0000147242.92691.71

Sanders, D. B., Wolfe, G. I., Benatar, M., Evoli, A., Gilhus, N. E., Illa, I., et al. (2016). International consensus guidance for management of myasthenia gravis: Executive summary. *Neurology* 87 (4), 419–425. doi:10.1212/WNL.0000000000002790

Shen, C., Lu, Y., Zhang, B., Figueiredo, D., Bean, J., Jung, J., et al. (2013). Antibodies against low-density lipoprotein receptor-related protein 4 induce myasthenia gravis. *J. Clin. Invest.* 123 (12), 5190–5202. doi:10.1172/JCI66039

Vincent, A., and Drachman, D. B. (2002). Myasthenia gravis. *Adv. Neurol.* 88, 159–188.

Wang, L., Zhang, Y., Zhu, M., Feng, J., Han, J., Zhu, J., et al. (2019). Effects of follicular helper T cells and inflammatory cytokines on myasthenia gravis. *Curr. Mol. Med.* 19 (10), 739–745. doi:10.2174/1566524019666190827162615

Xu, S., Wang, T., Lu, X., Zhang, H., Liu, L., Kong, X., et al. (2021). Identification of LINC00173 in myasthenia gravis by integration analysis of aberrantly methylated-differentially expressed genes and ceRNA networks. *Front. Genet.* 12, 726751. doi:10.3389/fgene.2021.726751

Yang, D., Su, Z., Wu, S., Bi, Y., Li, X., Li, J., et al. (2016). Low antioxidant status of serum bilirubin, uric acid, albumin and creatinine in patients with myasthenia gravis. *Int. J. Neurosci.* 126 (12), 1120–1126. doi:10.3109/00207454.2015.1134526

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

Zhang, F., Zhong, S. R., Yang, S. M., Wei, Y. T., Wang, J. J., Huang, J. L., et al. (2020). Identification of potential therapeutic targets of Alzheimer's disease by weighted gene Co-expression network analysis. *Chin. Med. Sci. J.* 35 (4), 330–341. doi:10.24920/003695

Zhao, X., Wang, Y., Cai, A., Mei, S., Liu, N., and Kong, X. (2021). A novel NAPB splicing mutation identified by Trio-based exome sequencing is associated with early-onset epileptic encephalopathy. *Eur. J. Med. Genet.* 64 (1), 104101. doi:10.1016/j.ejmg.2020.104101

Frontiers in Genetics

TYPE Original Research
PUBLISHED 07 April 2023
DOI 10.3389/fgene.2023.1110396

OPEN ACCESS

EDITED BY
Xiang Xue,
University of New Mexico, United States

REVIEWED BY
Xueqiong Han,
The Fifth Affiliated Hospital of Guangxi Medical University, China
Jin Zhou,
Qingdao University, China

*CORRESPONDENCE
Jianxian Chen,
✉ chjianxian@126.com

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

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

RECEIVED 28 November 2022
ACCEPTED 28 March 2023
PUBLISHED 07 April 2023

CITATION
Chen J, Meng Y, Huang X, Liao X, Tang X, Xu Y and Li J (2023), Potential effective diagnostic biomarker in patients with primary and metastatic small intestinal neuroendocrine tumors.
Front. Genet. 14:1110396.
doi: 10.3389/fgene.2023.1110396

Potential effective diagnostic biomarker in patients with primary and metastatic small intestinal neuroendocrine tumors

Jianxian Chen*[†], Yiliang Meng, Xiaojuan Huang, Xuegan Liao, Xiaochun Tang, Yuanchao Xu and Jie Li

Department of Oncology, The People's Hospital of Baise, Baise, Guangxi, China

Background: Small intestinal neuroendocrine tumors (SI-NETs) are the most common malignant tumors of the small intestine, with many patients presenting with metastases and their incidence increasing. We aimed to find effective diagnostic biomarkers for patients with primary and metastatic SI-NETs that could be applied for clinical diagnosis.

Methods: We downloaded GSE65286 (training set) and GSE98894 (test set) from the GEO database and performed differential gene expression analysis to obtain differentially expressed genes (DEGs) and differentially expressed long non-coding RNAs (DElncRNAs). The functions and pathways involved in these genes were further explored by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. In addition, a global regulatory network involving dysregulated genes in SI-NETs was constructed based on RNAInter and TRRUST v2 databases, and the diagnostic power of hub genes was identified by receiver operating characteristic curve (ROC).

Results: A total of 2,969 DEGs and DElncRNAs were obtained in the training set. Enrichment analysis revealed that biological processes (BPs) and KEGG pathways were mainly associated with cancer. Based on gene set enrichment analysis (GSEA), we obtained five BPs (cytokinesis, iron ion homeostasis, mucopolysaccharide metabolic process, platelet degranulation and triglyceride metabolic process) and one KEGG pathway (ppar signaling pathway). In addition, the core set of dysregulated genes obtained included MYL9, ITGV8, FGF2, FZD7, and FLNC. The hub genes were upregulated in patients with primary SI-NETs compared to patients with metastatic SI-NETs, which is consistent with the training set. Significantly, the results of ROC analysis showed that the diagnostic power of the hub genes was strong in both the training and test sets.

Conclusion: In summary, we constructed a global regulatory network in SI-NETs. In addition, we obtained the hub genes including MYL9, ITGV8, FGF2, FZD7, and FLNC, which may be useful for the diagnosis of patients with primary and metastatic SI-NETs.

KEYWORDS

hub genes, intestinal neuroendocrine tumor, primary small intestinal neuroendocrine tumor, metastatic small intestinal neuroendocrine rumors, lncRNA

Introduction

Neuroendocrine tumors (NETs) arise from specialized cells, which are widely dispersed throughout the gastroenteropancreatic tract and lungs (Colao et al., 2020). The incidence and prevalence of NETs have been rising that may be due to the emergence of early-stage disease and stage migration. Furthermore, there have been studies showing that epigenetics may help refine the diagnosis, as well as identify targeted therapies that interfere with epigenetically sensitive pathways (Dasari et al., 2017). SI-NETs are usually small, but frequently leading to lymph node metastases associated with a desmoplastic reaction of the mesentery. Moreover, although SI-NETs are slow growing tumors they frequently show liver metastases at the time of initial diagnosis (Bosch et al., 2018). Therefore, even small tumors with a favorable grading (commonly G1 or G2), can result in a deteriorated prognosis due to distant metastases. Furthermore, there remains significant variability in survival, even among those with metastatic disease. In patients with stage IV NETs of the small bowel, 25% of patients survive less than 2 years while 30% live more than 10 years (Ahmed et al., 2009).

It is usually difficult to make the diagnosis of SI-NETs at the early stage, due to the primary tumors tending to be little and generally without symptoms before the occurrence of bleeding, abdominal pain, obstruction, as well as carcinoid and metastatic syndrome (Norlen et al., 2012; Howe et al., 2017). Current clinical practice uses a documented proliferative index to describe the disease as a whole, including predicting progression of the liver metastases. Pathological examination of tumor specimens has been used to help determine prognosis (Jung et al., 2019). However, the reliability of a single tumor specimen is always subject to sampling error. A recent study demonstrated heterogeneity within an individual tumor (intertumoral) in well-differentiated NETs metastatic to the liver, as Ki67 indices varied widely in different areas within a single lesion (Dhall et al., 2012). Based on those findings, it was predicted that nine core biopsies would be required to obtain the true high Ki67 in a single lesion (Moris et al., 2017). Therefore, it is meaningful to find clinically usable diagnostic markers. Ferroptosis is an iron-dependent form of non-apoptotic cell death characterized by lipid peroxidation, which is widely involved in various diseases and cancers (Dixon et al., 2012). In addition, pharmacological modulation of ferroptosis shows great potential in the treatment of drug-resistant cancers (Jiang et al., 2021). It has been demonstrated that ferroptosis enhances the cytotoxic effects of gemcitabine in pancreatic cancer (Ye et al., 2021). However, the role of ferroptosis in SI-NETs has not been investigated. In order to move towards precision medicine, the genomic landscape of SI-NET has been increasingly studied over the past years with the aim of revealing the molecular events behind NET tumorigenesis, facilitating the identification of new therapeutic targets, rational (targeted) therapeutic management strategies and improving prognosis (Samsom et al., 2019). Recently, several biomarkers have been revealed to be associated with pathogenesis or tumor progression. For example, the possible role of EZH2 as a candidate oncogene for SI-NETs and suggests that CPI-1205 and metformin should be further evaluated as therapeutic options for patients with SI-NETs (Barazeghi et al., 2021). However the pathogenesis of SI-NETs remains largely unknown. Recent advances in the molecular mechanisms of SI-NETs development may improve the management of these tumors.

In the current study, we aimed to find the diagnostic marker used to diagnosis metastatic SI-NETs. We used a bioinformatics approach to identify key genes associated with metastatic SI-NETs, using the Gene Expression Omnibus Database (GEO) database. A global regulatory network was constructed to identify potential therapeutic targets. Furthermore, the diagnostic ability of core genes in the training and test sets was verified by receiver operating characteristic curve (ROC) analysis.

Material and methods

Data collection and processing

We downloaded gene expression data from the microarray study of Andersson et al. (2016) (Agilent-014850) (accession number: GSE65286) and Alvarez et al. (2018) (Illumina HiSeq 2500) (accession number: GSE98894). The data set of GSE65286 based on GPL4133 platform includes 21 patients with liver metastatic SI-NETs and 10 patients with primary SI-NETs. The GSE98894 data set based on GPL16791 platform includes 37 patients with liver metastatic SI-NETs and 44 patients with primary SI-NETs. In the original study, the GSE65286 dataset was used as a training dataset, while the GSE98894 was used as a test dataset. The justRMA method in the affy package (Gautier et al., 2004) was applied to normalize the raw data of the two data sets. If one gene corresponded to multiple probes, the average expression value of these probes was considered to be the expression value of the gene. The lncRNA expression data were obtained by reannotating the probes strategy according to previous study (Niu et al., 2019). The probe sets were mapped to Ensembl gene IDs based on the latest version of the NetAffx Annotation File (HuGene-1_0-st-v1 Probeset Annotations, CSV Format, Release 36).

Principal components analysis, differential gene expression analysis and bidirectional hierarchical clustering

PCA was performed prior to differential gene expression analysis using the expression profiles of all genes in R (version 3.4.2) (https://cran.r-project.org). The limma package (Yu et al., 2019a) was used to analyze differentially expressed genes and long non-coding RNA (DEGs and DElncRNA), between patients with primary and metastatic SI-NETs. The DEGs and DElncRNAs of the datasets with adjusted p-value <0.05 were considered for subsequent analysis. PCA was subsequently performed using the expression profiles of the DEGs and DElncRNAs. Samples were plotted in two-dimensional plots across the first two principal components. Bidirectional hierarchical clustering based on the expression profile in the GSE65286 dataset was performed by calculating the centered Person correlation coefficient. A heatmap was then constructed using the R package pheatmap (version 1.0.12) (Yu et al., 2019a). The DEGs and DElncRNAs with |log2 fold change (FC)| >1.5 and $p < 0.01$ were shown in heatmap.

Immune cell infiltration

The infiltration levels of immune cells were evaluated using ssGSEA in GSVA R software package in GSE65286 datasets. Differences in infiltration of immune cells between primary and metastatic SI-NETs were calculated with the limma R package. We also evaluated potential correlations between the feature genes and immune cells using Pearson correlation analysis. The CIBERSORT (https://cibersort.stanford.edu/) also used to evaluate proportion of immune cells in SI-NETs. Immune cells expressed as 0 were excluded from the analysis.

Enrichment analysis

The Clusterprofiler R package (Tan et al., 2020) was used to functionally analyze key DEGs in gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The threshold for significant differences was $p < 0.05$. The gene set variation analysis (GSVA) (Guo et al., 2019) was used to further explore significant differences between patients with primary and metastases SI-NETs, according to the enrichment score of gene sets defined by signaling pathways. The threshold was $p < 0.05$. Gene set enrichment analysis (GSEA) was performed using GSEA software (Subramanian et al., 2005). Gene sets used here were downloaded from the Molecular Signature Database [MSigDB (Liberzon et al., 2015)]. GSEA result satisfying a nominal *p*-value cut-off of <0.05 with a FDR >0.25 were considered statistically significant. Furthermore, GO and KEGG networks were drawn using Cytoscape (Pinto et al., 2019) and ClueGO plug-in (Bindea et al., 2009).

Constructing comprehensive regulatory network

Interactions between lncRNA and their target genes were downloaded from RNAInter database [(Agnello et al., 2018), http://www.rna-society.org/raid/]. The interactions between transcription factors (TFs) and stemness-highly relatecond mRNAs were downloaded from TRRUST v2 database (Han et al., 2018). The correlation analysis was performed to explore the correlation between the lncRNA/TF and its targets. Subsequently, combining with enrichment analysis result, a lncRNA/TF-target-KEGG pathway (transcriptional regulatory network) involved in dysregulated genes in SI-NETs was constructed.

Identifying hub genes

The GSVA function in GSVA package (Hanzelmann et al., 2013) was used as an unsupervised and non-parametric method for estimating the variation of gene set in patients with primary and metastatic SI-NETs. GSVA scores were calculated non-parametrically using a Kolmogorov Smirnoff-like random walk statistic and a negative value for a particular gene set, meaning that the gene set has a lower expression than the same gene set with a positive value. The receiver operator characteristic (ROC) curve anaysis was performed to obtain the diagnostic power of the hub genes, and the area under the ROC curve (AUC) of each cutoff was measured in accordance with previous sports (Guo et al., 2019).

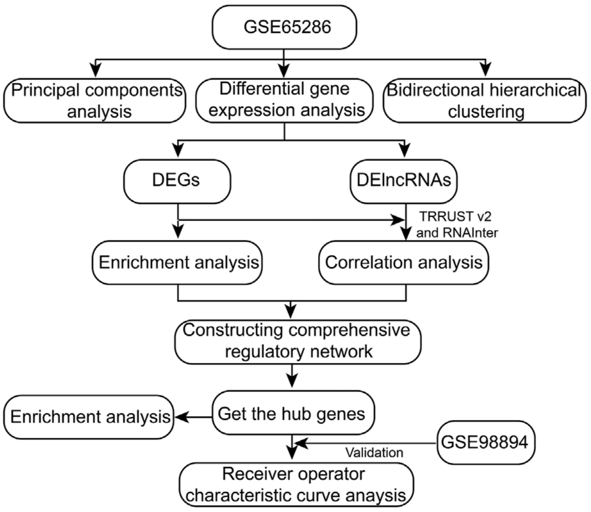


FIGURE 1
The workflow of this study.

Molecule docking

Docking was the process of bringing one molecule in vicinity with another molecule. The present research work was conducted the molecule to be docked with another molecule. We downloaded the 3D structure of the target protein from Protein Databank (www.rcsb.org). Docking was conducted using free software's Hex v6.0. And the results were visualized with Pymol software. Docking energy less than 0 means the two have binding potential, and the smaller the energy, the higher the binding potential.

Statistical analysis

Statistical analyses were performed using R (https://www.rproject.org/). The expression levels of genes were analyzed by an unpaired *t*-test, was considered statistically significant when $p < 0.05$. The Bioinforcloud platform is the main platform used for data analysis in this study (http://www.bioinforcloud.org.cn).

Results

Dysregulated genes in SI-NETs

The workflow of this study was showed in Figure 1. A total of 2,959 DEGs (1,642 down-expressed and 1,317 up-expressed DEGs, Supplementary Table S1) and 312 DElncRNA (157 down-expressed and 155 up-expressed DElncRNA, Supplementary Table S2) were obtained *via* difference analysis (Figure 2A).

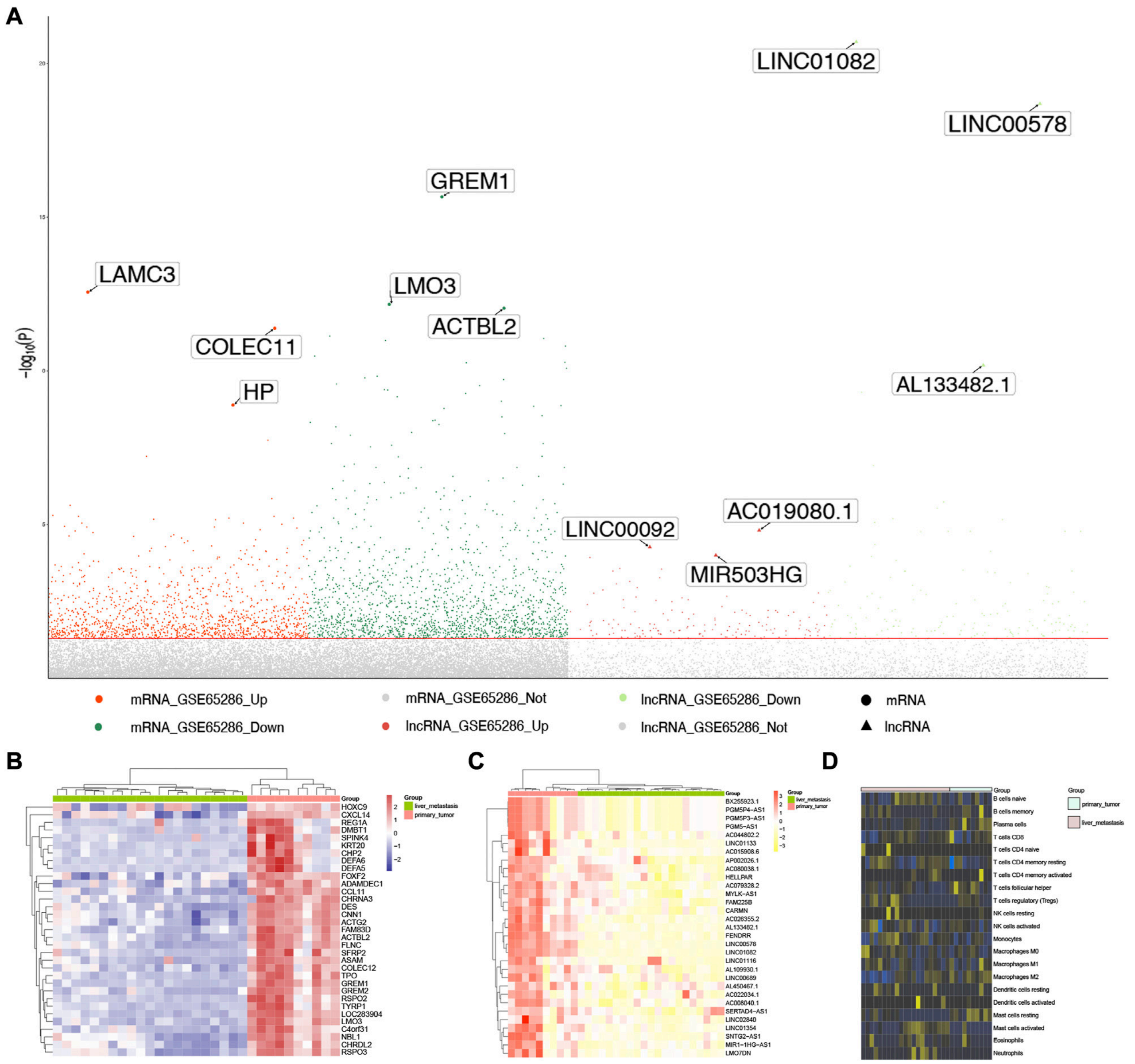


FIGURE 2
Expression disorders of SI-NETs. **(A)** Manhattan plot for DEGs and DElncRNA. The three most significantly DEGs and DElncRNA are marker and labeled with their names. **(B)** Hierarchical clustering dendrograms of the expression patterns of DEGs that distinguish between patients with primary and metastatic SI-NETs. **(C)** Hierarchical clustering dendrograms of the expression patterns of DElnRNA that distinguish between patients with primary and metastatic SI-NETs. DEGs: differentially expressed genes, DElncRNA: differentially expressed lncRNA, SI-NETs: small intestine neuroendocrine tumors. **(D)** Infiltration of immune cells between primary and metastatic SI-NETs.

Supplementary Figure S1 presented the results of PCA using all genes and Supplementary Figure S2 presents the results of PCA using the DEGs and DElncRNAs. As demonstrated in Supplementary Figure S2, samples with primary and metastatic SI-NETs samples are more easily distinguished using DEGs and DElncRNAs. We further explored the different expression for the most significant DEGs and DElncRNA in primary and metastatic SI-NET patients, they were shown in Figures 2B, C, respectively. In addition, we evaluated the infiltration of immune cells between primary and metastatic SI-NETs (Figure 2D).

The biological functions of dysregulated genes in SI-NETs

To evaluate the affected functions for DEGs in this study, we performed enrichment analysis. We found that biological processes were closely related to SI-NETs, such as positive regulation of cell adhesion, multicellular organismal homeostasis and regulation of neurotransmitter levels (Figure 3A). In addition, the KEGG pathways obtained were shown in Figure 3B. Among them, we found that wnt signaling pathway (Scarpa, 2019), focal adhesion (Francois et al., 2015),

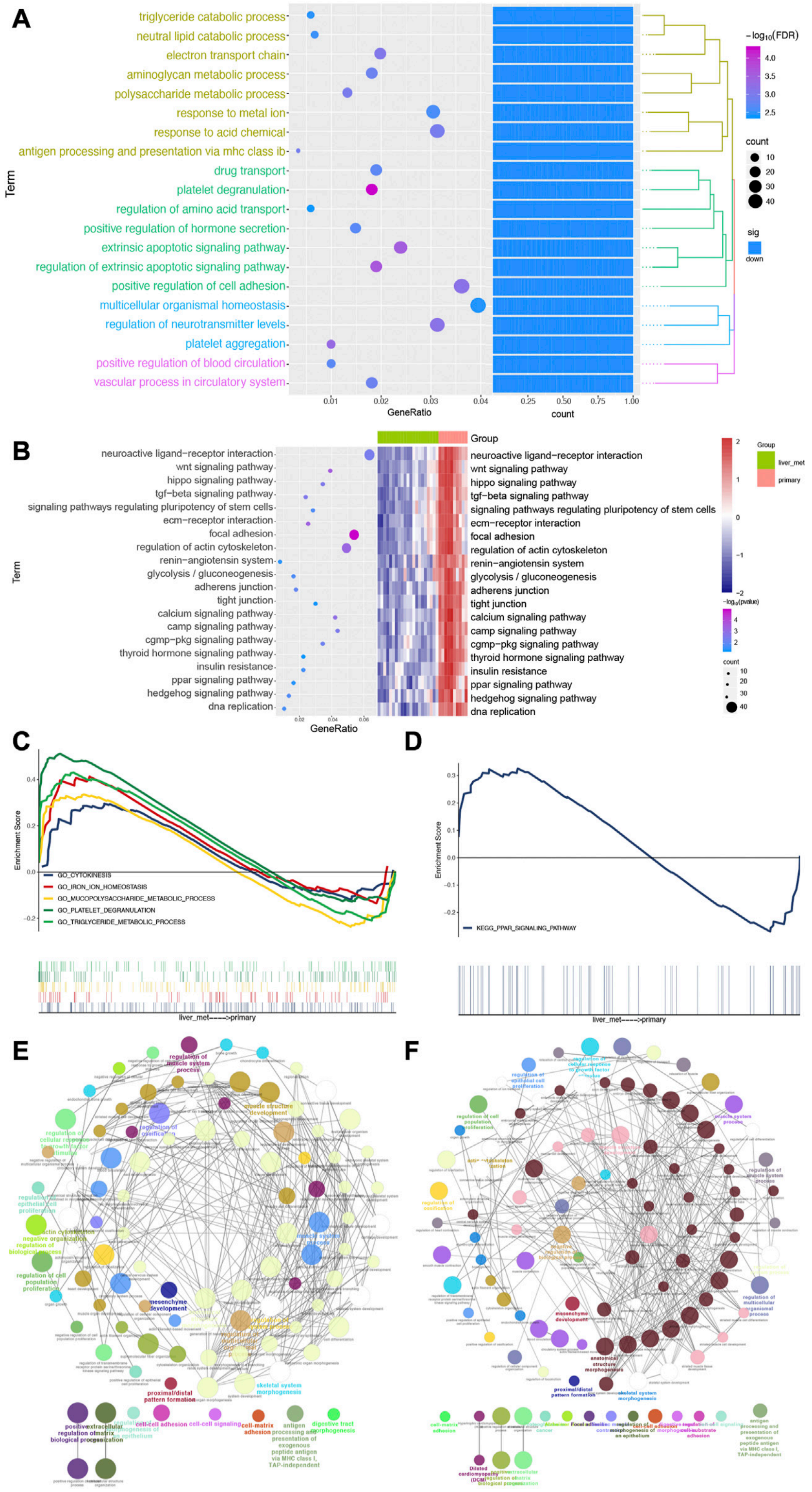


FIGURE 3
Biological processes and pathways of DEGs for patients with SI-NETs. **(A)** The top 20 biological processes for patients with SI-NETs. **(B)** The top 20 KEGG pathways for patients with SI-NETs. **(C)** GSEA results revealed the significantly enriched GO-BPs between patients with primary and metastatic SI-NETs. **(D)** GSEA results revealed the significantly enriched KEGG pathways between patients with primary and metastatic SI-NETs. **(E)** The ClueGo network for GO-BPs in patients with SI-NETs. **(F)** The ClueGO network for KEGG pathways in patients with SI-NETs. SI-NETs, small intestine neuroendocrine tumors; GSEA, gene set enrichment analysis; GO, gene ontology; BPs, biological processes; KEGG, Kyoto Encyclopedia of Genes and Genomes.

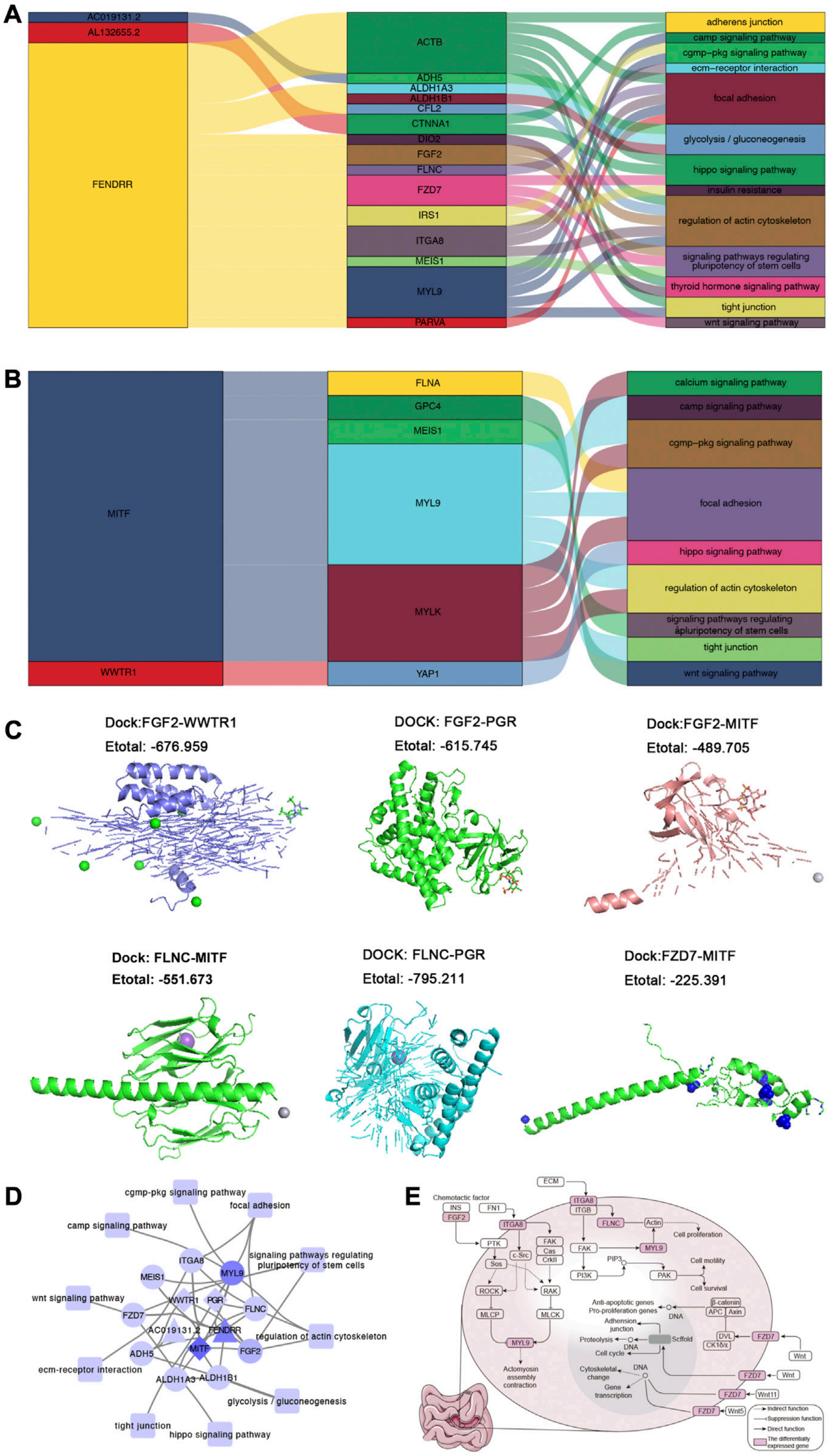


FIGURE 4
The global regulatory landscape in SI-NETs. **(A)** Sankey plot: Graphical summary of lncRNA regulating KEGG pathways through mRNA. **(B)** Sankey plot: Graphical summary of TFs regulating KEGG pathways through mRNA. **(C)** Molecular docking model: The prediction of targeted binding FGF2 and WWTR1/PGR/MITF, FLNC and MITF/PGR, as well as FZD7 and MITF. **(D)** The global regulatory landscape SI-NETs: lncRNA/TFs-mRNA-KEGG pathways network. **(E)** 3 KEGG pathways for further study. Node maps degree value. Light purple represents a smaller degree value, while dark purple representing a larger degree value. Degree value represents the number of genes interacting with other genes. SI-NETs, small intestine neuroendocrine tumors; KEGG, Kyoto Encyclopedia of Genes and Genomes; TFs, transcription factors.

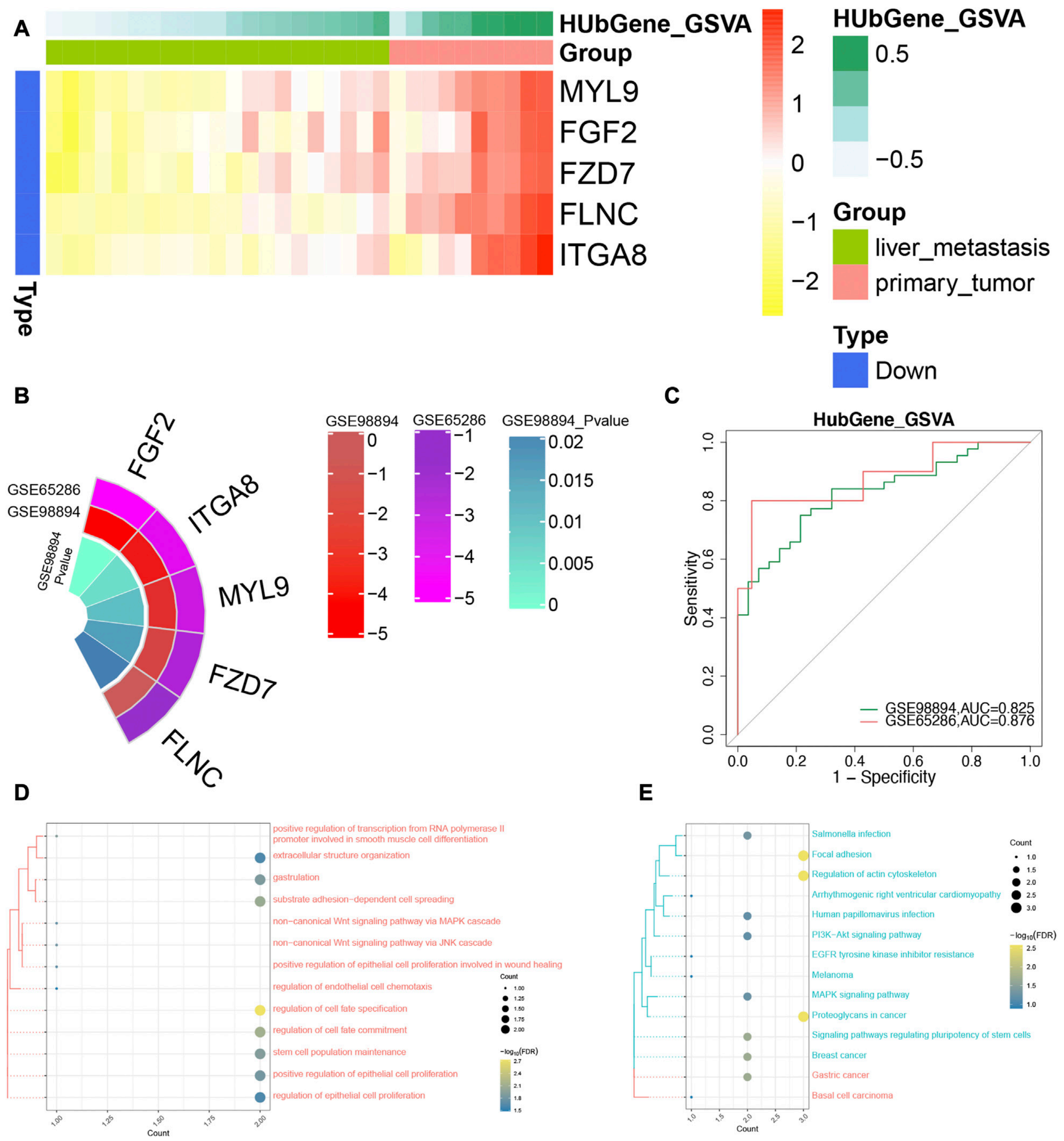


FIGURE 5
The identification of diagnostic power of the hub genes in SI-NETs. **(A)** The heatmap for the hub genes in GSE65286. **(B)** Circo plot: The expression level of the hub genes in GSE65286 and GSE98894. **(C)** The ROC analysis for the hub genes in GSE65286 and GSE98894. **(D)** The hub genes are significantly involved in biological processes. **(E)** The hub genes are significantly involved in the KEGG signaling pathway. SI-NETs, small intestine neuroendocrine tumors; ROC, receiver operator characteristic.

Hippo signaling pathway (Zhao et al., 2018) and regulation of actin cytoskeleton (Kim et al., 2016) were closely associated with SI-NETs. Furthermore, the GSEA results showed that cytokinesis, iron ion homeostasis, mucopolysaccharide metabolic process, platelet degranulation and triglyceride metabolic process were upregulated in metastatic SI-NET patients, compared to primary SI-NET patients (Figure 3C). Moreover, the ppar signaling pathway was upregulated in metastatic SI-NET patients, compared to primary SI-NET patients (Figure 3D). The ClueGo networks of GO and KEGG were shown in Figures 3E, F, respectively. Where each node represents a term, the connection between the nodes reflects the correlation between the terms, and the color of the node reflects the enrichment classification of that node.

The comprehensive regulatory landscape in SI-NETs

The correlation analysis result showed that lncRNA (Figure 4A) and TF (Figure 4B) can regulate the phenotype of organism *via* DEGs. To further explore if there was possibility of targeted binding of genes having correlation in Figures 4A, B, the molecule docking was performed. Among them, we found that the docking energy of FGF2 and WWTR1/PGR/MITF, FLNC, and MITF/PGR, as well as FZD7 and MITF was less than 0, which indicating there was possibility of targeted binding (Figure 4C). Combining the correlation analysis of TF-mRNA-KEGG pathway and lncRNA-mRNA-KEGG pathway, the TF/lncRNA-mRNA-KEGG pathway translational regulatory network was constructed (Figure 4D). As shown in Figure 4D, 2 lncRNA and 3 TFs can regulate 10 KEGG pathways *via* 10 mRNA. Combining the previous study (Kim et al., 2016; Zhao et al., 2018; Scarpa, 2019), We chose Wnt signaling pathway, hippo signaling pathway, focal adhesion and regulation of actin cytoskeleton as interested pathways for further study. We found FGF2 and ITGV8 may reduce actomyosin assembly contraction *via* acting on downstream genes in regulation of actin cytoskeleton. Meaningfully, ITGV8, FLNC, and MYL9 can promote the proliferation, motility and survival of cells in focal adhesion, which occurred with the help of acting on downstream genes and interactions among them. In Hippo signaling pathway, FZD7 can act on anti-apoptotic/pro-proliferation genes for promoting the proliferation of cells. While in Wnt signaling pathway, FZD7 can adjust interactions and activities of cells through scffold and DNA (Figure 4E).

Identification of the hub genes in SI-NETs

Combined with the interested pathways and DEGs, five genes (MYL9, FGF2, FZD7, FLNC, and ITGA8) were defined the hub genes in patients with SI-NET. In addition, we found that they were upregulated in patients with primary SI-NETs, compared to patients with metastatic SI-NETs (Figure 5A). In test dataset, we found the hub genes were also down-expressed for metastatic SI-NET patients in GSE98894, which was consistent with that of GSE65286 (Figure 5B). In addition, the ROC results showed that the diagnostic ability of the hub genes were all strong in the training set (GSE65286) (Supplementary Figure S3). This result was also validated in the test set (GSE98894) (Figure 5C). Furthermore, enrichment analysis of hub genes revealed that these genes were significantly involved in biological processes such as regulation of cell fate specification, substrate adhesion-dependent cell spreading (Figure 5D), and signaling pathways such as Focal adhesion, and Regulation of actin cytoskeleton (Figure 5E).

Discussion

SI-NETs are the most common form of neoplasm in the small bowel (Dasari et al., 2017; Norlen et al., 2018; Deguelte et al., 2020). The common diagnosis methods are pathological examination and Ki67 indices (Yao et al., 2008; Yang et al., 2011). There are some limitations, such as sampling error and differences in variable lesions. In the current study, we screened the DEGs and DElncRNA between patients with primary and metastatic SI-NETs. In addition, we explored the biological functions of DEGs in SI-NETs. We further constructed the comprehensive regulatory network and obtained the hub genes. At last, we evaluated the expression and diagnostic power of the hub genes in both training set and test set.

In the current study, 2 common biological processes (platelet degranulation and triglyceride metabolic process) and 1 common KEGG pathway (ppar signaling pathway) were obtained between enrichment analysis and GSEA. Notably, activation of ppar signaling pathway may reduce cell survival. Subsequently, the comprehensive regulatory network was constructed. We found that TFs (WWTR1, MITF and PGR) and lncRNA (FENDRR and AC019191.2) were the important regulators that can participate in mRNA transcriptional and post-transcriptional regulation. Ferroptosis is a form of iron-dependent non-apoptotic cell death that occurs through an increase in cellular phospholipid peroxidation in the presence of an impaired phospholipid peroxide repair system (Venkatesh et al., 2020). PPAR family can regulate iron death sensitivity (Venkatesh et al., 2020). Among them, WWTR1 is a downstream effector of the Hippo signaling and a transcriptional factor TEAD (Hong and Guan, 2012). In addition, WWTR1 can participate in many cancer cell signaling pathways, such WNT, mTOR and EMT signaling pathway (Wei et al., 2019). At the same time, MITF can regulate a range of biological processes, including cell metabolism, senescence, invasion, proliferation and differentiation (Goding and Arnheiter, 2019). PGR is a kind of nuclear progestin receptors that can control various physiological processes in mammals (Zhu et al., 2015). FENDDR plays the roles of acting both as oncogenic and tumor-suppressive factors as a kind of lncRNA (Yu et al., 2019b). Moreover, we found that AC019191.2 also plays an important role in transcriptional and post-transcriptional regulation for patients with primary and metastatic SI-NETs. In summary, TF and lncRNA can drive the transcription and post-transcriptional regulation, thereby affecting the phenotype of patients with SI-NETs.

In order to further explore the dysregulated mechanism in patients with SI-NETs, we chose 4 KEGG pathways (wnt signaling pathway, focal adhesion, Hippo signaling pathway and regulation of actin cytoskeleton) for the next study, combining with the previous studies. And the hub genes were obtained, including MYL9, FGF2, FZD7, FLNC, and ITGA8. MYL9 is necessary for experimental metastasis and cytoskeletal dynamics (Wang et al., 2017). And its overexpression is closely related to invasion-promoting functions of cancer-associated fibroblasts (Wang et al., 2017). FGF2 can function as a potential oncogenic protein in multiple malignancy tumors (Wang et al., 2015). FZD7 is a trans-membrane receptor. Some studies showed that FZD7 appears to promote tumorigenesis and cancer progression (Qiu et al., 2016). In addition, in ovarian cancer, FZD7 marks a cell population that is highly susceptible to ferroptosis (Wang et al., 2021). Several studies reports that FLNC is related to cancer, but the results are not inconsistent (Qiao et al., 2015). Moreover, we found that the different expression of ITGV8 mediated the dysregulation of biological functions in SI-NETs. In addition, we found that the hub genes were over-expressed in patients with primary SI-NETs compared to patients with metastatic SI-NETs. And the diagnostic power of the hub genes were high both in training set and test set, which indicated it may be used as a diagnostic biomarker for patients with SI-NETs.

There are some limitations. Firstly, the study is based on a retrospective data, which needs to be verified in a larger prospective cohort. Secondly, this study is mainly based bioinformatics analysis, which needs the further biological experiments.

Conclusion

We construct a global regulatory network in patients with primary and metastatic SI-NETs. In addition, we obtained the hub genes that consists of MYL9, FGF2, FZD7, FLNC, and ITGA8, which may be meaningful for diagnosis of patients with primary and metastatic SI-NETs.

Data availability statement

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

Author contributions

JC analyzed the data and drafted the manuscript. YM and XH designed the study and revised the manuscript. XL, XT, and YX produced and edited the chart. All authors contributed to data analysis, drafting or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Funding

This study was funded by the Guangxi Key Research and Development Plan (No. GUIKEAB16380215).

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.2023.1110396/full#supplementary-material

SUPPLEMENTARY FIGURE S1
The PCA for lncRNA **(A)** and mRNA **(B)** in patients with primary and metastatic SI-NETs. PCA, principal component analysis; SI-NETs, small intestine neuroendocrine tumors.

SUPPLEMENTARY FIGURE S2
The PCA for DElncRNA **(A)** and DEGs **(B)** in patients with primary and metastatic SI-NETs. PCA, principal component analysis; SI-NETs, small intestine neuroendocrine tumors; DElncRNA, differentially expressed lncRNA; DEGs, differentially expressed genes.

SUPPLEMENTARY FIGURE S3
ROC curves of the hub genes in GSE65286.

SUPPLEMENTARY TABLE S1
List of differentially expressed genes.

SUPPLEMENTARY TABLE S2
List of differentially expressed lncRNAs.

References

Agnello, R., Barbisan, M., Furno, I., Guittienne, P., Howling, A. A., Jacquier, R., et al. (2018). Cavity ring-down spectroscopy to measure negative ion density in a helicon plasma source for fusion neutral beams. *Rev. Sci. Instrum.* 89 (10), 103504. doi:10.1063/1.5044504

Ahmed, A., Turner, G., King, B., Jones, L., Culliford, D., McCance, D., et al. (2009). Midgut neuroendocrine tumours with liver metastases: Results of the UKINETS study. *Endocr. Relat. Cancer* 16 (3), 885–894. doi:10.1677/ERC-09-0042

Alvarez, M. J., Subramaniam, P. S., Tang, L. H., Grunn, A., Aburi, M., Rieckhof, G., et al. (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat. Genet.* 50 (7), 979–989. doi:10.1038/s41588-018-0138-4

Andersson, E., Arvidsson, Y., Sward, C., Hofving, T., Wangberg, B., Kristiansson, E., et al. (2016). Expression profiling of small intestinal neuroendocrine tumors identifies subgroups with clinical relevance, prognostic markers and therapeutic targets. *Mod. Pathol.* 29 (6), 616–629. doi:10.1038/modpathol.2016.48

Barazeghi, E., Hellman, P., Norlen, O., Westin, G., and Stalberg, P. (2021). EZH2 presents a therapeutic target for neuroendocrine tumors of the small intestine. *Sci. Rep.* 11 (1), 22733. doi:10.1038/s41598-021-02181-7

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25 (8), 1091–1093. doi:10.1093/bioinformatics/btp101

Bosch, F., Bruewer, K., D'Anastasi, M., Ilhan, H., Knoesel, T., Pratschke, S., et al. (2018). Neuroendocrine tumors of the small intestine causing a desmoplastic reaction of the mesentery are a more aggressive cohort. *Surgery* 164 (5), 1093–1099. doi:10.1016/j.surg.2018.06.026

Colao, A., de Nigris, F., Modica, R., and Napoli, C. (2020). Clinical epigenetics of neuroendocrine tumors: The road ahead. *Front. Endocrinol. (Lausanne)* 11, 604341. doi:10.3389/fendo.2020.604341

Dasari, A., Shen, C., Halperin, D., Zhao, B., Zhou, S., Xu, Y., et al. (2017). Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol.* 3 (10), 1335–1342. doi:10.1001/jamaoncol.2017.0589

Deguelte, S., Perrier, M., Hammoutene, C., Cadiot, G., and Kianmanesh, R. (2020). Surgery and perioperative management in small intestinal neuroendocrine tumors. *J. Clin. Med.* 9 (7), 2319. doi:10.3390/jcm9072319

Dhall, D., Mertens, R., Bresee, C., Parakh, R., Wang, H. L., Li, M., et al. (2012). Ki-67 proliferative index predicts progression-free survival of patients with well-differentiated ileal neuroendocrine tumors. *Hum. Pathol.* 43 (4), 489–495. doi:10.1016/j.humpath.2011.06.011

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 (5), 1060–1072. doi:10.1016/j.cell.2012.03.042

Francois, R. A., Maeng, K., Nawab, A., Kaye, F. J., Hochwald, S. N., and Zajac-Kaye, M. (2015). Targeting focal adhesion kinase and resistance to mTOR inhibition in pancreatic neuroendocrine tumors. *J. Natl. Cancer Inst.* 107 (8), djv123. doi:10.1093/jnci/djv123

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

Goding, C. R., and Arnheiter, H. (2019). MITF-the first 25 years. *Genes Dev.* 33 (15-16), 983–1007. doi:10.1101/gad.324657.119

Guo, Q., Guan, G. F., Cheng, W., Zou, C. Y., Zhu, C., Cheng, P., et al. (2019). Integrated profiling identifies caveolae-associated protein 1 as a prognostic biomarker of malignancy in glioblastoma patients. *CNS Neurosci. Ther.* 25 (3), 343–354. doi:10.1111/cns.13072

Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46 (D1), D380–D386. doi:10.1093/nar/gkx1013

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

Hong, W., and Guan, K. L. (2012). The YAP and TAZ transcription co-activators: Key downstream effectors of the mammalian hippo pathway. *Semin. Cell Dev. Biol.* 23 (7), 785–793. doi:10.1016/j.semcdb.2012.05.004

Howe, J. R., Cardona, K., Fraker, D. L., Kebebew, E., Untch, B. R., Wang, Y. Z., et al. (2017). The surgical management of small bowel neuroendocrine tumors: Consensus guidelines of the north American neuroendocrine tumor society. *Pancreas* 46 (6), 715–731. doi:10.1097/MPA.0000000000000846

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

Jung, J., Hwang, S., Hong, S. M., Kim, K. H., Ahn, C. S., Moon, D. B., et al. (2019). Long-term postresection prognosis of primary neuroendocrine tumors of the liver. *Ann. Surg. Treat. Res.* 97 (4), 176–183. doi:10.4174/astr.2019.97.4.176

Kim, M. H., Kim, J., Hong, H., Lee, S. H., Lee, J. K., Jung, E., et al. (2016). Actin remodeling confers BRAF inhibitor resistance to melanoma cells through YAP/TAZ activation. *EMBO J.* 35 (5), 462–478. doi:10.15252/embj.201592081

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Moris, D., Tsilimigras, D. I., Ntanasis-Stathopoulos, I., Beal, E. W., Felekouras, E., Vernadakis, S., et al. (2017). Liver transplantation in patients with liver metastases from neuroendocrine tumors: A systematic review. *Surgery* 162 (3), 525–536. doi:10.1016/j.surg.2017.05.006

Niu, X., Zhang, J., Zhang, L., Hou, Y., Pu, S., Chu, A., et al. (2019). Weighted gene Co-expression network analysis identifies critical genes in the development of heart failure after acute myocardial infarction. *Front. Genet.* 10, 1214. doi:10.3389/fgene.2019.01214

Norlen, O., Montan, H., Hellman, P., Stalberg, P., and Sundin, A. (2018). Preoperative (68)Ga-DOTA-Somatostatin analog-PET/CT hybrid imaging increases detection rate of intra-abdominal small intestinal neuroendocrine tumor lesions. *World J. Surg.* 42 (2), 498–505. doi:10.1007/s00268-017-4364-1

Norlen, O., Stalberg, P., Oberg, K., Eriksson, J., Hedberg, J., Hessman, O., et al. (2012). Long-term results of surgery for small intestinal neuroendocrine tumors at a tertiary referral center. *World J. Surg.* 36 (6), 1419–1431. doi:10.1007/s00268-011-1296-z

Pinto, P. I. S., Andrade, A. R., Thorne, M. A. S., Estevao, M. D., Canario, A. V. M., and Power, D. M. (2019). Estradiol and genistein effects on the sea bass (*Dicentrarchus labrax*) scales: Transcriptome dataset. *Data Brief.* 27, 104587. doi:10.1016/j.dib.2019.104587

Qiao, J., Cui, S. J., Xu, L. L., Chen, S. J., Yao, J., Jiang, Y. H., et al. (2015). Filamin C, a dysregulated protein in cancer revealed by label-free quantitative proteomic analyses of human gastric cancer cells. *Oncotarget* 6 (2), 1171–1189. doi:10.18632/oncotarget.2645

Qiu, X., Jiao, J., Li, Y., and Tian, T. (2016). Overexpression of FZD7 promotes glioma cell proliferation by upregulating TAZ. *Oncotarget* 7 (52), 85987–85999. doi:10.18632/oncotarget.13292

Samsom, K. G., van Veenendaal, L. M., Valk, G. D., Vriens, M. R., Tesselaar, M. E. T., and van den Berg, J. G. (2019). Molecular prognostic factors in small-intestinal neuroendocrine tumours. *Endocr. Connect.* 8 (7), 906–922. doi:10.1530/EC-19-0206

Scarpa, A. (2019). The landscape of molecular alterations in pancreatic and small intestinal neuroendocrine tumours. *Ann. Endocrinol. Paris.* 80 (3), 153–158. doi:10.1016/j.ando.2019.04.010

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

Tan, L., Xu, Q., Wang, Q., Shi, R., and Zhang, G. (2020). Identification of key genes and pathways affected in epicardial adipose tissue from patients with coronary artery disease by integrated bioinformatics analysis. *PeerJ* 8, e8763. doi:10.7717/peerj.8763

Venkatesh, D., O'Brien, N. A., Zandkarimi, F., Tong, D. R., Stokes, M. E., Dunn, D. E., et al. (2020). MDM2 and MDMX promote ferroptosis by PPARα-mediated lipid remodeling. *Genes Dev.* 34 (7-8), 526–543. doi:10.1101/gad.334219.119

Wang, F., Yang, L., Shi, L., Li, Q., Zhang, G., Wu, J., et al. (2015). Nuclear translocation of fibroblast growth factor-2 (FGF2) is regulated by Karyopherin-β2 and Ran GTPase in human glioblastoma cells. *Oncotarget* 6 (25), 21468–21478. doi:10.18632/oncotarget.4097

Wang, J. H., Zhang, L., Huang, S. T., Xu, J., Zhou, Y., Yu, X. J., et al. (2017). Expression and prognostic significance of MYL9 in esophageal squamous cell carcinoma. *PLoS One* 12 (4), e0175280. doi:10.1371/journal.pone.0175280

Wang, Y., Zhao, G., Condello, S., Huang, H., Cardenas, H., Tanner, E. J., et al. (2021). Frizzled-7 identifies platinum-tolerant ovarian cancer cells susceptible to ferroptosis. *Cancer Res.* 81 (2), 384–399. doi:10.1158/0008-5472.CAN-20-1488

Wei, J., Wang, L., Zhu, J., Sun, A., Yu, G., Chen, M., et al. (2019). The Hippo signaling effector WWTR1 is a metastatic biomarker of gastric cardia adenocarcinoma. *Cancer Cell Int.* 19, 74. doi:10.1186/s12935-019-0796-z

Yang, Z., Tang, L. H., and Klimstra, D. S. (2011). Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: Implications for prognostic stratification. *Am. J. Surg. Pathol.* 35 (6), 853–860. doi:10.1097/PAS.0b013e31821a0696

Yao, J. C., Hassan, M., Phan, A., Dagohoy, C., Leary, C., Mares, J. E., et al. (2008). One hundred years after "carcinoid": Epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J. Clin. Oncol.* 26 (18), 3063–3072. doi:10.1200/JCO.2007.15.4377

Ye, Z., Zhuo, Q., Hu, Q., Xu, X., Mengqi, L., Zhang, Z., et al. (2021). FBW7-NRA41-SCD1 axis synchronously regulates apoptosis and ferroptosis in pancreatic cancer cells. *Redox Biol.* 38, 101807. doi:10.1016/j.redox.2020.101807

Yu, J., Wu, X., Huang, K., Zhu, M., Zhang, X., Zhang, Y., et al. (2019). Bioinformatics identification of lncRNA biomarkers associated with the progression of esophageal squamous cell carcinoma. *Mol. Med. Rep.* 19 (6), 5309–5320. doi:10.3892/mmr.2019.10213

Yu, Z., Zhao, H., Feng, X., Li, H., Qiu, C., Yi, X., et al. (2019). Long non-coding RNA FENDRR acts as a miR-423-5p sponge to suppress the treg-mediated immune escape of hepatocellular carcinoma cells. *Mol. Ther. Nucleic Acids* 17, 516–529. doi:10.1016/j.omtn.2019.05.027

Zhao, F., Yang, Z., Chen, Y., Zhou, Q., Zhang, J., Liu, J., et al. (2018). Deregulation of the hippo pathway promotes tumor cell proliferation through YAP activity in human sporadic vestibular schwannoma. *World Neurosurg.* 117, e269–e279. doi:10.1016/j.wneu.2018.06.010

Zhu, Y., Liu, D., Shaner, Z. C., Chen, S., Hong, W., and Stellwag, E. J. (2015). Nuclear progestin receptor (pgr) knockouts in zebrafish demonstrate role for pgr in ovulation but not in rapid non-genomic steroid mediated meiosis resumption. *Front. Endocrinol. (Lausanne)* 6, 37. doi:10.3389/fendo.2015.00037

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 07 April 2023
DOI 10.3389/fgene.2023.1156230

OPEN ACCESS

EDITED BY
Min Sun,
Hubei University of Medicine, China

REVIEWED BY
Yunfeng Jin,
Fudan University, China
Changhui Sun,
Ruijin Hospital Lu Wan Branch, China

*CORRESPONDENCE
Dexin Zhang,
✉ dexin1994@mail.xjtu.edu.cn

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

RECEIVED 01 February 2023
ACCEPTED 20 March 2023
PUBLISHED 07 April 2023

CITATION
Zhang D and Zhao Y (2023), Identification of natural killer cell associated subtyping and gene signature to predict prognosis and drug sensitivity of lung adenocarcinoma.
Front. Genet. 14:1156230.
doi: 10.3389/fgene.2023.1156230

Identification of natural killer cell associated subtyping and gene signature to predict prognosis and drug sensitivity of lung adenocarcinoma

Dexin Zhang[1]* and Yujie Zhao[2]

[1]Respiratory Department of the Second Affiliated Hospital of Xi'an Jiaotong University Medical College, Xi'an, China, [2]Regional Marketing Department, Yuce Biotechnology Co, Ltd., Dabaihui Center, Shenzhen, China

Introduction: This research explored the immune characteristics of natural killer (NK) cells in lung adenocarcinoma (LUAD) and their predictive role on patient survival and immunotherapy response.

Material and methods: Molecular subtyping of LUAD samples was performed by evaluating NK cell-associated pathways and genes in The Cancer Genome Atlas (TCGA) dataset using consistent clustering. 12 programmed cell death (PCD) patterns were acquired from previous study. Riskscore prognostic models were constructed using Least absolute shrinkage and selection operator (Lasso) and Cox regression. The model stability was validated in Gene Expression Omnibus database (GEO).

Results: We classified LUAD into three different molecular subgroups based on NK cell-related genes, with the worst prognosis in C1 patients and the optimal in C3. Homologous Recombination Defects, purity and ploidy, TMB, LOH, Aneuploidy Score, were the most high-expressed in C1 and the least expressed in C3. ImmuneScore was the highest in C3 type, suggesting greater immune infiltration in C3 subtype. C1 subtypes had higher TIDE scores, indicating that C1 subtypes may benefit less from immunotherapy. Generally, C3 subtype presented highest PCD patterns scores. With four genes, ANLN, FAM83A, RHOV and PARP15, we constructed a LUAD risk prediction model with significant differences in immune cell composition, cell cycle related pathways between the two risk groups. Samples in C1 and high group were more sensitive to chemotherapy drug. The score of PCD were differences in high- and low-groups. Finally, we combined Riskscore and clinical features to improve the performance of the prediction model, and the calibration curve and decision curve verified that the great robustness of the model.

Conclusion: We identified three stable molecular subtypes of LUAD and constructed a prognostic model based on NK cell-related genes, maybe have a greater potential for application in predicting immunotherapy response and patient prognosis.

KEYWORDS

lung adenocarcinoma, NK cells, programmed cell death, immunity, survival, prediction model, chemotherapy drug

1 Background

Lung cancer is a leading cause of cancer mortality in the world (Hirsch et al., 2017). Statistics reported that in 2022 in the United States will die from cancer, and approximately 350 of them die from lung cancer every day (Siegel et al., 2022). Adenocarcinoma (lung adenocarcinoma, LUAD) is currently the predominant histologic type, which accounts for approximately 50% of all lung cancer cases, and is notable for its high incidence, high mortality, and poor prognosis (Succony et al., 2021). Currently, surgery is recommended for early-stage lung cancer and is considered the most effective treatment option, while those with advanced disease are often further supplemented with radiotherapy, chemotherapy, targeted therapy, and immunotherapy (Hoy et al., 2019). Regardless of the interventions used, the overall 5-year survival of LUAD patients remains below 20% (Duma et al., 2019). Therefore, it is necessary to develop current understanding on the pathogenesis of LUAD to provide a theoretical basis for reducing the occurrence of LUAD, improving the treatment of LUAD and its prognosis.

The development of LUAD involves external environment, gene mutation, tumor immunity, and family genetics, and is a multistep, cascade process (Suster and Mino-Kenudson, 2020). As a component of the tumor microenvironment, tumor immune cells are present in all stages of LUAD and play an important role in shaping tumor development (Saab et al., 2020). For example, tumor-associated macrophages can accelerate tumor progression by promoting tumor angiogenesis, metastasis and immune escape. Regulatory T cells inhibit anti-tumor immune responses, thereby promoting the development of immunosuppressive tumor microenvironments and promoting cancer progression (Hsieh et al., 2012). Cytotoxic $CD8^+$ memory T cells kill tumor cells by recognizing specific antigens on them and stimulating an immune response (Arneth, 2019). Dendritic cells are antigen-presenting cells, which are an important bridge between innate and adaptive immunity. Dendritic cells can not only induce cellular immunity and humoral immunity, but also activate natural killer (NK) cells and NK T cells (Sadeghzadeh et al., 2020). NK cells are anti-tumor immune cells that kill cancer cells in the body, but in the tumor microenvironment NK cells are generally reduced in number and impaired in function (Russell et al., 2022). Basic experiments and clinical studies together have shown that NK cells are in the first line of defense against tumors and do not require pre-stimulation to cause NK cells to migrate to the lesion and play an immunomodulatory role (Guillerey, 2020). Phenotypically, NK cell subpopulations display potent antitumor immune cytotoxicity *via* MEK/ERK and PI3K/Akt/mTOR pathways upon stimulation through cytokines such as interleukin (IL) (Valipour et al., 2019). Although patient's immune system can recognize neoantigens produced by tumors with high mutational load (immunogenic "hot" tumors), in terms of its mutational load, lung cancer is immunogenic, only moderately, to some extent. Therefore, the highly complex interaction between LUAD and NK cells is a major challenge to improve immunotherapy.

Studies on the pathogenesis of NK cells in LUAD have delved into the genetic-molecular field, and it is mostly believed that the development of LUAD is the result of a multigene, multistage involvement (Crinier et al., 2020). However, the genetic landscape and immune profile of NK cells in LUAD are unclear, also the prognosis and immune efficacy prediction of LUAD based on NK cells have not been reported. This study first identified stable molecular subtypes of LUAD by consistent clustering of NK cell-associated genes, and further compared clinicopathological, mutational, immunological, and pathway characteristics among the subtypes. Then, we constructed a risk model and a clinical prognostic model, which can be used to evaluate personalized treatment for LUAD patients.

2 Materials and methods

2.1 Source of clinical information and gene expression profile data of NK cells

The clinical information and mRNA transcriptome data of LUAD patients were downloaded from the TCGA GDC API (Colaprico et al., 2016). To verify the accuracy of the results, we also downloaded the clinical and mRNA gene expression data of LUAD patients from the Gene Expression Omnibus database (GEO) database (Toro-Domínguez et al., 2019), including GSE72094, GSE31210 datasets. The TCGA dataset contained 500 LUAD samples as the training set, while the GSE72094 and GSE31210 datasets contained 398 and 226 LUAD samples, respectively, as the validation set.

To ensure the quality and reliability of the downloaded data, quality control was performed, and the inclusion and exclusion criteria were (Hirsch et al., 2017) to remove samples with incomplete clinical information; (Siegel et al., 2022); to remove samples with unknown survival time and survival status; (Succony et al., 2021); to remove probes with one probe matching to multiple genes, and the mean value was taken as the expression value of that gene when multiple probes matched to one gene.

NK cell-associated genes were obtained from three parts, including the ImmPort official website (https://www.immport.org/resource), the MSigDB database (Molecular Signatures Database, https://ngdc.cncb.ac.cn/databasecommons/database/id/1077) and the LM22 database (Newman et al., 2015), containing 134 cell-associated genes, 18 NK cell-associated pathways, and 79 NK cell-associated genes, respectively.

2.2 Subtyping of LUAD patients based on NK cell-associated genes

A total of 213 NK cell-associated genes and 18 NK cell-associated pathways were obtained from the three databases, and we used the single sample gene set enrichment analysis (ssGSEA) method to evaluate these 213 NK cell-associated genes and 18 NK cell-associated pathways in the TCGA and GEO datasets, respectively. The samples were then clustered by ConsensusClusterPlus using these pathway scores in the TCGA and GEO cohorts, and the "K-M" algorithm and "1-Pearson correlation" as the metric distance (Azman et al., 2006). We conducted 500 bootstraps, with each one including 80% patients of in the training set and 20% those of the validation set. Finally, based on the cumulative distribution function (CDF), the optimal number clusters were decided, and the optimal classification and the sample molecular subtyping was obtained by calculating the consistency matrix and the consistency cumulative distribution function (Zhang et al., 2021a).

2.3 Immunological features and pathway analysis among different molecular subtypes

We obtained the molecular characteristics of LUAD genomic alterations from published literature, including LOH, Aneuploidy Score, tumor mutation burden (TMB), purity, and ploidy, Homologous Recombination Defects, Intratumor heterogeneity. The relative abundance of 22 immune cells were calculated using CIBERSORT R package. At the same time, we used the ESTIMATE algorithm R package to calculate the proportion of immune cells and finally compared the inflammatory and immune activity scores (Chakraborty and Hossain, 2018; Chen et al., 2018).

We performed gene set enrichment analysis (GSEA) on all NK cell-associated genes in the Hallmark database, and then used the ssGSEA method to calculate the pathway scores for both TCGA and GEO datasets in the GSVA package (Barbie et al., 2009). A false discovery rate (FDR) of <0.05 in this study was considered statistically significant.

2.4 Drug sensitivity analysis between molecular subtypes

Immune checkpoint inhibitor (ICI)-based therapy has become one of the standard treatments for advanced lung cancer (Zhang et al., 2021b). We first assessed the expression of genes associated with immunotherapy, such as CTLA4, PD-L1, and PD-1, among various molecular subtypes to determine whether there were differences in immunotherapy responsiveness among them. Next, we used the TIDE software (http://tide.dfci.harvard.edu/) to assess the potential clinical effects of immunotherapy in our defined molecular subtypes. Greater likelihood of immune escape was correlated with a higher TIDE prediction score, suggesting that patients may benefit less from immunotherapy (Jiang et al., 2018). Finally, we performed drug sensitivity prediction for LUAD in the "pRRophetic" package (Geeleher et al., 2014).

2.5 Identification of key NK cell-related genes among molecular subtypes

The differentially expressed genes among different molecular typing were calculated by the "limma" package, using FDR <0.05 and |log2FC| > 1 as the statistical difference criteria, and visualized the differentially expressed genes by "pheatmap" and "ggplot2" R packages in a heatmap and volcano map. Then, all genes with statistically significant differences were enriched using the "clusterProfiler" package.

Next, we performed univariate Cox regression analysis for differentially expressed genes between molecular subtypes, and then reduced the prognosis-related genes by Lasso regression (Sun et al., 2021), which can better solve the problem of multicollinearity in regression analysis by compressing some coefficients and setting some coefficients to zero at the same time. With the gradual increase of lambda, we selected the number of factors when the coefficients of independent variables tended to zero. Then, we used the AIC deficit pool information criterion through stepwise regression, which has the advantage of considering the statistical fit of the model and the number of parameters used to fit it, and at the same time indicates a sufficient fit of the model obtains with fewer parameters (Zhang, 2016).

2.6 Construction and validation of the prognostic model

We calculated the NK cell-related prognostic RiskScore for each sample according to the formula defined by the sample RiskScore (below), and normalized it (Nie et al., 2021).

$$\begin{aligned}\text{RiskScore} &= \text{coefficient1*gene1 expression} + \ldots \\ &\quad + \text{coefficientN*geneN expression.}\end{aligned}$$

After that, LUAD patients were divided into high- and low-risk groups based on the relationship between RiskScore and 0, where those with RiskScore >0 were considered as having a high risk and those with RiskScore <0 were considered as having a low risk. Finally, the survival differences between the two groups were compared by log-rank test. In order to verify the robustness of the model, we performed immune signature analysis, survival curve, and drug treatment difference analysis for the patients in the two groups.

2.7 Improvement of prognostic models and survival prediction in LUAD patients

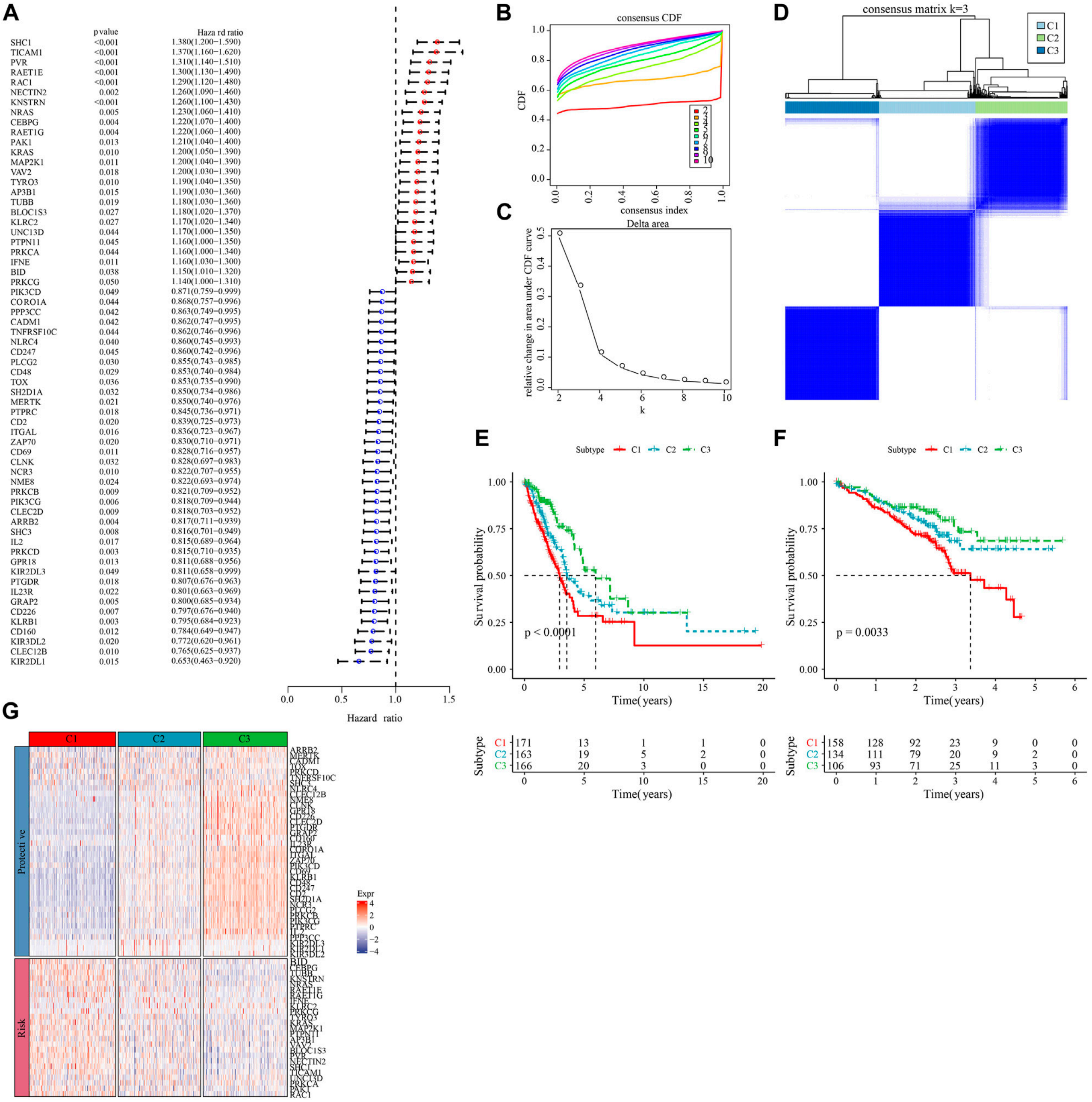
To more accurately quantify the risk assessment and survival probability of LUAD patients, we combined the RiskScore with other clinicopathological characteristics of LUAD patients and constructed a nomogram using the "nomogramEx" R package. To validate the accuracy of the model, a calibration curve was plotted by the "PredictABEL" function to visualize the goodness-of-fit. This was followed by decision curve analysis (DCA) to describe the change in net benefit as the threshold probability changed under the intervention of the predicted value by the model (Van Calster et al., 2016; Van Calster et al., 2018).

2.8 Programmed cell death (PCD) analysis

12 PCD patterns (apoptosis, necroptosis, pyroptosis, ferroptosis, cuproptosis, entotic cell death, netotic cell death, parthanatos, lysosome-dependent cell death, autophagy-dependent cell death, alkaliptosis, and oxeiptosis) have been taken from the previous research (Zou et al., 2022). ssGSEA analysis based on the expression data of PCD related genes using the R package GSVA. Spearman analysis was conducted to know the relationship among PCD patterns, clinical features, RiskScore in LUAD samples.

2.9 Statistical analysis

Unless otherwise specified, all statistical tests were bilateral and conducted using R software (version 4.1.3, https://www.r-project.org/), and $p < 0.05$ was considered statistically significant.



	p value	Haza rd ratio
SHC1	<0.001	1.380(1.200−1.590)
TICAM1	<0.001	1.370(1.160−1.620)
PVR	<0.001	1.310(1.140−1.510)
RAET1E	<0.001	1.300(1.130−1.490)
RAC1	<0.001	1.290(1.120−1.480)
NECTIN2	0.002	1.260(1.090−1.460)
KNSTRN	<0.001	1.260(1.100−1.430)
NRAS	0.005	1.230(1.060−1.410)
CEBPG	0.004	1.220(1.070−1.400)
RAET1G	0.004	1.220(1.060−1.400)
PAK1	0.013	1.210(1.040−1.400)
KRAS	0.010	1.200(1.050−1.390)
MAP2K1	0.011	1.200(1.040−1.390)
VAV2	0.018	1.200(1.030−1.390)
TYRO3	0.010	1.190(1.040−1.350)
AP3B1	0.015	1.190(1.030−1.360)
TUBB	0.019	1.180(1.030−1.360)
BLOC1S3	0.027	1.180(1.020−1.370)
KLRC2	0.027	1.170(1.020−1.340)
UNC13D	0.044	1.170(1.000−1.350)
PTPN11	0.045	1.160(1.000−1.350)
PRKCA	0.044	1.160(1.000−1.340)
IFNE	0.011	1.160(1.030−1.300)
BID	0.038	1.150(1.010−1.320)
PRKCG	0.050	1.140(1.000−1.310)
PIK3CD	0.049	0.871(0.759−0.999)
CORO1A	0.044	0.868(0.757−0.996)
PPP3CC	0.042	0.863(0.749−0.995)
CADM1	0.042	0.862(0.747−0.995)
TNFRSF10C	0.044	0.862(0.746−0.996)
NLRC4	0.040	0.860(0.745−0.993)
CD247	0.045	0.860(0.742−0.996)
PLCG2	0.030	0.855(0.743−0.985)
CD48	0.029	0.853(0.740−0.984)
TOX	0.036	0.853(0.735−0.990)
SH2D1A	0.032	0.850(0.734−0.986)
MERTK	0.021	0.850(0.740−0.976)
PTPRC	0.018	0.845(0.736−0.971)
CD2	0.020	0.839(0.725−0.973)
ITGAL	0.016	0.836(0.723−0.967)
ZAP70	0.020	0.830(0.710−0.971)
CD69	0.011	0.828(0.716−0.957)
CLNK	0.032	0.828(0.697−0.983)
NCR3	0.010	0.822(0.707−0.955)
NME8	0.024	0.822(0.693−0.974)
PRKCB	0.009	0.821(0.709−0.952)
PIK3CG	0.006	0.818(0.709−0.944)
CLEC2D	0.009	0.818(0.703−0.952)
ARRB2	0.004	0.817(0.711−0.939)
SHC3	0.008	0.816(0.701−0.949)
IL2	0.017	0.815(0.689−0.964)
PRKCD	0.003	0.815(0.710−0.935)
GPR18	0.013	0.811(0.688−0.956)
KIR2DL3	0.049	0.811(0.658−0.999)
PTGDR	0.018	0.807(0.676−0.963)
IL23R	0.022	0.801(0.663−0.969)
GRAP2	0.005	0.800(0.685−0.934)
CD226	0.007	0.797(0.676−0.940)
KLRB1	0.003	0.795(0.684−0.923)
CD160	0.012	0.784(0.649−0.947)
KIR3DL2	0.020	0.772(0.620−0.961)
CLEC12B	0.010	0.765(0.625−0.937)
KIR2DL1	0.015	0.653(0.463−0.920)

FIGURE 1
Molecular subtyping based on natural killer cell-associated genes. **(A)** Forest plot of prognostically significant natural killer cell-associated genes in the TCGA-LUAD cohort. **(B)** CDF curves of the TCGA-LUAD cohort. **(C)** CDF Delta area curves of the TCGA-LUAD cohort. **(D)** heat map of sample clustering at consensus $k = 3$ in the TCGA-LUAD cohort. **(E)** KM curves of the relationship between overall survival (OS) prognosis of the three subtypes in the TCGA-LUAD cohort. **(F)** Prognostic differences between the three molecular subtypes in the GSE72094 cohort. **(G)** Heatmap of prognosis significant natural killer cell genes expression in different subtypes of TCGA-LUAD.

3 Results

3.1 Molecular subtyping of LUAD based on NK cell-associated genes

We first calculated the NK cell-related genes showing close relationship with LUAD survival chance by univariate Cox regression analysis, and screened 63 prognostically significant genes ($p < 0.05$, Figure 1A), including the prognostic (Protective) genes SHC1, TICAM1, PVR, RAET1E, RAC1 (HR > 1), and KLRB1, CD160, KIR3DL2, CLEC12B, and KIR2DL1 (HR < 1). Then, we used these 63 differential genes for consistent clustering, and determined the best cluster number according to CDF. And we could see from Figures 1B, C that Cluster = 3 had more stable clustering results, hence, $k = 3$ was selected to obtain three molecular subtypes (C1, C2, and C3) (Figure 1D). Then, we

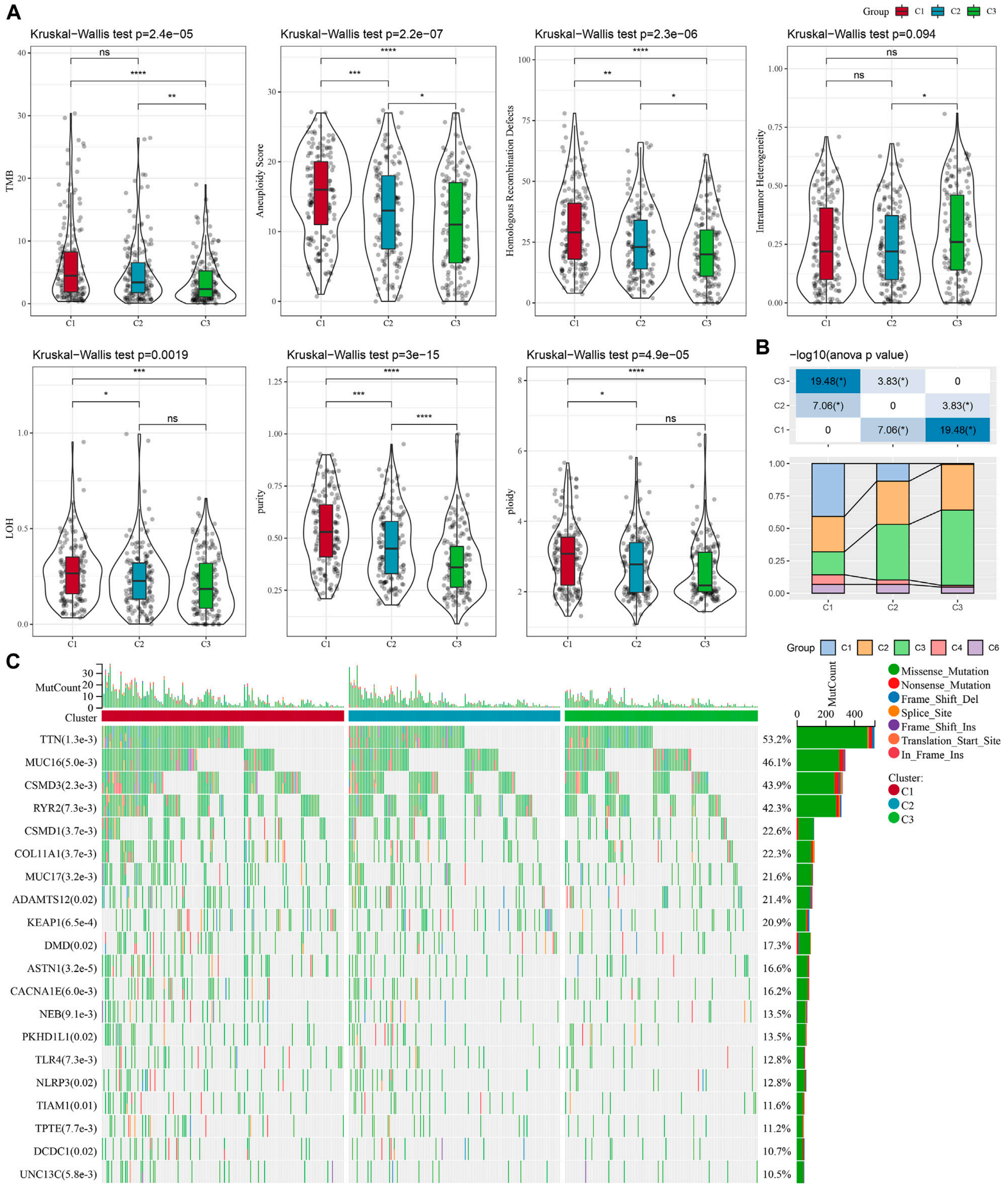


FIGURE 2
Genomic alterations in the molecular subtypes of the TCGA cohort. **(A)** Comparison of differences in Homologous Recombination Defects, Aneuploidy Score, Fraction Altered, Number of Segments, and Tumor mutation burden in the TCGA cohort molecular subtypes. **(B)** Comparison of the three molecular subtypes with immune molecular subtypes. **(C)** Somatic mutations in the three molecular subtypes (chi-square test). **p* < 0.05; ***p* < 0.01; ****p* < 0.001; *****p* < 0.0001.

performed survival analysis of patients with these three molecular subtypes using the K-M survival method, and the results identified a significant difference in prognosis among the three molecular subtypes, with C1 patients having the worst prognosis and C3 patients having the optimal prognosis (Figure 1E). The results were also validated in the GSE72094 dataset (Figure 1F). Meanwhile, we found that the "Risk" genes were high-expressed in the C1 subtype and the "Protective" genes were high-expressed in the C3 subtype in the heat map (Figure 1G). These results suggested that the molecular subtyping based on NK cell-related genes was reasonable, and there were significant differences in gene expression and prognosis among patients with different subtypes.

3.2 Genetic landscape between molecular subtypes of LUAD

To explore the differences in specific gene expression profiles among different molecular subtypes, we compared the molecular profiles among C1, C2, and C3 subtypes of LUAD samples, respectively, and it is obvious from Figure 2A that purity, and ploidy, TMB, Aneuploidy Score, LOH, Homologous Recombination Defects expression were the highest in C1 but the lowest in C3, which was consistent with previous studies (Thorsson et al., 2018). In addition, we compared the differences between the molecular subtyping of published studies and that in this study. Here it was found that the C3 subclass occupied the most of the C3 subtypes we defined, suggesting that the C3 subtype was the major subtype of LUAD (Figure 2B). In addition, a significant correlation between molecular subtypes and gene mutations was detected after analyzing the correlation between gene mutations and molecular subtypes, and observed. TTN, MUC16, CSMD3, and RYR2 were the most widely mutated genes in LUAD (Figure 2C), and this finding indicated that the development of LUAD was closely related to the above-mentioned gene mutations.

3.3 Pathways enrichment analysis among the molecular subtyping of LUAD

To investigate pathway differences in LUAD among different molecular subtypes, we performed GSEA enrichment analysis among molecular subtypes. As shown in Figure 3A, we enriched a total of 33 significant pathways in the TCGA-LUAD dataset, including MYC_TARGETS_V2, E2F_TARGETS, NFLAMMATORY_ RESPONSE, MYOGENESIS, INTERFERON_GAMMA_RESPONSE, MYC_TARGETS_V1, GLYCOLYSIS, G2M_CHECKPOINT, EPITHELIAL_MESENCHYMAL_TRANSITION, ALLOGRAFT_REJECTION, suggesting that these NK cell genes were mainly associated with cell cycle and immunity in C1 and C3. Additionally, pathways different between C1 and C3 subtypes, between C2 and C3 subtypes, between C1 and C2, were analyzed (Figure 3B). Overall, the cell cycle pathway was activated in C1 patients, while the immune-related pathway was suppressed, therefore we hypothesized that these NK cell genes might play an important role in the cell cycle pathway as well as in the tumor microenvironment. To validate these results, we presented the pathway differences between C1 and C2, and C2 and C3 as radar plots, and the results showed that they both had significant consistency in cell cycle (MYC_TARGETS_V2, MTORC1_SIGNALING, MYC_TARGETS_V1) and immune-related pathways (G2M_CHECKPOINT, E2F_TARGETS, UNFOLDED_PROTEIN_RESPONSE) (Figure 3C).

3.4 Immune characteristics among different molecular typologies of LUAD

The immune system plays a dual role in the development of LUAD, as it can recognize and destroy tumor cells, while tumor cells can also evade host immune attack by forming a complex immunosuppressive network under the pressure of immune selection using the immune system's own negative regulatory mechanisms, thus the TME is in a constant state of change (Anichini et al., 2020; Spella and Stathopoulos, 2021). To explore the immune landscape among different molecular subtypes of LUAD, we first assessed the differences in the components of immune cells in the TCGA-LUAD cohort using the CIBERSORT algorithm and observed that most immune cells (B cells, T cells, NK cells, *etc.*) were significantly different ($p < 0.05$) (Figure 4A). We then used the ESTIMATE algorithm to assess immune cell infiltration, and the results showed that StromalScore, ImmuneScore and EstimateScore were significantly different between C1, C2, and C3 ($p < 0.05$)), with ImmuneScore accounting for the largest proportion of C3 types, suggesting a higher degree of immune infiltration in C3 subtypes (Figure 4B). Similarly, we obtained results in the GSE72094-LUAD cohort that were consistent with the TCGA-LUAD cohort (Figures 4C, D). In addition, we assessed the inflammatory activity of C3, C2, C1, except for IgG, the remaining six out of 7 metagenes clusters (HCK, Interferon, LCK, MCH I, MCH II, and STAT1) showed significantly different enrichment scores, with the C4 subtype having higher inflammatory activity (Figure 4E). The findings were consistent in the GSE72094-LUAD cohort (Figure 4F).

3.5 Differences in immunotherapy between molecular subtypes

In recent years, immunotherapy has led to new opportunities in the treatment of small cell lung cancer. Clinical trials of some immune checkpoint inhibitors have demonstrated their efficacy and safety in LUAD (Hua et al., 2021). Based on this, we first evaluated the expression of the representative molecules of immunotherapy (PD-1, PD-L1, CTLA4) among the three molecular subtypes, and observed that PD-1, PD-L1, and CTLA4 were significantly more expressed in C3 subtype ($p < 0.05$) (Figure 5A). We also applied the "T-cell-inflamed GEP

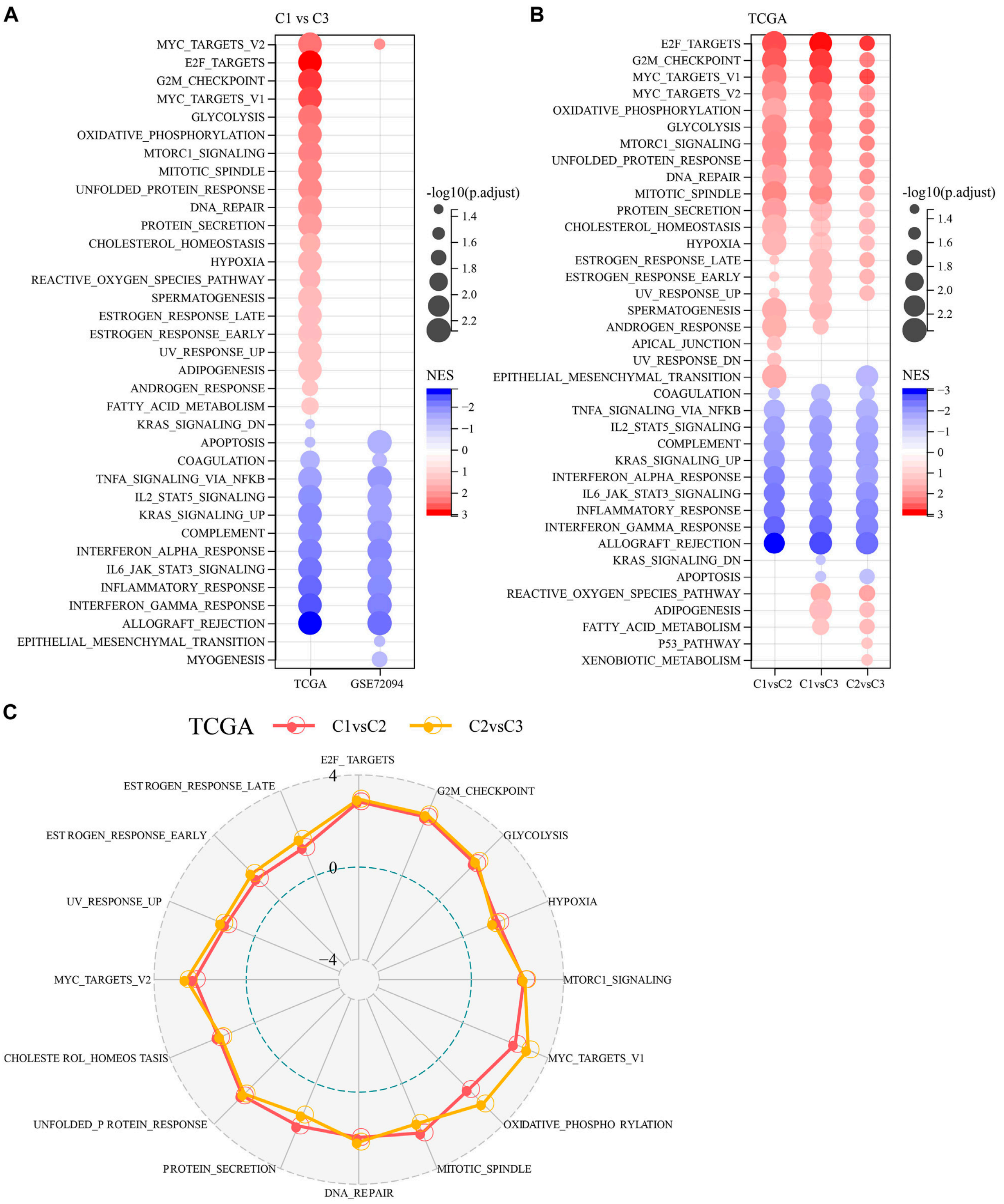


FIGURE 3
Pathway analysis between molecular subtypes. **(A)** Bubble plots of GSEA results for C1 vs. C3 subtypes in two lung adenocarcinoma cohorts. **(B)** Bubble plots of GSEA results for different molecular subtypes compared to each other in the TCGA-LUAD cohort. **(C)** Radar plots of C1 vs. C2 and C2 vs. C3 activation pathways in the TCGA-LUAD cohort.

score” to assess the predictive potential of different molecular subtypes to cancer immunotherapy, and the results also showed that the score was highest in C3 (Figure 5B). Considering that IFN-γ is a cytokine that plays a key role in immunomodulation and immunotherapy, we downloaded the GOBP_RESPONSE_TO_INTERFERON_GAMMA gene set from the GO database for

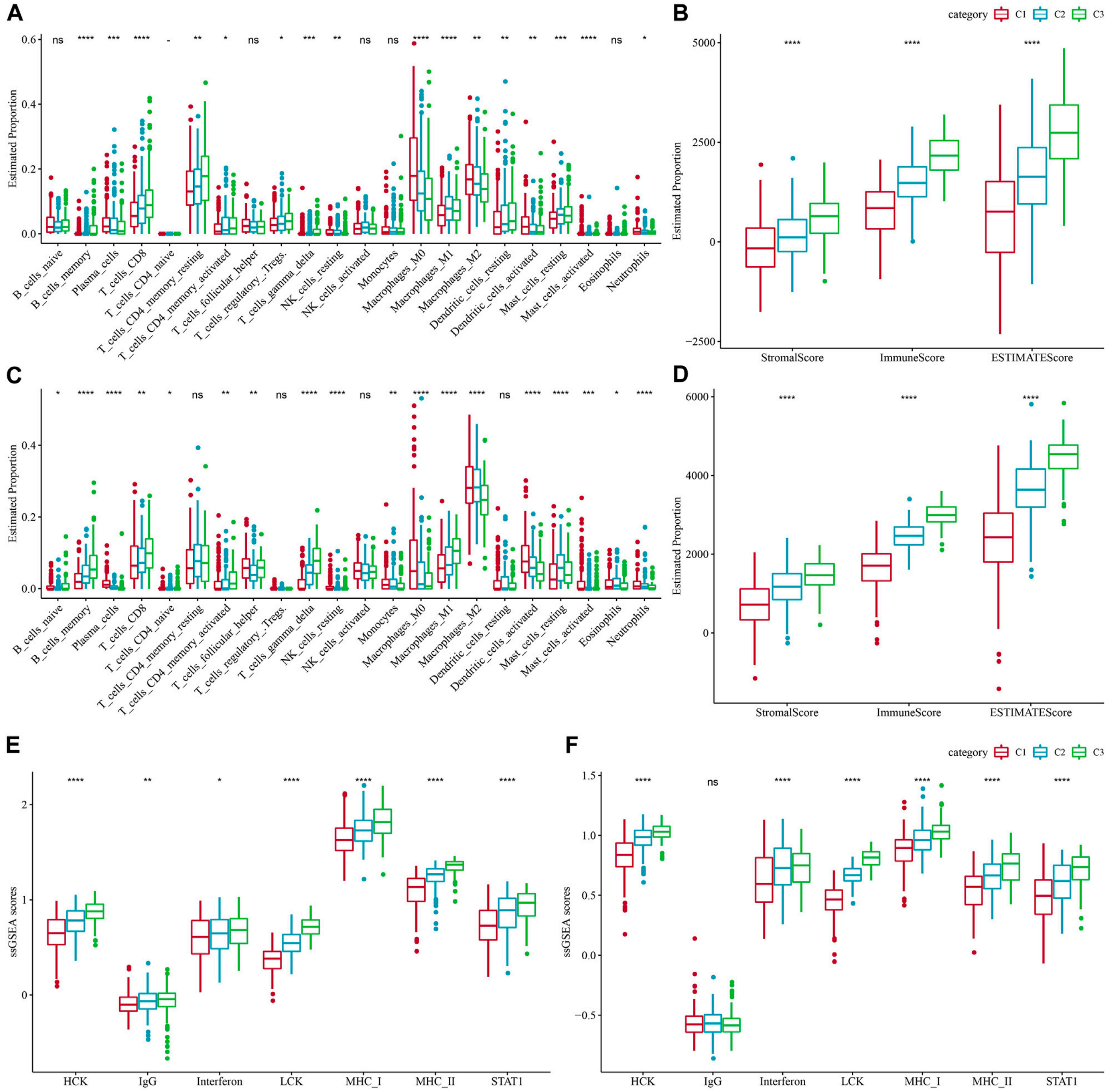


FIGURE 4
Proportions of immune cell components in the two lung adenocarcinoma cohorts. **(A)** Differences in 22 immune cell scores between different molecular subtypes in the TCGA-LUAD cohort. **(B)** Differences in ESTIMATE immune infiltration between different molecular subtypes in the TCGA-LUAD cohort. **(C)** Differences in the GSE72094 cohort 22 immune cell scores between different molecular subtypes. **(D)** Differences in ESTIMATE immune infiltration between different molecular subtypes in the GSE72094 cohort. **(E)** Differences in seven inflammation-related gene cluster scores across molecular subtypes in TCGA-LUAD cohort. **(F)** Differences in gene cluster scores between different molecular subtypes in seven inflammatory-related genes in GSE72094 cohort.

ssGSEA analysis, and found that the IFN-γ response was significantly enhanced in the C1 subtype (Figure 5C). We also compared the differences in INFG gene expression in the three subtypes and found that INFG was noticeably high-expressed in the C3 subtype (Figure 5D). Moreover, CYT score, which reflects the cytotoxic effect, was significantly higher in the C3 subtype than in the other subtypes (Figure 5E). In addition, the TIDE prediction data indicated that the C1 subtype had a higher TIDE score, suggesting that the C1 subtype was less likely to benefit from immunotherapy (Figure 5F). The estimated IC50 of docetaxel, vincristine, paclitaxel and cisplatin among 3 subtypes showed that C1 was more sensitive to the four chemotherapy drugs (Figure 5G). The above results indicated that predicting immunotherapy for LUAD based on NK cell-related genes was a practical approach.

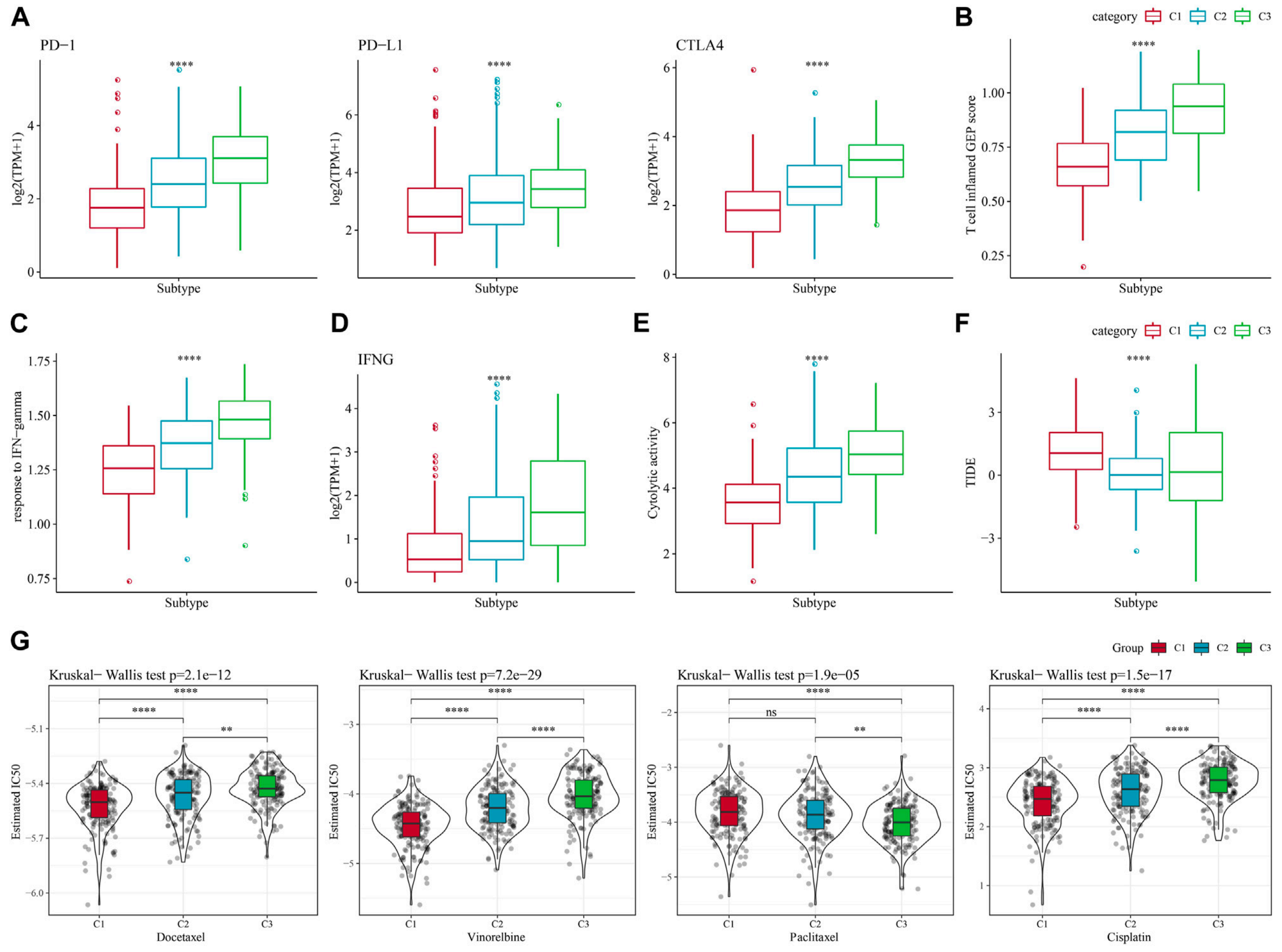


FIGURE 5
Differences in immunotherapy/chemotherapy treatment between molecular subtypes. **(A)** Differences in expression of immune checkpoint genes between molecular subtypes. **(B)** difference in "T cell inflamed GEP score" between molecular subtypes. **(C)** Differences in "response to IFN-gamma" between different molecular subtypes. **(D)** Differences in INFG gene expression in different isoforms. **(E)** Differences in "Cytolytic activity" between different molecular subtypes. **(F)** Differences in TIDE scores between subtypes. **(G)** Box plots of estimated IC50 of docetaxel, vincristine, paclitaxel and cisplatin in TCGA-LUAD.

3.6 The analysis of PCD patterns among molecular subtypes

The ssGSEA analysis calculated the score of 12 PCD patterns in each sample in TCGA dataset and GSE72094 dataset. We found that 9 PCD scores had differences among 3 subtypes in both two datasets (Figures 6A, B). In TCGA dataset, Stage, Gender, especially, Age had closely associated to PCD patterns (Figure 6C), but in GSE72094 dataset, clinical features had litter associated to PCD patterns (Figure 6D). Autophagy score were increased in early Stage, the scores of Pyroptosis, Autophagy, Necroptosis and Oxeiptosis were enhanced in Male samples, and samples with age > 60 had higher Pyroptosis, Entotic. cell.death scores in TCGA dataset (Figure 6E). In GSE72094 dataset, Oxeiptosis score was highest in StageIII, and Ferroptosis and Necroptosis scores were greater in patients with age>60 (Figure 6F).

3.7 Establishment of LUAD risk model

We first calculated the NK cell-related genes significantly differentially expressed among the three molecular subtypes by the limma package, significant expression differences of NK cell-related genes among C1, C2, and C3 were detected, including 11 upregulated genes and 180 downregulated genes (Supplementary Figures S1A, B). Differentially expressed downregulated genes were related to immune-related pathways, as shown by the results of enrichment analysis (Supplementary Figure S1C). Genes with upregulated level were related to inflammatory and immune pathways (Supplementary Figure S1D). 173 genes with high prognostic impact ($p < 0.05$), including 159"Protective" and 14"Risk" genes, were identified from those genes by conducting one-way Cox regression analysis (Supplementary Figure S2A). Further, we observed the trajectory of each gene with lambda using Lasso analysis, and the model was optimal when lambda = 0.0382, which corresponded to 9 differential genes (Supplementary Figures S2B, C). After that, we reduced the

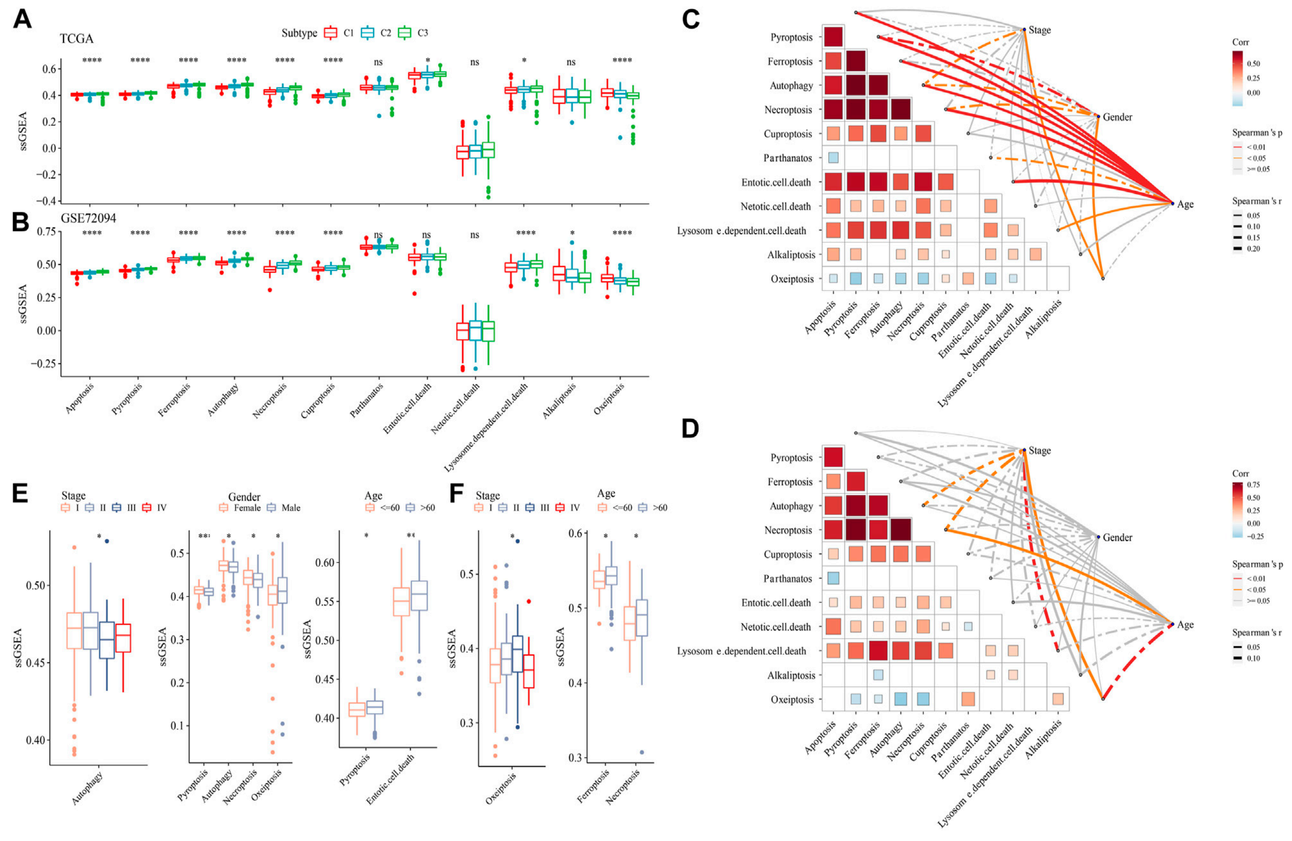


FIGURE 6
The PCD characteristic among 3 molecular subtypes. **(A)** the ssGSEA analysis of 12 PCD patterns among 3 molecular subtypes in TCGA-LUAD dataset. **(B)** the ssGSEA analysis of 12 PCD patterns among three molecular subtypes in GSE72094 dataset. **(C)** The spearman analysis between clinical feature and PCD in TCGA-LUAD dataset. **(D)** The spearman analysis between clinical feature and PCD in GSE72094 dataset. **(E)** The ssGSEA analysis of PCD in TCGA-LUAD samples with Stage, Gender and Age. **(F)** The ssGSEA analysis of PCD in GSE72094 samples with Stage, and Age.

genes to four, namely, ANLN, FAM83A, RHOV, and PARP15, by the stepAIC method in the MASS package (Supplementary Figure S2D).

Then, we calculated the Riskscore score for each TCGA-LUAD patient using these four genes and the above formula (Figure 7A). We classified those RiskScore with 0 ≤ as low-risk group and with RiskScore >0 as high-risk group. Then, we performed a prognostic classification ROC analysis in the "timeROC" package for analyzing 1-year, 2-year, 3-year, and 5-year prognostic prediction classification efficiency, and we found that the model had a high AUC (0.71, 0.69, 0.7, and 0.67) (Figure 7B). The results of survival analysis showed that patients in the low-risk group developed a significantly better prognosis ($p < 0.001$) (Figure 7C). To confirm the robustness of this clinical prognostic model, we validated it in the GSE72094 and GSE31210 cohorts and used the same approach to calculate the RiskScore of patients (Figures 7D–G).

3.8 Pathological characteristics of high- and low-risk groups

To investigate the reliability of this risk model classification method, we first compared the clinical characteristics of patients in both high- and low-risk groups. The results showed that the RiskScore scores of patients with Stage III-IV, M Stage, N Stage, T Stage were significantly higher than Stage I-II ones. In addition, we also found that male patients had a higher RiskScore (Figure 8A). Also, we compared the differences in RiskScore by molecular subtype and found that the RiskScore for the C1 subtype with poorer prognosis was significantly higher than C3 with a better prognostic outcome (Figure 8B), and that the majority of the samples with high RiskScore were "C1" patients (Figure 8C). In addition, we also compared whether there was a prognostic difference in the—high- and low-risk groups between the different clinicopathological characteristics subgroups in the TCGA-LUAD cohort. Across different clinical subgroups, the risk grouping performed equally well, pointing to the reliability of the grouping (Figure 8D). This finding also applied to the GSE72094-LUAD cohort (Supplementary Figure S3).

3.9 Immune infiltration and pathway characteristics of low-risk and high-risk patients

We compared the relative abundance of 22 immune cell types in the two subgroups of the TCGA-LUAD cohort in high- and low-risk

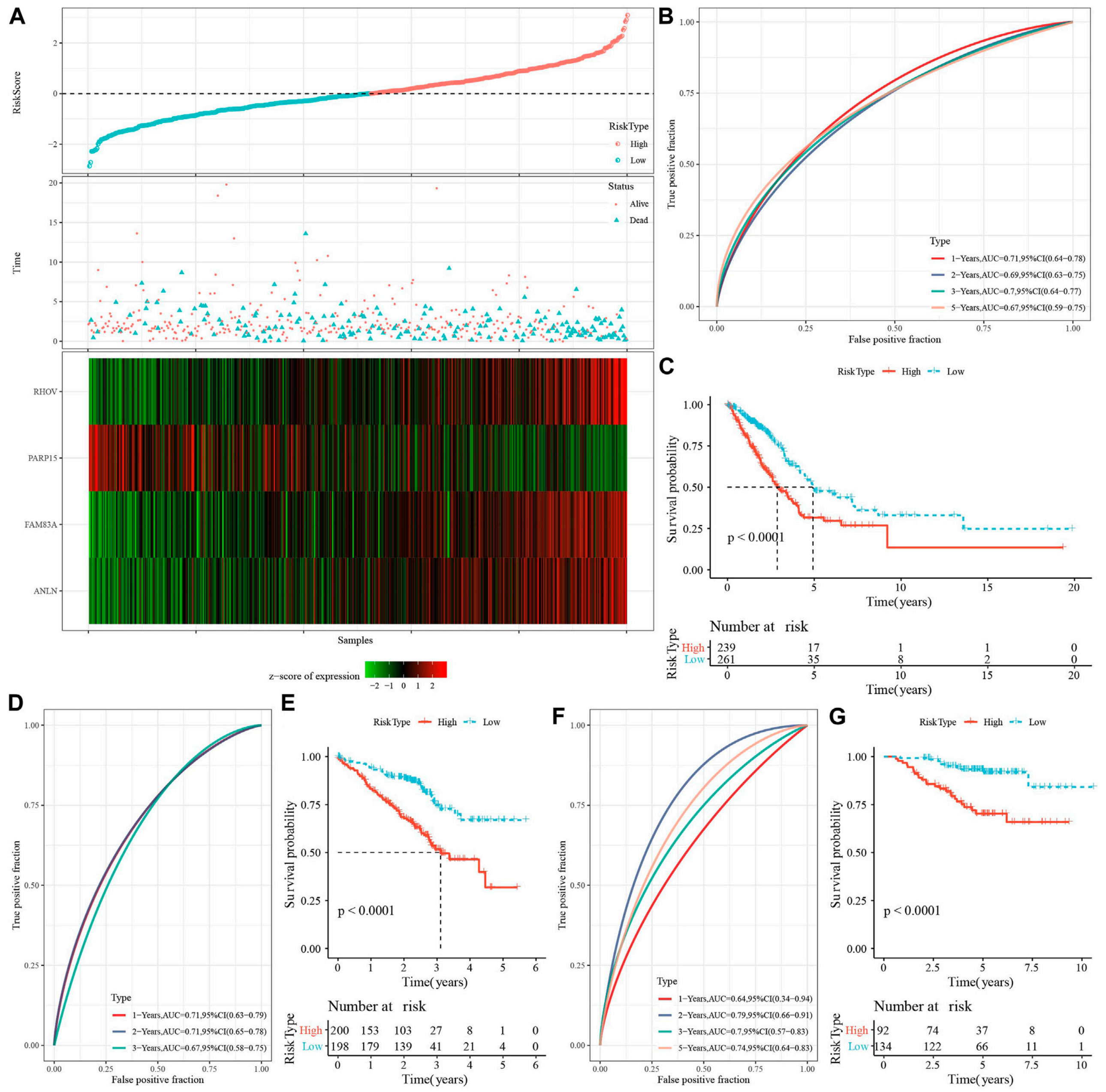


FIGURE 7
Risk modeling and validation. **(A)** RiskScore, survival time vs. survival status and expression of necroptosis-related genes in TCGA-LUAD dataset. **(B)** ROC curves with AUC for RiskScore classification in the TCGA-LUAD dataset. **(C)** Distribution of KM survival curves for RiskScore in the TCGA-LUAD dataset. **(D,E)**: ROC curves and KM survival curves for RiskScore in the GSE72094 cohort. **(F,G)**: ROC curves and KM survival curves of RiskScore in the GSE31210 cohort.

groups. We discovered that the majority of immune cells (B cells, macrophages, T cells, and mast cells) were significantly different in high- and low-risk groups (p 0.05, Figure 9A). It is worth noting that activated NK cells had no significance between high- and low-group. We also examined the connection between the RiskScore and 22 immune cell components (Figure 9B). Also, we assessed the immune cell infiltration using the ESTIMATE method. The three scores were significantly different between two risk groups ($p < 0.05$), and the low-Riskscore group had higher immune infiltration (Figure 9C). The relationship between biological function in different samples with RiskScore was analyzed by "ssGSEA" analysis and found that the high risk group was significantly enriched to some cell cycle-related pathways, such as HALLMARK_SPERMATOGENESIS, and HALLMARK_REPAIR, SPERMATOGENESIS, HALLMARK_DNA_REPAIR, ALLMARK_ MYC_TARGETS_V2, HALLMARK_UNFOLDED_PROTEIN_ RESPONSE, *etc.* (Figure 9D). Further, we selected functional pathways with correlations greater than 0.4, from which we could see that RiskScore showed positive correlation with cell cycle-related pathways, such as HALLMARK_MYC_TARGETS_V1,

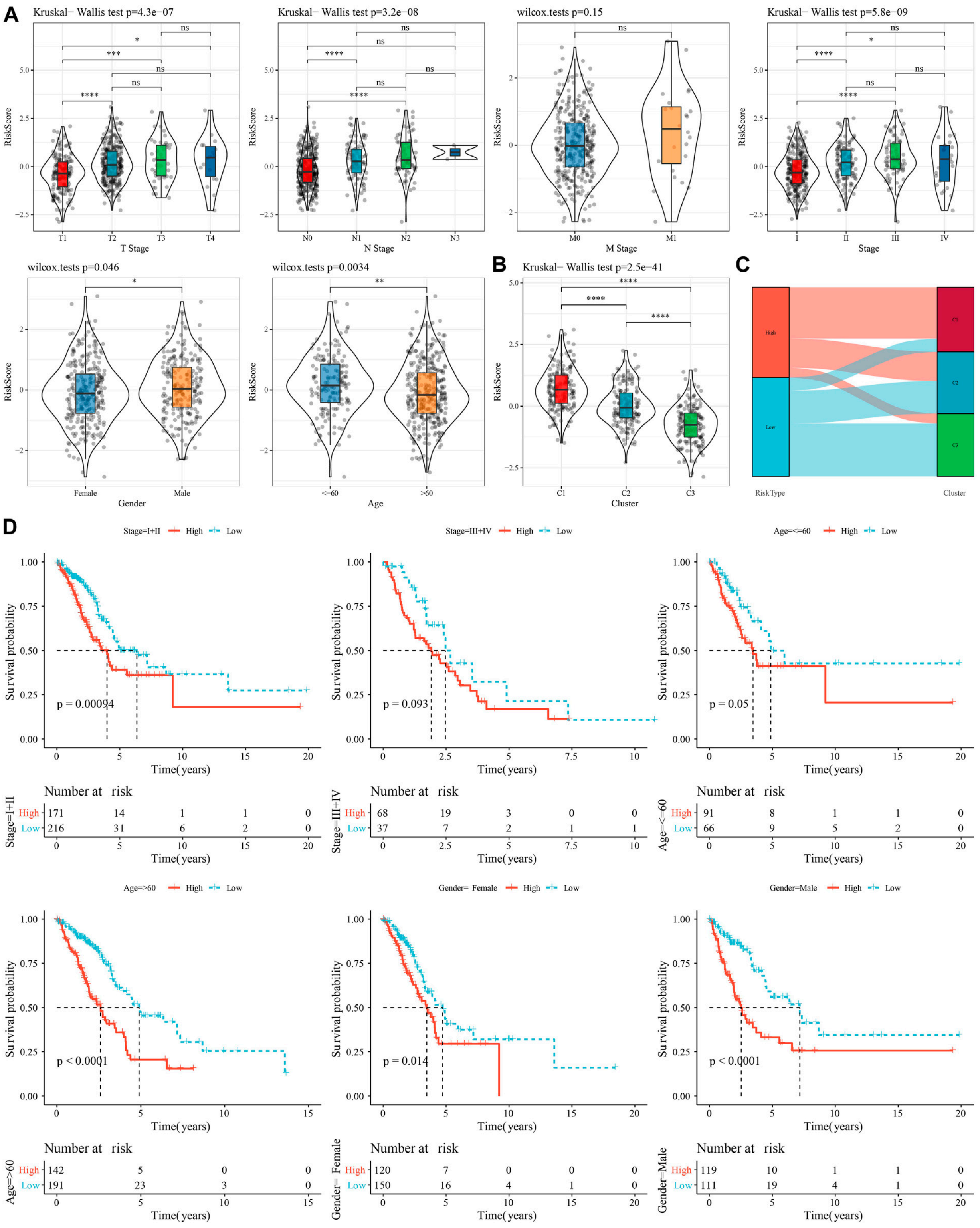


FIGURE 8
Performance of RiskScore in TCGA-LUAD cohort with different clinicopathological characteristics. **(A)** Differences between RiskScore between different clinicopathological subgroups in the TCGA-LUAD cohort. **(B)** Differences in RiskScore between different molecular subtypes in the TCGA-LUAD cohort. **(C)** Differences between molecular subtypes and RiskScore subgroups in the TCGA-LUAD cohort. **(D)** KM curves between high- and low-risk groups in the TCGA-LUAD cohort between different clinicopathological subgroups.

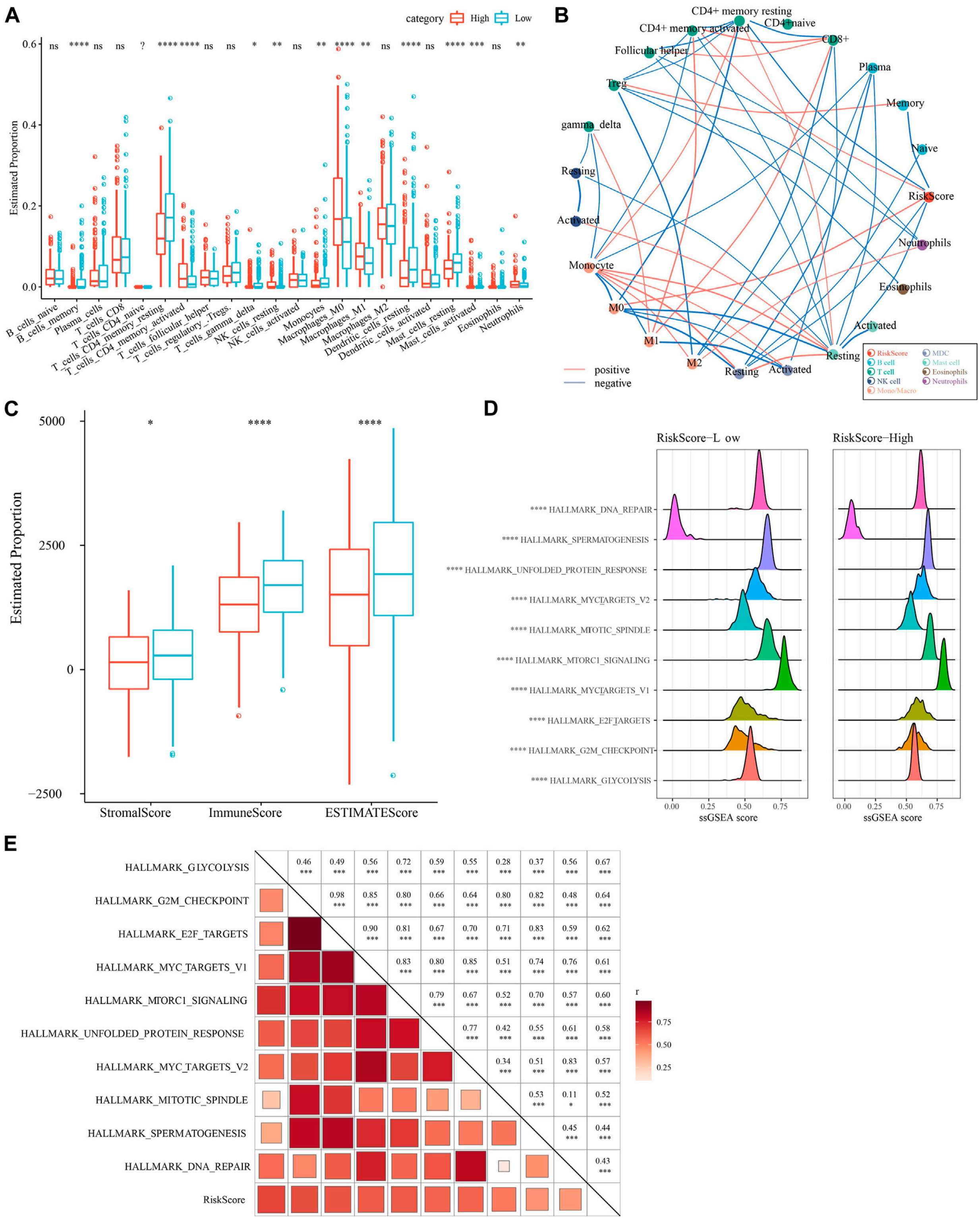


FIGURE 9
Immune infiltration/pathway characteristics between RiskScore subgroups. **(A)** Proportion of immune cell components in the TCGA cohort. **(B)** Correlation analysis of 22 immune cell components in the TCGA cohort with the RiskScore. **(C)** Proportion of immune cell components in the TCGA cohort calculated by ESTIMATE software. **(D)** The top 10 pathways with the most significant differences between high- and low-groups. **(E)** Results of correlation analysis between the RiskScore and KEGG pathways with correlation greater than 0.4.

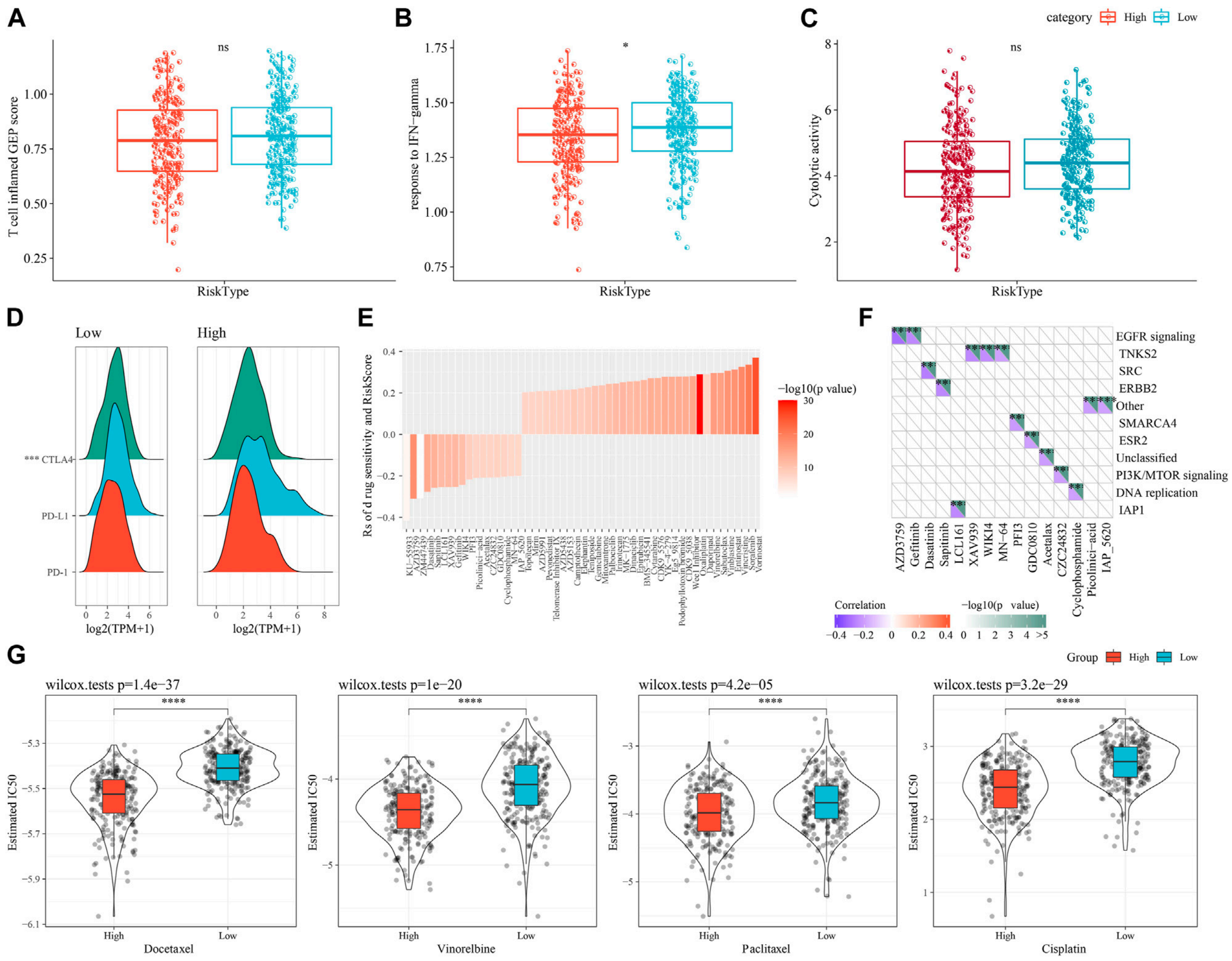


FIGURE 10
Differences in immunotherapy/chemotherapy between RiskScore subgroups. **(A)** Difference in "T cell inflamed GEP score" between molecular subtypes. **(B)** Difference in "response to IFN-γ" between molecular subtypes. **(C)** Differences in "Cytolytic activity" between molecular subtypes. **(D)** Differences in expression of immune checkpoint genes between molecular subtypes. **(E)** 15 pairs drugs were significantly associated to RiskScore. **(F)** 15 pairs drugs mainly target EGFR signaling and TNKS2 pathways. **(G)** IC50 box plots of docetaxel, vincristine, paclitaxel and cisplatin in TCGA-LUAD dataset.

HALLMARK_MTORC1_SIGNALING, HALLMARK_E2F_TARGETS (Figure 9E).

3.10 Differences in immunotherapy/chemotherapy for patients in high- and low-risk groups

First, we used the "T-cell-inflamed GEP score" to assess the predictive potential of the different RiskScore subgroups in cancer immunotherapy. The results showed that the "T-cell-inflamed GEP score" was elevated in the low-risk group, but the difference was not statistically significant (Figure 10A), however, in the low-risk group the IFN-γ response was noticeably elevated (Figure 10B). The CYT score, which reflects cytotoxic effects, was elevated in the low-risk group, showed no statistically significant differences (Figure 10C). The expression of representative molecules of immunotherapy (CTLA4, PD-L1, and PD-1) was calculated in the risk groups and showed that CTLA4 was significantly more expressed in the low-risk group ($p < 0.05$), while the difference in PD-1 and PD-L1 expression was not significant (Figure 10D). We looked at the connection between RiskScore and medication response in cancer cell lines to better understand the impact of RiskScore on drug response. We found 49 substantially linked relationships between RiskScore and drug sensitivity in the Genomics of Drug Sensitivity in Cancer (GDSC, http://cancer.sanger.ac.uk/cell_lines#) database using Spearman correlation analysis. Of these 49 pairs, 15 pairs were significantly associated with Riskscore correlations, such as Vinorelbine, Sabutoclax, Vinblastine, Entinostat, Vincristine, and Sorafenib (Figure 10E). We found that these drugs mainly target the EGFR signaling and TNKS2 pathways through the study on the signaling pathways of the

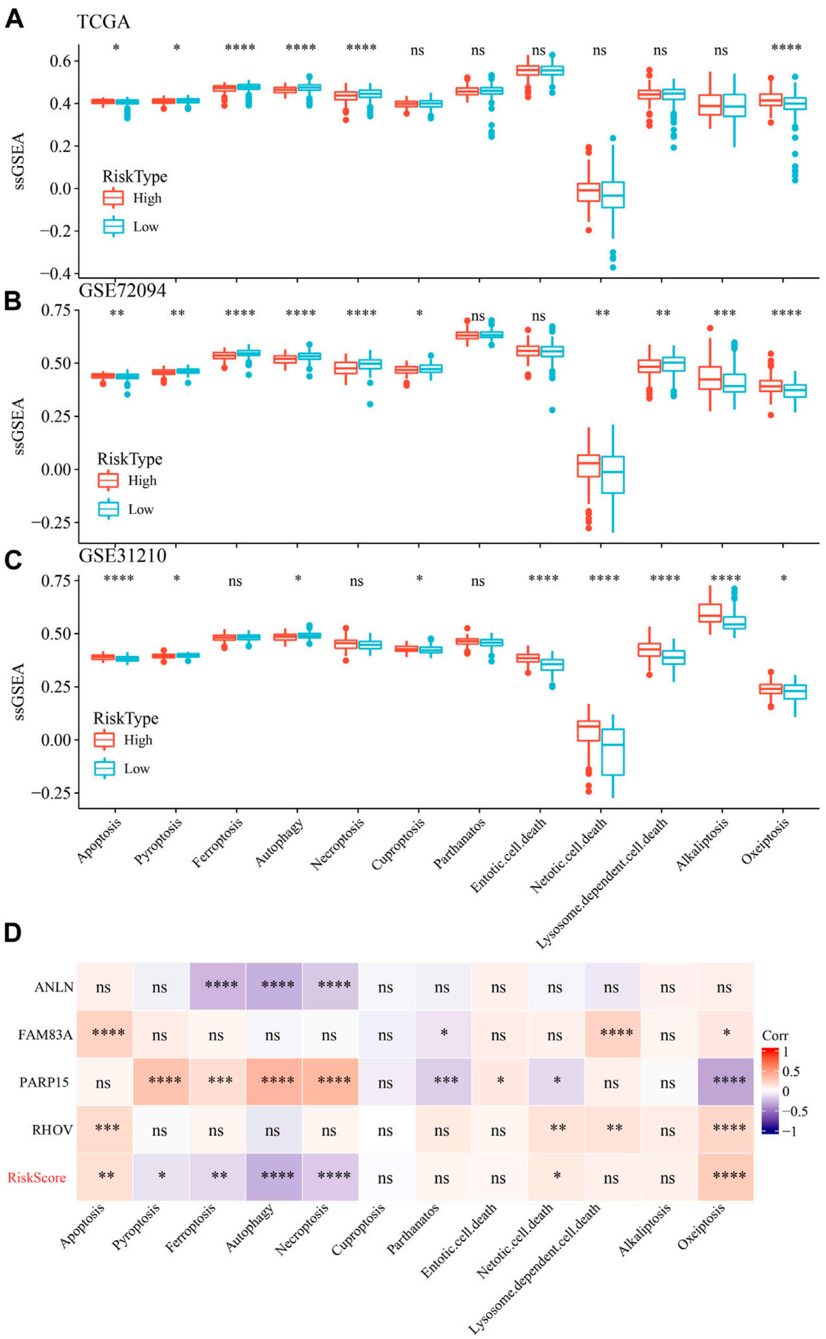


FIGURE 11
The ssGSEA analysis of 12 PCD patterns in high- and low-groups. **(A)** ssGSEA analysis of 12 PCD in high- and low-group in TCGA-LUAD dataset. **(B)** ssGSEA analysis of 12 PCD in high- and low-group in GSE72094 dataset. **(C)** ssGSEA analysis of 12 PCD in high- and low-group in GSE31210 dataset. **(D)** the association among RiskScore, model genes and 12 PCD patterns.

genes targeted by these drugs (Figure 10F). In addition, we also explored the response of different molecular subtypes in the TCGA-LUAD cohort to the traditional chemotherapeutic agents Docetaxel, Vinorelbine, Paclitaxel and Cisplatin, and found that overall patients in the high-risk group were more sensitive to all the four chemotherapeutic agents (Figure 10G), suggesting that patients in the high-risk group may benefit from these four drugs.

3.11 PCD characteristics in high- and low-risk groups

We also determine the PCD characteristics in high- and low-risk groups using ssGSEA analysis. 6 of 12 PCD styles had differences between high- and low-risk groups in TCGA dataset (Figure 11A). In GSE72094 dataset, 10 PCD patterns scores presented

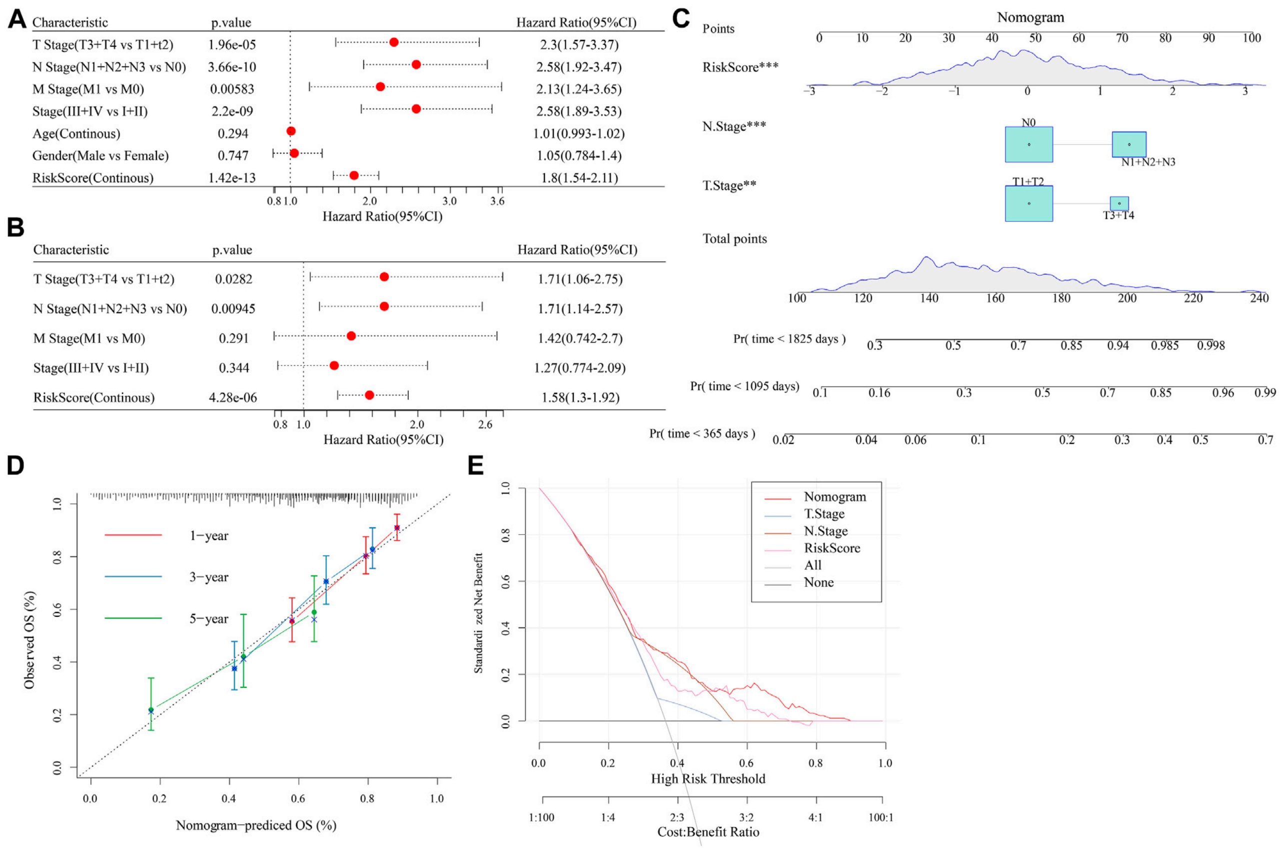


FIGURE 12
Establishment of nomogram. **(A,B)**: Univariate and Multivariate Cox analysis of RiskScore and clinicopathological characteristics; **(C)**: The nomogram model; **(D)**: Calibration curves for 1, 3, and 5 years for the nomogram; **(E)**: Decision curves for the nomogram.

differentiation in high- and low-risk groups (Figure 11B). Moreover, the differences of 9 PCD scores between high- and low-groups was observed in GSE31210 dataset (Figure 11C). RiskScore as well as four model genes were obviously related to PCD patterns (Figure 11D).

3.12 RiskScore combined with clinicopathological features to further improve prognostic models and survival prediction

Univariate and multifactorial Cox regression analyses revealed RiskScore as the most significant prognostic factor (Figures 12A, B). We created a nomogram (Figure 12C) combining RiskScore and other clinicopathological traits for the risk assessment and prediction of survival probability for LUAD patients. The model results revealed that RiskScore had the biggest influence on survival prediction. The prediction calibration curves at the three calibration points of 1, 3, and 5 year(s) nearly overlapped with the standard curve, which indicated that the nomogram plot had excellent prediction performance. We further assessed the prediction accuracy of the model using the calibration curve (Figure 12D). We also used DCA (Decision curve) to test the model's dependability, and it was shown that RiskScore and Nomogram performed much better than the extreme curve and had the strongest ability to predict survival among other clinicopathological factors (Figure 12E).

4 Discussion

Lung cancer is currently the most aggressive malignancy in the world, of which LUAD is the most common histological subtype of primary lung cancer, accounting for 64% of peripheral lung cancers, and has been reclassified from invasive precancerous lesions to invasive adenocarcinoma (Denisenko et al., 2018; Hutchinson et al., 2019). Despite the current advances in the treatment of LUAD, the median survival is only 8.6 months and immune escape is considered one of the main factors leading to treatment failure in LUAD (Yotsukura et al., 2021). In contrast to the remarkable efficacy of immune checkpoint inhibitor (ICI) in metastatic melanoma, Hodgkin's lymphoma, and bladder cancer, not all patients with LUAD are sensitive to ICI (Zhang et al., 2020). Mechanisms of immune escape that lack adaptive immune response include hypoxia-driven immunosuppressive factors, anti-apoptotic pathways, chronic inflammation, metabolic damage, and immune cells such as regulatory T (Treg) cells, tumor-associated M2 macrophages (TAM), myeloid-derived suppressor cells (MDSC) (Yu et al., 2021). Recent studies have shown that T and NK cell dysfunction and depletion or deficiency of antitumor-specific effector cells are involved in LUAD immune escape (Hong et al., 2019), and although the exact mechanism is unclear, it points to new ideas for the study of immune escape in LUAD and provides new targets for immunotherapy in LUAD.

LUAD is usually resistant to chemotherapy and/or radiotherapy and leads to the development of distant metastases (Jiang et al., 2021). NK cell dysfunction and failure in patients with LUAD could be caused by immune escape mechanisms mediated by lung cancer cells or tumor microenvironment, leading to failure of immunotherapy. The reason for this is related to tumor upregulation of inhibitory ligands (e.g., HLA-C molecules) and recognition by autoinhibitory KIR receptors carrying ITIM motifs (Daëron et al., 2008). Cellular experiments showed that other inhibitory receptors, for instance, KLRG-1, LAG-3, CD94/NKG2A, TIM3, TIGIT, and their ligands were also frequently upregulated on NK cells from LUAD patients (Lee et al., 1998; Nayyar et al., 2019), which was consistent with our study, where we found significantly different NK cell-related gene expression in different subtypes. CTLA-4 (ipilimumab) improved clinical prognosis of patients with LUAD (Paulsen et al., 2017) in addition to the common PDL-1 inhibitors (avelumab, atezolizumab, durvalumab) and PD-1 (camrelizumab, spartalizumab, nivolumab, pembrolizumab). Our study identified the expression patterns of PD-1/PD-L1 and CTLA-4 in different subtypes, confirming a possible immune escape mechanism of NK cells in LUAD and providing a new perspective for blocking immune dysregulation.

The tumor microenvironment (TME) consists of associated fibroblasts (CAF), tumor cells, other immune cells, and endothelial cell constituents (ECs) (Vitale et al., 2019). Ghiringhelli F et al. showed that suppressive immune cells such as Treg cells, CTLA-4+ regulatory, and that N2 neutrophils and M2 macrophages can disrupt the anti-lung cancer activity of NK cells (Domagala-Kulawik et al., 2014). Similarly, our data showed significant differences in the proportion of NK cells, B cells, and T cell content between different molecular subtypes, suggesting that other immune cells may impair the cytotoxic and migratory activity of NK cells with numerical and functional advantages, and thus causing NK cell depletion (Bi and Tian, 2017). But we found that activated NK cells had no differences between high- and low-group, maybe caused by insufficient samples.

Changes in NK cell counts, including peripheral blood, circulation and TME in healthy individuals, can be used as prognostic markers in patients with head and neck and lung tumors (Lin et al., 2017; Lin et al., 2020; Zhong et al., 2021). We constructed the prognosis model by NK cell-related genes (ANLN, FAM83A, RHOV, and PARP15), which is a powerful tool to assist clinical decision-making with effective prediction of patient survival and drug sensitivity. ANLN is an actin-binding protein, and previous studies have demonstrated that ANLN is associated with actin cytoskeleton dynamics (Xu et al., 2019). Xu J et al. showed that ANLN overexpression promotes distant metastasis of lung cancer cells and is associated with epithelial mesenchymal transformation (EMT) of LUAD cells transformation (EMT) in LUAD cells. Similar to previous bioinformatic analyses, our study found that upregulated FAM83A in LUAD tissues, which was relate to LUAD prognosis (Suzuki et al., 2005; Deng et al., 2021). Knockdown of FAM83A inhibited proliferation, migration and invasion of LUAD cells. In addition, the lncRNA FAM83A-AS1 regulates FAM83A expression by acting as a competing endogenous RNA for miR-495-3p (Wang et al., 2021). These results suggested that FAM83A plays an oncogenic role in LUAD and that FAM831-AS1 can regulate FAM83 expression by taking up miR-495-3p. Similar to FAM83A, invasion, migration and proliferation of LUAD cells could be stimulated by RHOV overexpression, while knockdown of RHOV inhibits the functionalistic behavior of the cells. In addition, RHOV knockdown inhibits metastasis and LUAD tumor growth of nude mice, which may be related to RHOV activation of the JNK/c-Jun signaling pathway (Zhang et al., 2021c). There are fewer basic studies on PARP15 in LUAD, and genomic data with large sample sizes suggested that RHOV is a useful marker for immunotherapy and survival in LUAD (Han et al., 2020). The above studies revealed a novel regulatory mechanism of NK cells in LUAD tumor development, which may be a new biomarker and therapeutic target for LUAD.

Docetaxel, Vinorelbine, Paclitaxel and Cisplatin are currently widely used chemotherapy drugs for lung cancer, which cause cell cycle arrest (Clegg et al., 2001; Dasari and Tchounwou, 2014). However, resistance can develop, leading to further tumor development and side effects such as myelosuppression, drug nephritis, nausea, vomiting, hearing loss and polyneuropathy, which will significantly reduce the patient's quality of life (Dasari and Tchounwou, 2014). Acquired chemotherapy resistance is a major problem faced by clinicians and a major cause of treatment failure. Regardless of the type of resistance, loss of tumor sensitivity to the drug leaves very little time for therapy to correct, with the goal of improving patient survival. Patients' clinical outcomes can be significantly improved by personalizing treatment regimens and predicting the effects of drug therapy. The results of this study showed that patients in C1 subtype and high-risk group were more sensitive to and benefited from four chemotherapy drugs. We speculated that may be the number of NK cells affects drug sensitivity.

Although this study reveals the immune signature of NK cell-related genes in LUAD and confirms the role in prognosis and immunotherapy of LUAD, the following limitations remain: (Hirsch et al., 2017): The wide variety, rapid development of bioinformatics tools can help predict potential key molecules and pathways, narrow the scope and improve the efficiency of the study, but the final findings should be validated based on real genetic data in basic and clinical settings; (Siegel et al., 2022); The database used to conduct functional and signaling pathway enrichment analysis has comprehensive and complete data, but its slow updates may have some unpredictable effects on the results; (Succony et al., 2021); The results were based on extrapolation of the raw signal algorithm and should be supported by further laboratory and clinical evidence.

5 Conclusion

Based on NK cell-related genes, we identified three stable molecular subtypes of LUAD, which differed significantly in terms of immunity, pathways, prognosis and drug sensitivity among different molecular subtypes. Based on NK cell-related genes, this study developed a prognostic model, which was highly robust and had a greater potential for application in predicting immunotherapeutic response and patient prognosis.

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 author.

Author contributions

All authors contributed to this present work: DZ, designed the study. YZ, acquired the data. YZ, drafted the manuscript. DZ, revised the manuscript. All authors read and approved the manuscript.

Funding

This study was supported by Clinical Study on Diagnosis and Treatment of Peripheral Pulmonary Nodules by Bronchoscopic Navigation and Thoracic Wall Navigation (No. S2023-YF-YBSF-0407).

Conflict of interest

Author YZ was employed by Yuce Biotechnology Co, Ltd.

The remaining author declares 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.2023.1156230/full#supplementary-material

References

Anichini, A., Perotti, V. E., Sgambelluri, F., and Mortarini, R. (2020). Immune escape mechanisms in non small cell lung cancer. *Cancers* 12 (12), 3605. doi:10.3390/cancers12123605

Arneth, B. (2019). Tumor microenvironment. *Med. Kaunas. Lith.* 56 (1), 15. doi:10.3390/medicina56010015

Azman, J., Frković, V., Bilić-Zulle, L., and Petrovecki, M. (2006). [Correlation and regression]. *Acta medica Croat. cas. Hravatske akad. Med. znan.* 60 (1), 81–91.

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462 (7269), 108–112. doi:10.1038/nature08460

Bi, J., and Tian, Z. (2017). NK cell exhaustion. *Front. Immunol.* 8, 760. doi:10.3389/fimmu.2017.00760

Chakraborty, H., and Hossain, A. (2018). R package to estimate intracluster correlation coefficient with confidence interval for binary data. *Comput. methods programs Biomed.* 155, 85–92. doi:10.1016/j.cmpb.2017.10.023

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. Clift. NJ)* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Clegg, A., Scott, D. A., Sidhu, M., Hewitson, P., and Waugh, N. (2001). A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer. *Health Technol. Assess. Winch. Engl.* 5 (32), 1–195. doi:10.3310/hta5320

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

Crinier, A., Narni-Mancinelli, E., Ugolini, S., and Vivier, E. (2020). SnapShot: Natural killer cells. *Cell.* 180 (6), 1280–1280.e1. doi:10.1016/j.cell.2020.02.029

Daëron, M., Jaeger, S., Du Pasquier, L., and Vivier, E. (2008). Immunoreceptor tyrosine-based inhibition motifs: A quest in the past and future. *Immunol. Rev.* 224, 11–43. doi:10.1111/j.1600-065X.2008.00666.x

Dasari, S., and Tchounwou, P. B. (2014). Cisplatin in cancer therapy: Molecular mechanisms of action. *Eur. J. Pharmacol.* 740, 364–378. doi:10.1016/j.ejphar.2014.07.025

Deng, F., Xu, Z., Zhou, J., Zhang, R., and Gong, X. (2021). ANLN regulated by miR-30a-5p mediates malignant progression of lung adenocarcinoma. *Comput. Math. methods Med.* 2021, 9549287. doi:10.1155/2021/9549287

Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell. death Dis.* 9 (2), 117. doi:10.1038/s41419-017-0063-y

Domagala-Kulawik, J., Osinska, I., and Hoser, G. (2014). Mechanisms of immune response regulation in lung cancer. *Transl. lung cancer Res.* 3 (1), 15–22. doi:10.3978/j.issn.2218-6751.2013.11.03

Duma, N., Santana-Davila, R., and Molina, J. R. (2019). Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment. *Mayo Clin. Proc.* 94 (8), 1623–1640. doi:10.1016/j.mayocp.2019.01.013

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

Guillerey, C. (2020). NK cells in the tumor microenvironment. *Adv. Exp. Med. Biol.* 1273, 69–90. doi:10.1007/978-3-030-49270-0_4

Han, L., Shi, H., Luo, Y., Sun, W., Li, S., Zhang, N., et al. (2020). Gene signature based on B cell predicts clinical outcome of radiotherapy and immunotherapy for patients with lung adenocarcinoma. *Cancer Med.* 9 (24), 9581–9594. doi:10.1002/cam4.3561

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 (London, Engl.* 389 (10066), 299–311. doi:10.1016/S0140-6736(16)30958-8

Hong, G., Chen, X., Sun, X., Zhou, M., Liu, B., Li, Z., et al. (2019). Effect of autologous NK cell immunotherapy on advanced lung adenocarcinoma with EGFR mutations. *Precis. Clin. Med.* 2 (4), 235–245. doi:10.1093/pcmedi/pbz023

Hoy, H., Lynch, T., and Beck, M. (2019). Surgical treatment of lung cancer. *Crit. care Nurs. Clin. N. Am.* 31 (3), 303–313. doi:10.1016/j.cnc.2019.05.002

Hsieh, C. S., Lee, H. M., and Lio, C. W. (2012). Selection of regulatory T cells in the thymus. *Nat. Rev. Immunol.* 12 (3), 157–167. doi:10.1038/nri3155

Hua, Z. D., Liu, X. B., Sheng, J. H., Li, C., Li, P., Cai, X. Q., et al. (2021). UBE2V2 positively correlates with PD-L1 expression and confers poor patient survival in lung adenocarcinoma. *Appl. Immunohistochem. Mol. Morphol. AIMM* 29 (8), 585–591. doi:10.1097/PAI.0000000000000928

Hutchinson, B. D., Shroff, G. S., Truong, M. T., and Ko, J. P. (2019). Spectrum of lung adenocarcinoma. *Seminars ultrasound, CT, MR* 40 (3), 255–264. doi:10.1053/j.sult.2018.11.009

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

Jiang, T., Fang, Z., Tang, S., Cheng, R., Li, Y., Ren, S., et al. (2021). Mutational landscape and evolutionary pattern of liver and brain metastasis in lung adenocarcinoma. *J. Thorac. Oncol. official Publ. Int. Assoc. Study Lung Cancer* 16 (2), 237–249. doi:10.1016/j.jtho.2020.10.128

Lee, N., Llano, M., Carretero, M., Ishitani, A., Navarro, F., López-Botet, M., et al. (1998). HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. *Proc. Natl. Acad. Sci. U. S. A.* 95 (9), 5199–5204. doi:10.1073/pnas.95.9.5199

Lin, M., Liang, S. Z., Shi, J., Niu, L. Z., Chen, J. B., Zhang, M. J., et al. (2017). Circulating tumor cell as a biomarker for evaluating allogenic NK cell immunotherapy on stage IV non-small cell lung cancer. *Immunol. Lett.* 191, 10–15. doi:10.1016/j.imlet.2017.09.004

Lin, M., Luo, H., Liang, S., Chen, J., Liu, A., Niu, L., et al. (2020). Pembrolizumab plus allogeneic NK cells in advanced non-small cell lung cancer patients. *J. Clin. investigation* 130 (5), 2560–2569. doi:10.1172/JCI132712

Nayyar, G., Chu, Y., and Cairo, M. S. (2019). Overcoming resistance to natural killer cell based immunotherapies for solid tumors. *Front. Oncol.* 9, 51. doi:10.3389/fonc.2019.00051

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

Nie, J., Shan, D., Li, S., Zhang, S., Zi, X., Xing, F., et al. (2021). A novel ferroptosis related gene signature for prognosis prediction in patients with colon cancer. *Front. Oncol.* 11, 654076. doi:10.3389/fonc.2021.654076

Paulsen, E. E., Kilvaer, T. K., Rakaee, M., Richardsen, E., Hald, S. M., Andersen, S., et al. (2017). CTLA-4 expression in the non-small cell lung cancer patient tumor microenvironment: Diverging prognostic impact in primary tumors and lymph node metastases. *CII* 66 (11), 1449–1461. doi:10.1007/s00262-017-2039-2

Russell, É., Conroy, M. J., and Barr, M. P. (2022). Harnessing natural killer cells in non-small cell lung cancer. *Cells* 11 (4), 605. doi:10.3390/cells11040605

Saab, S., Zalzale, H., Rahal, Z., Khalifeh, Y., Sinjab, A., and Kadara, H. (2020). Insights into lung cancer immune-based biology, prevention, and treatment. *Front. Immunol.* 11, 159. doi:10.3389/fimmu.2020.00159

Sadeghzadeh, M., Bornehdeli, S., Mohahammadrezakhani, H., Abolghasemi, M., Poursaei, E., Asadi, M., et al. (2020). Dendritic cell therapy in cancer treatment; the state-of-the-art. *Life Sci.* 254, 117580. doi:10.1016/j.lfs.2020.117580

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

Spella, M., and Stathopoulos, G. T. (2021). Immune resistance in lung adenocarcinoma. *Cancers* 13 (3), 384. doi:10.3390/cancers13030384

Succony, L., Rassl, D. M., Barker, A. P., McCaughan, F. M., and Rintoul, R. C. (2021). Adenocarcinoma spectrum lesions of the lung: Detection, pathology and treatment strategies. *Cancer Treat. Rev.* 99, 102237. doi:10.1016/j.ctrv.2021.102237

Sun, J., Yue, W., You, J., Wei, X., Huang, Y., Ling, Z., et al. (2021). Identification of a novel ferroptosis-related gene prognostic signature in bladder cancer. *Front. Oncol.* 11, 730716. doi:10.3389/fonc.2021.730716

Suster, D. I., and Mino-Kenudson, M. (2020). Molecular pathology of primary non-small cell lung cancer. *Archives Med. Res.* 51 (8), 784–798. doi:10.1016/j.arcmed.2020.08.004

Suzuki, C., Daigo, Y., Ishikawa, N., Kato, T., Hayama, S., Ito, T., et al. (2005). ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. *Cancer Res.* 65 (24), 11314–11325. doi:10.1158/0008-5472.CAN-05-1507

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48 (4), 812–830.e14. doi:10.1016/j.immuni.2018.03.023

Toro-Domínguez, D., Martorell-Marugán, J., López-Domínguez, R., García-Moreno, A., González-Rumayor, V., Alarcón-Riquelme, M. E., et al. (2019). ImaGEO: Integrative gene expression meta-analysis from GEO database. *Bioinforma. Oxf. Engl.* 35 (5), 880–882. doi:10.1093/bioinformatics/bty721

Valipour, B., Velaei, K., Abedelahi, A., Karimipour, M., Darabi, M., and Charoudeh, H. N. (2019). NK cells: An attractive candidate for cancer therapy. *J. Cell. physiology* 234 (11), 19352–19365. doi:10.1002/jcp.28657

Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M. J., and Steyerberg, E. W. (2016). A calibration hierarchy for risk models was defined: From utopia to empirical data. *J. Clin. Epidemiol.* 74, 167–176. doi:10.1016/j.jclinepi.2015.12.005

Van Calster, B., Wynants, L., Verbeek, J. F. M., Verbakel, J. Y., Christodoulou, E., Vickers, A. J., et al. (2018). Reporting and interpreting decision curve analysis: A guide for investigators. *Eur. Urol.* 74 (6), 796–804. doi:10.1016/j.eururo.2018.08.038

Vitale, I., Manic, G., Coussens, L. M., Kroemer, G., and Galluzzi, L. (2019). Macrophages and metabolism in the tumor microenvironment. *Cell. metab.* 30 (1), 36–50. doi:10.1016/j.cmet.2019.06.001

Wang, G., Li, X., Yao, Y., Jiang, Z., Zhou, H., Xie, K., et al. (2021). FAM83A and FAM83A-AS1 both play oncogenic roles in lung adenocarcinoma. *Oncol. Lett.* 21 (4), 297. doi:10.3892/ol.2021.12558

Xu, J., Zheng, H., Yuan, S., Zhou, B., Zhao, W., Pan, Y., et al. (2019). Overexpression of ANLN in lung adenocarcinoma is associated with metastasis. *Thorac. cancer* 10 (8), 1702–1709. doi:10.1111/1759-7714.13135

Yotsukura, M., Asamura, H., Motoi, N., Kashima, J., Yoshida, Y., Nakagawa, K., et al. (2021). Long-term prognosis of patients with resected adenocarcinoma *in situ* and minimally invasive adenocarcinoma of the lung. *J. Thorac. Oncol. official Publ. Int. Assoc. Study Lung Cancer* 16 (8), 1312–1320. doi:10.1016/j.jtho.2021.04.007

Yu, Y., Wang, Z., Zheng, Q., and Li, J. (2021). GREB1L overexpression correlates with prognosis and immune cell infiltration in lung adenocarcinoma. *Sci. Rep.* 11 (1), 13281. doi:10.1038/s41598-021-92695-x

Zhang, C., Zhang, G., Sun, N., Zhang, Z., Zhang, Z., Luo, Y., et al. (2020). Comprehensive molecular analyses of a TNF family-based signature with regard to prognosis, immune features, and biomarkers for immunotherapy in lung adenocarcinoma. *EBioMedicine* 59, 102959. doi:10.1016/j.ebiom.2020.102959

Zhang, H., Luo, Y. B., Wu, W., Zhang, L., Wang, Z., Dai, Z., et al. (2021). The molecular feature of macrophages in tumor immune microenvironment of glioma patients. *Comput. Struct. Biotechnol. J.* 19, 4603–4618. doi:10.1016/j.csbj.2021.08.019

Zhang, Q., Tang, L., Zhou, Y., He, W., and Li, W. (2021). Immune checkpoint inhibitor-associated pneumonitis in non-small cell lung cancer: Current understanding in characteristics, diagnosis, and management. *Front. Immunol.* 12, 663986. doi:10.3389/fimmu.2021.663986

Zhang, D., Jiang, Q., Ge, X., Shi, Y., Ye, T., Mi, Y., et al. (2021). RHOV promotes lung adenocarcinoma cell growth and metastasis through JNK/c-Jun pathway. *Int. J. Biol. Sci.* 17 (10), 2622–2632. doi:10.7150/ijbs.59939

Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Ann. Transl. Med.* 4 (7), 136. doi:10.21037/atm.2016.03.35

Zhong, R., Chen, D., Cao, S., Li, J., Han, B., and Zhong, H. (2021). Immune cell infiltration features and related marker genes in lung cancer based on single-cell RNA-seq. *Clin. Transl. Oncol. official Publ. Fed. Span. Oncol. Soc. Natl. Cancer Inst. Mexico* 23 (2), 405–417. doi:10.1007/s12094-020-02435-2

Zou, Y., Xie, J., Zheng, S., Liu, W., Tang, Y., Tian, W., et al. (2022). Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery. *Int. J. Surg. Lond. Engl.* 107, 106936. doi:10.1016/j.ijsu.2022.106936

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 20 June 2023
DOI 10.3389/fgene.2023.1112251

OPEN ACCESS

EDITED BY
Min Sun,
Hubei University of Medicine, China

REVIEWED BY
Pranabananda Dutta,
Charles R. Drew University of Medicine and Science, United States
Francesk Mulita,
General University Hospital of Patras, Greece

*CORRESPONDENCE
Xueqiong Han,
✉ hanxueqiong@stu.gxmu.edu.cn

RECEIVED 30 November 2022
ACCEPTED 14 April 2023
PUBLISHED 20 June 2023

CITATION
Deng Y, Li Z, Pan M, Wu H, Ni B and Han X (2023), Implications of inflammatory cell death-related IFNG and co-expressed RNAs (AC006369.1 and CCR7) in breast carcinoma prognosis, and anti-tumor immunity.
Front. Genet. 14:1112251.
doi: 10.3389/fgene.2023.1112251

Implications of inflammatory cell death-related IFNG and co-expressed RNAs (AC006369.1 and CCR7) in breast carcinoma prognosis, and anti-tumor immunity

Yongran Deng, Zhenlong Li, Mingmei Pan, Huayun Wu, Bingqiang Ni and Xueqiong Han*

Department of Oncology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China

Objective: Interferon-γ (IFN-γ) encoded by IFNG gene is a pleiotropic molecule linked with inflammatory cell death mechanisms. This work aimed to determine and characterize IFNG and co-expressed genes, and to define their implications in breast carcinoma (BRCA).

Methods: Transcriptome profiles of BRCA were retrospectively acquired from public datasets. Combination of differential expression analysis with WGCNA was conducted for selecting IFNG-co-expressed genes. A prognostic signature was generated through Cox regression approaches. The tumor microenvironment populations were inferred utilizing CIBERSORT. Epigenetic and epitranscriptomic mechanisms were also probed.

Results: IFNG was overexpressed in BRCA, and connected with prolonged overall survival and recurrence-free survival. Two IFNG-co-expressed RNAs (AC006369.1, and CCR7) constituted a prognostic model that acted as an independent risk factor. The nomogram composed of the model, TNM, stage, and new event owned the satisfying efficacy in BRCA prognostication. IFNG, AC006369.1, and CCR7 were closely linked with the tumor microenvironment components (e.g., macrophages, CD4/CD8 T cells, NK cells), and immune checkpoints (notably PD1/PD-L1). Somatic mutation frequencies were 6%, and 3% for CCR7, and IFNG, and high amplification potentially resulted in their overexpression in BRCA. Hypomethylated cg05224770 and cg07388018 were connected with IFNG and CCR7 upregulation, respectively. Additionally, transcription factors, RNA-binding proteins, and non-coding RNAs possibly regulated IFNG and co-expressed genes at the transcriptional and post-transcriptional levels.

Abbreviations: BRCA, breast carcinoma; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IFN-γ, interferon-γ; NK, natural killer; TME, tumor microenvironment; TCGA, The Cancer Genome Atlas; miRNA, microRNA; WGCNA, weighted correlation network analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; lncRNAs, long non-coding RNAs; OS, overall survival; RFS, recurrence-free survival; Tregs, T cells regulatory.

Conclusion: Collectively, our work identifies IFNG and co-expressed genes as prognostic markers for BRCA, and as possible therapeutic targets for improving the efficacy of immunotherapy.

KEYWORDS

breast carcinoma, IFNg, inflammatory cell death, AC006369.1, CCR7, prognosis, tumor microenvironment, immune checkpoint

Introduction

Breast carcinoma (BRCA) has a high incidence globally, with over two million cases per year (Sung et al., 2021). This malignancy represents a remarkable threat to female health and affects one in seven women over the course of a lifetime (Corti et al., 2022). Based upon the expression status of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), four molecular subtypes have been widely accepted: luminal A, luminal B, HER2-enriched and basal-like tumors (Bidard et al., 2022; Curigliano et al., 2022; Shepherd et al., 2022). Despite the progress in early diagnosis and treatment, most patients still succumb to various complex malignant phenotypes (Martin et al., 2022; Mayer et al., 2022; Shepherd et al., 2022). Within 10 years following breast conservation surgical resection with post-operative radiotherapy, the recurrence rate is still as high as 3%–15% (Gadaleta et al., 2022). Emerging immunotherapy has exhibited promising results in BRCA, but with low response rates (Loibl et al., 2019; Schmid et al., 2020; Huober et al., 2022). Such alarming situation has prompted to determine innovative and effective therapeutic targets for BRCA.

Interferon-γ (IFN-γ) encoded by IFNG gene is the only member of the type II interferon family, which is an essential cytokine generated from activated T cells, natural killer (NK), and NK T cells in the tumor microenvironment (TME) (Dörrie et al., 1999; Wu H. et al., 2022; Wei et al., 2022). Cell death can provide host defense and maintain homeostasis (Niu et al., 2021; Wang Z. et al., 2022). IFN-γ can prime diverse inflammatory cell death mechanisms. For instance, IFN-γ secreted from $CD8^+$ T cells rewires lipid metabolism of malignant cells through ACSL4, thus activating polyunsaturated fatty acids and sensitizing malignant cells to ferroptotic cell death (Liao et al., 2022). IFN-γ can also initiate macrophages for pathogen ligand-induced killing through caspase-8 and mitochondrial cell death signaling (Simpson et al., 2022). Moreover, the diverse implications of IFN-γ in BRCA (e.g., prognostication, therapeutic efficacy) have been demonstrated in prior studies (Witek Janusek et al., 2019). Non-etheless, IFN-γ-co-expressed genes and underlying molecular mechanisms remain indistinct in BRCA. For solving these problems, this work was implemented for determining and characterizing IFNG and co-expressed genes, and clarifying their implications in BRCA and probing possible epigenetic and epitranscriptomic mechanisms.

Materials and methods

Collection of BRCA datasets

BRCA transcriptome RNA-sequencing data (Htseq-FPKM) and matched clinical parameters were gathered from The Cancer Genome Atlas (TCGA) database. Somatic mutation, copy-number alteration, DNA methylation, and microRNA (miRNA) data were also extracted. External microarray datasets from the Gene Expression Omnibus database were online analyzed on the Kaplan-Meier Plotter platform.

Selection of IFNG-co-expressed genes

Utilizing limma method (Ritchie et al., 2015), aberrant expressed genes in BRCA *versus* control specimens were selected with adjusted $p < 0.05$. Based upon the same threshold, genes with different expression between lowly and highly expressed IFNG BRCA were acquired. Above genes were intersected and named as BRCA- and IFNG-relevant genes. Next, weighted correlation network analysis (WGCNA) was implemented through WGCNA package (Langfelder and Horvath, 2008). Firstly, a clustering dendrogram was plotted, with removal of outliers via hierarchical clustering analysis. By Pearson's test, interactions between genes were analyzed, and interaction pairs with $p < 0.05$ were used for constructing a similarity matrix. Afterwards, soft thresholding value was adopted for transforming the similarity matrix to the adjacency matrix. A scale-free network and topological overlap matrix were built, respectively. Next, a hierarchical clustering dendrogram was produced for detecting modules. At last, modules were merged with dynamic tree cutting approach. The module with the strongest connection to IFNG was chosen or subsequent analysis.

Functional enrichment analysis

Enrichment on Gene Ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was analyzed based upon module genes by use of clusterProfiler approach (Yu et al., 2012).

Cox regression analysis and nomogram establishment

Univariate-cox regression analysis on genes in the black module with prognosis was conducted. Genes with $p < 0.05$ were selected for the construction of a multivariate cox regression model. Based upon 1:1, TCGA-BRCA cases were randomized into the discovery and verification sets. Survival difference was then estimated. The predictive independency was analyzed utilizing cox regression analysis. A nomogram was defined with rms package, and predictive efficiency was demonstrated by calibration curves.

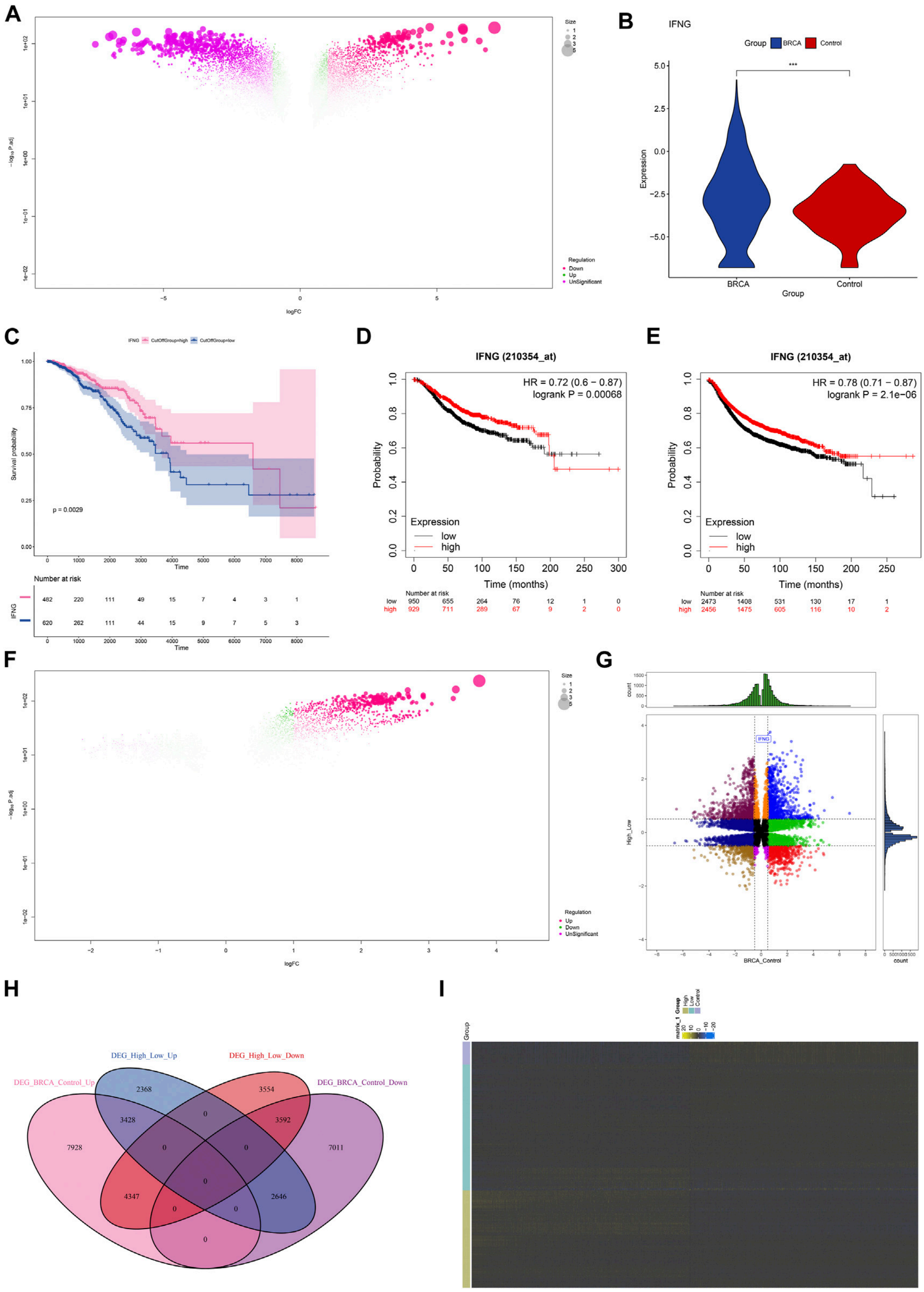


FIGURE 1

Expression and prognostic implication of IFNG and selection of IFNG-relevant genes in BRCA. **(A)** Genes with aberrant expression in BRCA *versus* control specimens. **(B)** IFNG transcript level in BRCA and controls. **(C)** OS probability of groups with low or high IFNG expression. **(D, E)** Verification of OS and RFS of two groups in multiple microarray datasets. **(F)** Genes with different expression in BRCA samples with low *versus* high IFNG expression. **(G, H)** The shared expression patterns in BRCA *versus* controls and high *versus* low IFNG expression. **(I)** The transcript level of IFNG-relevant genes in controls, BRCA with low or high IFNG expression.

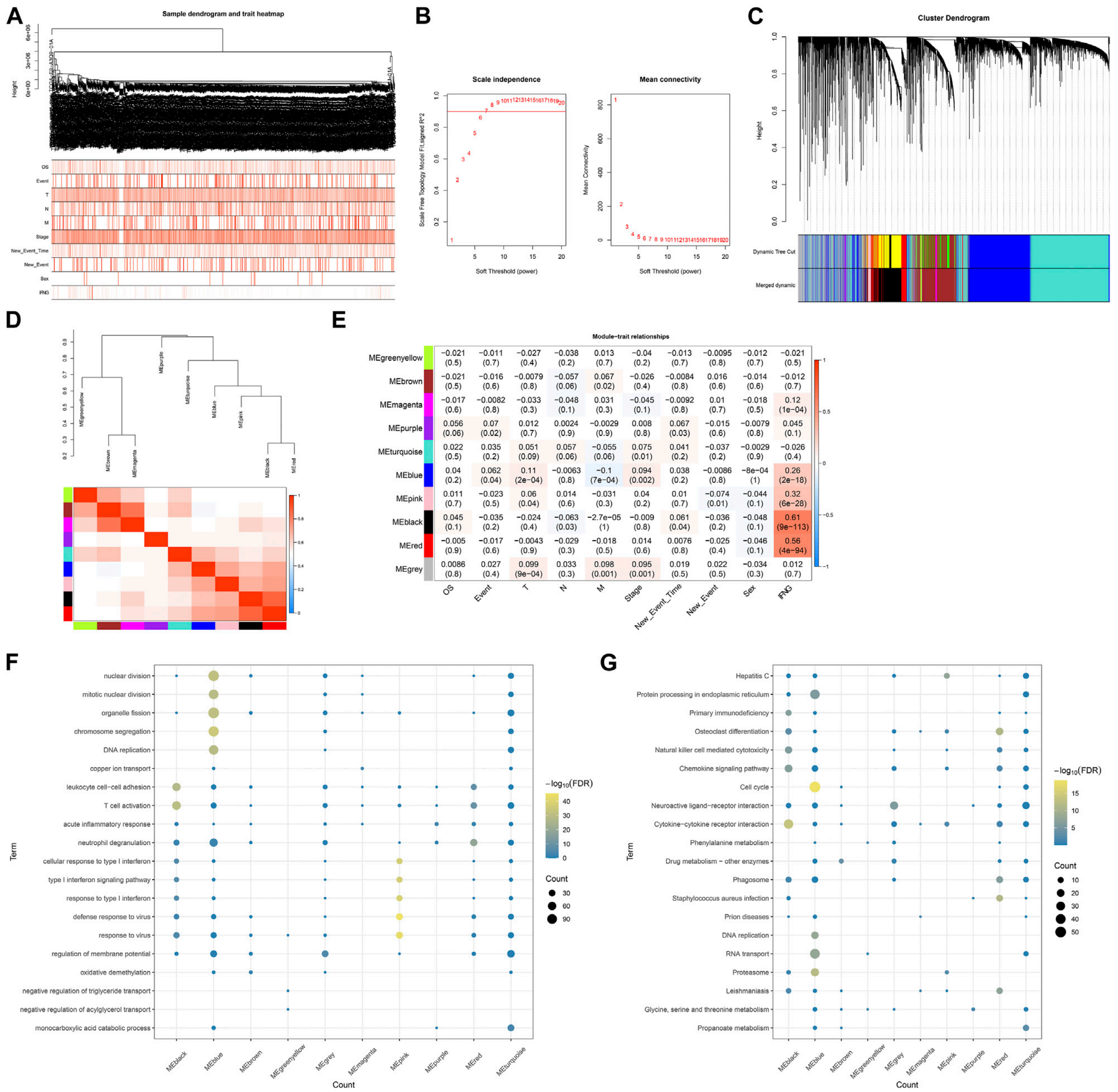


FIGURE 2
Establishment of IFNG-based co-expression modules. **(A)** Sample dendrogram and heatmap of clinical characteristics. **(B)** Scale independence along with mean connectivity under diverse soft-thresholding values. **(C)** Clustering dendrogram and merged modules. **(D)** Eigengene dendrogram and heatmap of eigengene adjacency. **(E)** Relationships of co-expression modules with clinical characteristics. **(F, G)** GO and KEGG pathways of genes in each module.

Quantification of the TME components

CIBERSORT is an algorithm for characterization of the cellular compositions within bulk tissues based upon transcriptome profiling (Newman et al., 2015). The components within the TME were quantified by use of this algorithm.

Genetic alteration assessment

Somatic variants were estimated by use of maftools package (Mayakonda et al., 2018). The mutated frequency of IFNG and co-expressed genes was extracted. GISTIC2.0 was adopted for copy-number alterations of above genes (Mermel et al., 2011).

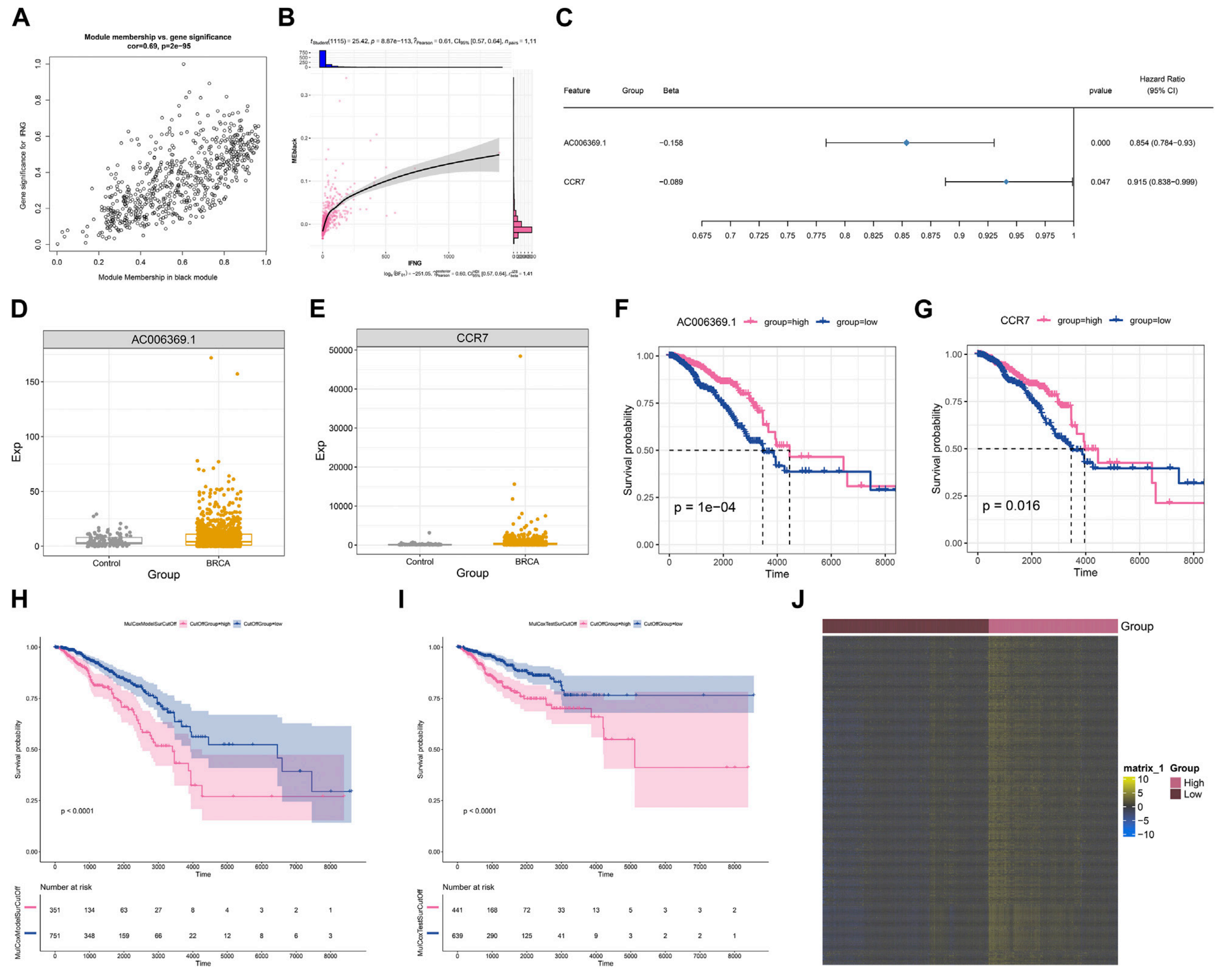


FIGURE 3
Generation of an IFNG-co-expressed prognostic signature. **(A)** Relationships of module membership in black module with gene significance for IFNG. **(B)** Correlation analysis on IFNG expression with black module. **(C)** Univariate-cox regression results of AC006369.1, and CCR7 with BRCA survival. **(D, E)** The transcript level of AC006369.1, and CCR7 in BRCA and controls. **(F, G)** OS analysis of AC006369.1, and CCR7 across BRCA patients. **(H, I)** OS difference between low- and high-score groups in the discovery and verification sets. **(J)** Heatmap of the expression patterns of IFNG-co-expressed genes in low- and high-score groups.

DNA methylation analysis

DNA methylation levels (beta-values) were normalized by use of preprocessCore package. Interactions of IFNG and co-expressed genes with methylation sites were then assessed.

Non-coding RNA analysis

MiRNAs with different expression were screened between BRCA *versus* controls and lowly *versus* highly expressed IFNG BRCA following adjusted $p < 0.05$. Above miRNAs were intersected, and determined as BRCA- and IFNG-relevant miRNAs. Correlation analysis on long non-coding RNAs (lncRNAs) with IFNG and co-expressed genes was then carried out.

Statistical analysis

For continuous variables, Student's t-test, or one-way ANOVA test was utilized for comprising between groups. Chi-square or Fisher's exact test was employed for analysis of categorical data. Kaplan-Meier curves of overall survival (OS) and recurrence-free survival (RFS) were plotted, with log-rank test for estimating survival difference. Correlation analysis was conducted with Pearson's test. All analyses were achieved based upon the R platform (version 4.0.3). $p < 0.05$ indicated statistically significant.

TABLE 1 Univariate-cox regression results of IFNG-relevant genes with BRCA prognosis.

IFNG-relevant genes	Beta	z	p	Hazard ratio	Lower	Upper
AC006369.1	−0.027	−2.67695	0.00743	0.973336	0.954265	0.992788
CCR7	0	2.671615	0.007549	1.000088	1.000023	1.000153
RPL4P1	−0.03	−2.44974	0.014296	0.970893	0.948216	0.994112
TRBV5.5	−0.045	−2.43696	0.014811	0.955617	0.921356	0.991154
TRDV1	−0.014	−2.42653	0.015244	0.985938	0.974724	0.997281
PSMB8	0	−2.36695	0.017935	0.999942	0.999894	0.99999
DEF6	0	−2.35957	0.018296	0.999729	0.999504	0.999954
SHISAL2A	−0.005	−2.23588	0.025359	0.9949	0.99045	0.999369
TRBC2	0	−2.19914	0.027868	0.999841	0.999699	0.999983
HCST	−0.001	−2.19251	0.028343	0.998989	0.998085	0.999893
GZMM	−0.002	−2.16979	0.030023	0.998335	0.996834	0.999839
RAC2	0	−2.08823	0.036778	0.999904	0.999814	0.999994
ARMH1	−0.002	−2.07626	0.037869	0.997758	0.995646	0.999874
IL12B	−0.006	−1.96702	0.049181	0.993893	0.987845	0.999978
TRBV4.2	−0.006	−1.92912	0.053717	0.993905	0.987751	1.000098
SPIB	0	−1.92572	0.054139	0.999817	0.99963	1.000003
RELB	0	−1.82117	0.068581	0.999859	0.999708	1.000011
CD2	0	−1.79848	0.072101	0.999872	0.999732	1.000012
TRBV18	−0.007	−1.79285	0.072996	0.992864	0.985121	1.000668
PIM2	0	−1.79045	0.073381	0.999913	0.999819	1.000008
KLHDC7B	0	−1.75477	0.079299	0.999942	0.999877	1.000007
AL606834.2	−0.007	−1.69148	0.090745	0.993121	0.985209	1.001096
TESPA1	−0.001	−1.68354	0.092271	0.999415	0.998734	1.000096
IFNG	−0.003	−1.65036	0.09887	0.996621	0.992622	1.000635
HLA.DQB2	0	−1.6428	0.100424	0.999883	0.999743	1.000023
CCL22	0	−1.62196	0.104813	0.999762	0.999475	1.00005
CD37	0	−1.598	0.110044	0.999891	0.999757	1.000025
IGHG4	0	−1.58194	0.113664	0.999988	0.999974	1.000003
TRBV12.4	−0.007	−1.51643	0.12941	0.993283	0.984668	1.001973
FASLG	−0.001	−1.50584	0.132109	0.998747	0.997118	1.000378
GZMA	0	−1.47864	0.139236	0.999745	0.999406	1.000083
AC004585.1	−0.002	−1.47704	0.139664	0.998327	0.99611	1.000548
TRAV8.4	−0.006	−1.47261	0.140856	0.993756	0.985506	1.002075
ITGAL	0	−1.40573	0.159805	0.999934	0.999842	1.000026
TRAV12.2	−0.004	−1.37904	0.167883	0.99609	0.990559	1.001652
CAMK4	0	−1.32771	0.184273	0.999609	0.999032	1.000186
NAPSB	0	−1.31747	0.187682	0.999835	0.999589	1.000081
CD1A	0	−1.29969	0.193708	0.999669	0.999169	1.000168

(Continued on following page)

TABLE 1 (*Continued*) **Univariate-cox regression results of IFNG-relevant genes with BRCA prognosis.**

IFNG-relevant genes	Beta	z	p	Hazard ratio	Lower	Upper
AC007569.1	−0.012	−1.29509	0.195289	0.987607	0.969143	1.006423
IGHGP	0	−1.1592	0.246375	0.99997	0.99992	1.000021
RGS1	0	−1.14014	0.254229	0.999972	0.999925	1.00002
SELL	0	−1.08242	0.279068	0.999948	0.999855	1.000042
LINC00494	0.001	1.061845	0.288306	1.000693	0.999414	1.001974
FCRLA	0	−1.01453	0.310331	0.999625	0.9989	1.00035
TRAV1.2	−0.003	−0.66639	0.50516	0.99687	0.98772	1.006104
GPR18	0	−0.57179	0.567466	0.999553	0.99802	1.001087
NME8	0.001	0.569963	0.568703	1.000928	0.99774	1.004127
TRBV13	−0.002	−0.47433	0.635265	0.997881	0.989171	1.006667
IGHG2	0	−0.39298	0.694333	0.999999	0.999994	1.000004
RAB37	0	−0.23861	0.811409	0.999931	0.999364	1.000498

Results

Expression and prognostic implication of IFNG and selection of IFNG-relevant genes in BRCA

The investigation on the transcriptional alterations in BRCA was conducted. With adjusted $p < 0.05$, 28,953 genes presented the differential expression in BRCA relative to controls (Figure 1A; Supplementary Table S1). Among them, we focused on IFNG that was prominently upregulated in BRCA (Figure 1B). Its prognostic significance was then evaluated. With the cutoff value, the classification of BRCA patients as low or high IFNG expression group was performed. As illustrated in Figure 1C, patients with high IFNG expression owned the notable survival superiority. The prognostic significance was further verified in multiple microarray datasets via the Kaplan-Meier Plotter. Consistently, IFNG upregulation was connected with better OS and RFS (Figures 1D, E). Above data unveiled the involvement of IFNG in BRCA pathogenesis. Afterwards, the relevant molecules of IFNG were probed. Consequently, 19,935 genes presenting different expression between low and high IFNG expression groups were selected (Figure 1F; Supplementary Table S2). After intersecting, 7020 IFNG-relevant genes were obtained (Figures 1G–I).

Establishment of IFNG-based co-expression modules

BRCA specimens with matched clinical and IFNG characteristics were included for WGCNA (Figure 2A). The appropriate soft-thresholding value was set as 6 through considering scale independence and mean connectivity (Figure 2B). Utilizing dynamic tree cutting method, highly connected genes were merged into ten modules (Figures 2C, D). Black module exhibited the strongest connection with IFNG (Figure 2E), which was regarded as IFNG-relevant module. It was noted that genes in the black module were prominently linked with immunity (e.g., T cell activation, leukocyte cell-cell adhesion, and cytokine-cytokine receptor interaction) (Figures 2F, G).

Generation of an IFNG-co-expressed prognostic signature

Module membership in black module exhibited a notably positive connection with gene significance for IFNG (Figure 3A). It was also demonstrated that black module was positively linked with IFNG (Figure 3B). Such evidence proved that genes in black module were IFNG-co-expressed genes. Most of them owned the significant survival significance of BRCA (Table 1). Notably, AC006369.1, and CCR7 presented the aberrant expression in BRCA *versus* controls, and their upregulation was in relation to OS outcomes (Figures 3C–G). They were incorporated into the multivariate-cox regression model, and worse OS was investigated in high-score patients both in the discovery and verification sets (Figures 3H, I). Most IFNG-co-expressed genes had the higher expression in high-than low-score groups (Figure 3J), indicating their subtype specific expression.

The IFNG-co-expressed prognostic signature as an independent risk factor of BRCA and definition of a nomogram

Next, it was observed that there was a positive connection of the IFNG-co-expressed prognostic signature with event (Figures 4A, B). In addition, IFNG was negatively linked with N stage (Figure 4C). Through considering uni- and multivariate-cox regression results, the prognostic model acted as an independent risk factor of BRCA (Figures 4D, E). The nomogram composed of the prognostic model and clinical traits was defined, and the excellent predictive efficacy was proven by calibration curves (Figures 4F, G).

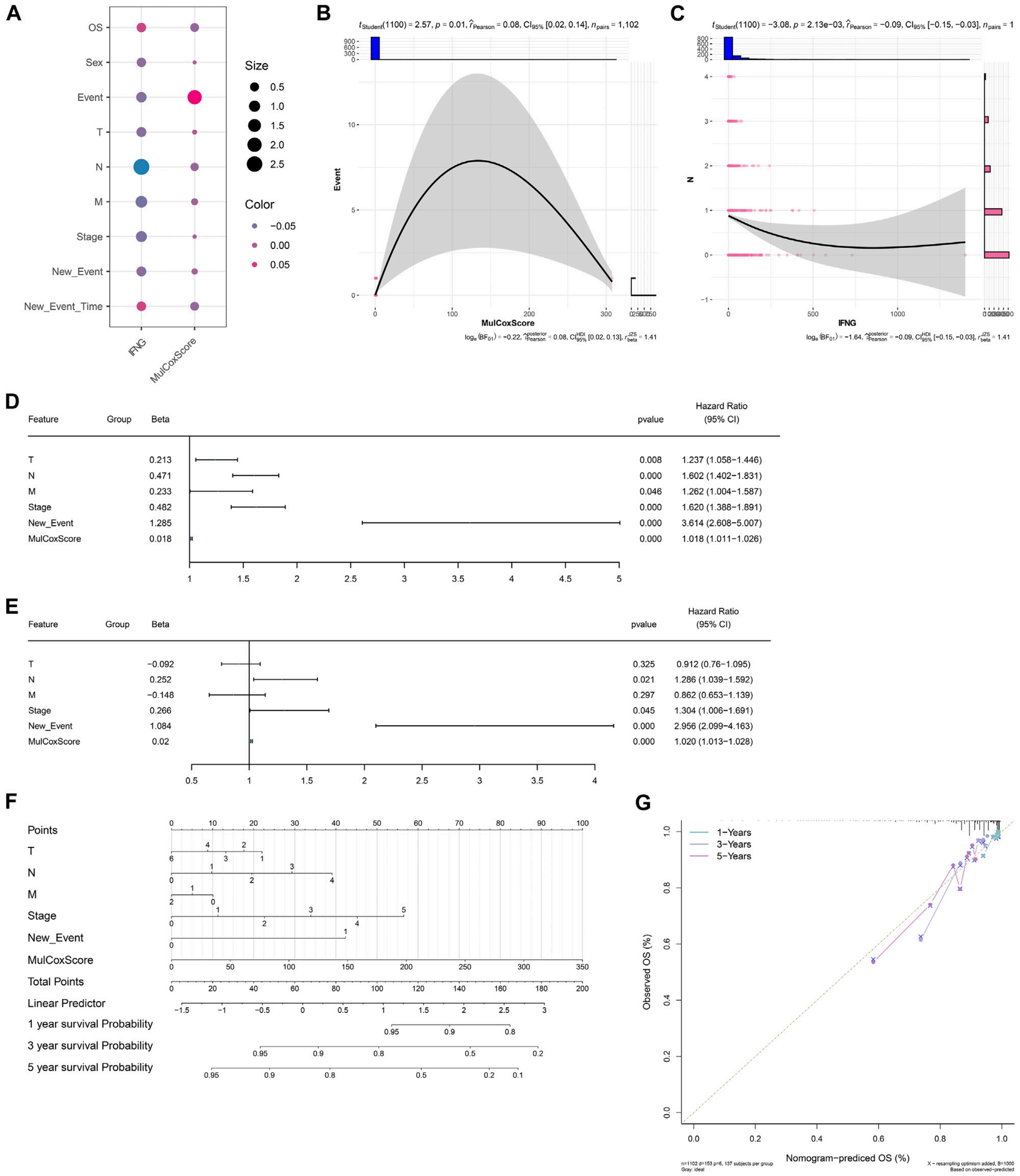


FIGURE 4
Associations of clinical traits with the IFNG-co-expressed prognostic signature and construction of a nomogram. **(A)** Correlation analysis on IFNG and the IFNG-co-expressed prognostic model with clinical parameters. **(B)** Relationship of the prognostic model *versus* event. **(C)** Relationship of IFNG *versus* N stage. **(D, E)** Uni- or multivariate-cox regression results on the prognostic signature and clinical variables with BRCA survival. **(F)** The nomogram for survival prediction. **(G)** Calibration curves depicting the model-predictive and observed OS.

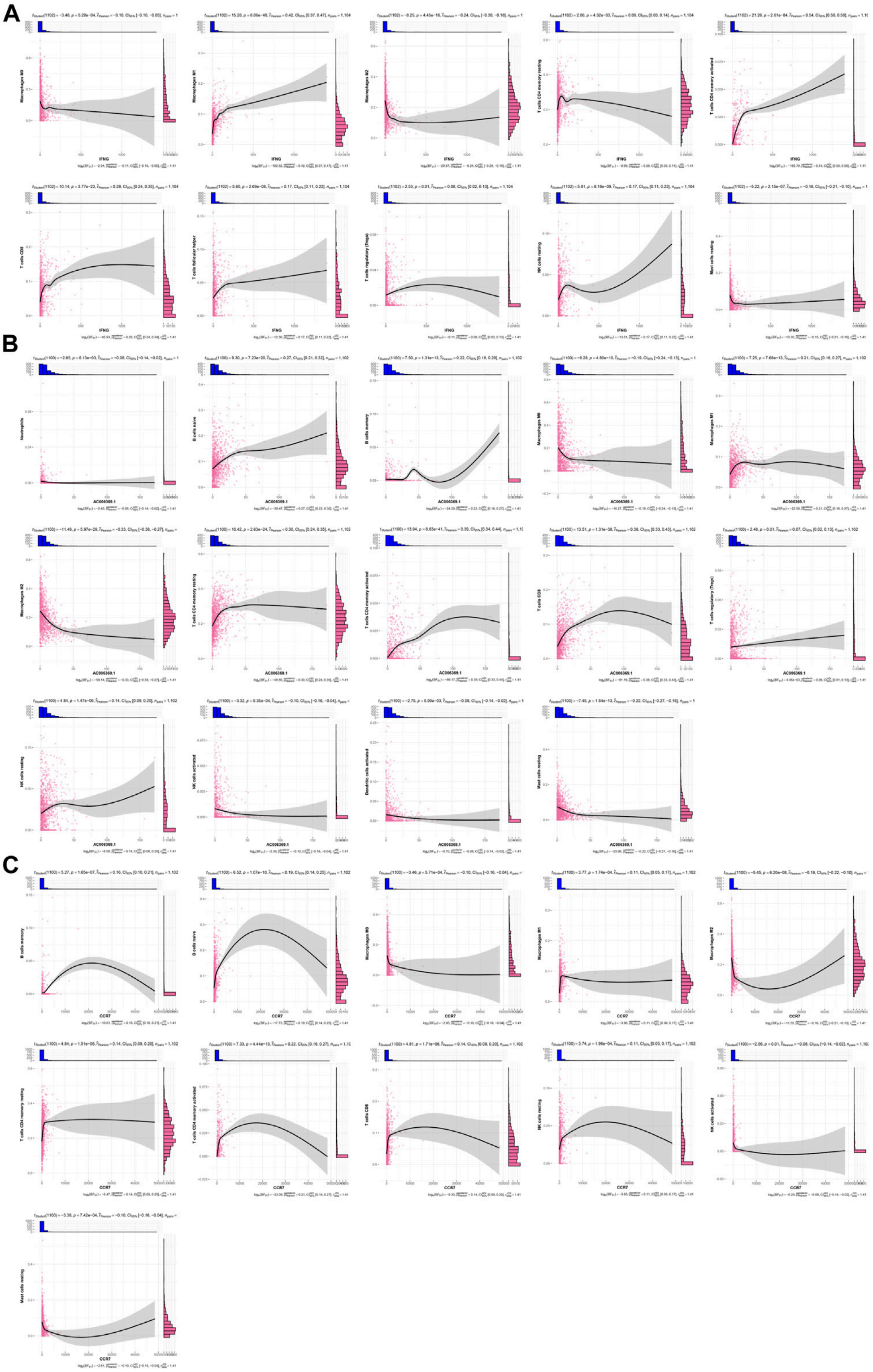


FIGURE 5
Associations of IFNG, and co-expressed AC006369.1, and CCR7 with the TME components. **(A–C)** Correlation analysis on **(A)** IFNG, **(B)** AC006369.1, and **(C)** CCR7 with the abundance of immune cells within the TME.

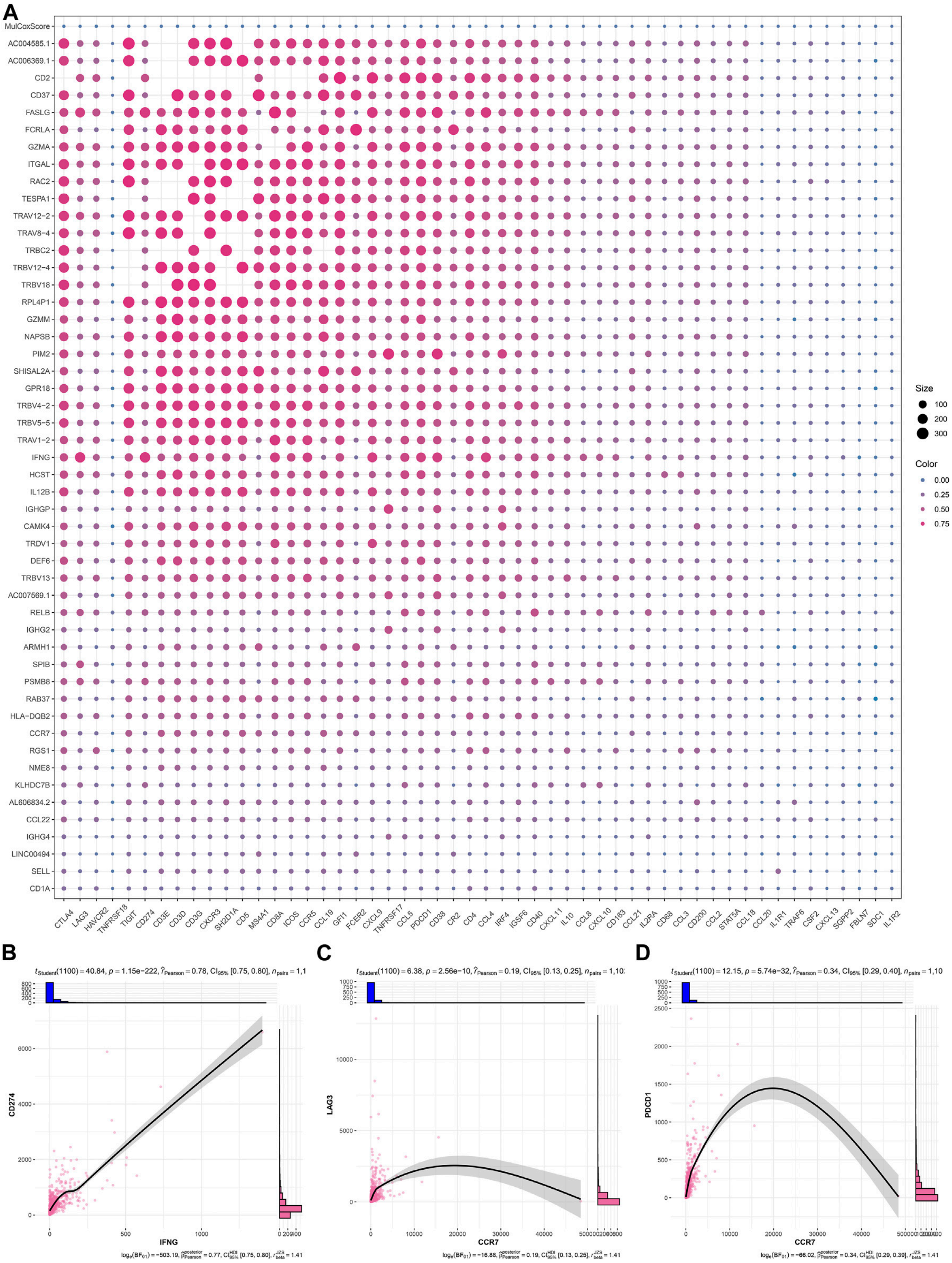


FIGURE 6
Interactions of IFNG, and co-expressed genes with immune checkpoints. **(A)** Correlation analysis on IFNG, co-expressed genes, and the prognostic signature with immune checkpoints. **(B–D)** Visualization of the relationships of IFNG with CD274, LAG3, and PDCD1.

TABLE 2 Correlation analyses of IFNG, AC006369.1, and CCR7 with immune checkpoints in BRCA.

Immune checkpoint	IFNG		AC006369.1		CCR7	
	r	p	r	p	r	p
CCL18	0.325517	1.29E-28	0.251434	2.38E-17	0.142041	2.20E-06
CCL19	0.248134	6.33E-17	0.776397	8.92E-223	0.472312	2.59E-62
CCL2	0.286609	2.79E-22	0.234465	3.15E-15	0.132134	1.08E-05
CCL20	0.095234	0.001551	0.004758	0.874644	0.00426	0.887675
CCL21	0.051963	0.084676	0.341864	1.44E-31	0.277394	6.43E-21
CCL3	0.362793	1.29E-35	0.268843	1.06E-19	0.112889	0.000173
CCL4	0.734428	2.24E-187	0.50303	9.82E-72	0.216751	3.49E-13
CCL5	0.678464	1.69E-149	0.652101	1.93E-134	0.316081	5.44E-27
CCL8	0.505601	1.44E-72	0.148589	7.25E-07	0.07855	0.00909
CCR5	0.721719	5.85E-178	0.623922	6.29E-120	0.311697	2.96E-26
CD163	0.450361	3.81E-56	0.227338	2.20E-14	0.102062	0.000691
CD200	0.222103	8.79E-14	0.35874	8.27E-35	0.189568	2.25E-10
CD274	0.776278	1.15E-222	0.345991	2.42E-32	0.168349	1.88E-08
CD38	0.732854	3.52E-186	0.5259	2.12E-79	0.288228	1.59E-22
CD3D	0.577263	6.38E-99	0.873946	0	0.468492	3.30E-61
CD3E	0.543867	7.60E-86	0.887998	0	0.493715	8.93E-69
CD3G	0.67702	1.24E-148	0.773085	1.04E-219	0.42113	1.33E-48
CD4	0.53137	2.54E-81	0.591527	5.57E-105	0.299854	2.49E-24
CD40	0.573782	1.73E-97	0.509643	6.85E-74	0.254698	8.89E-18
CD5	0.507463	3.56E-73	0.843876	1.03E-299	0.480718	8.50E-65
CD68	0.192994	1.05E-10	0.194672	7.15E-11	0.060922	0.043178
CD8A	0.732739	4.30E-186	0.735715	2.33E-188	0.354788	4.94E-34
CR2	0.150807	4.92E-07	0.482392	2.67E-65	0.294252	1.89E-23
CSF2	0.208605	2.67E-12	0.121952	4.93E-05	0.070092	0.019964
CTLA4	0.608277	1.71E-112	0.730355	2.68E-184	0.408044	1.85E-45
CXCL10	0.487788	6.12E-67	0.196638	4.56E-11	0.09579	0.001455
CXCL11	0.527657	5.17E-80	0.295427	1.24E-23	0.147004	9.53E-07
CXCL13	0.072472	0.016117	0.083451	0.005572	0.039526	0.189804
CXCL9	0.739629	2.18E-191	0.62981	7.77E-123	0.327891	4.94E-29
CXCR3	0.614729	1.65E-115	0.808784	7.48E-256	0.424934	1.52E-49
FBLN7	−0.07183	0.017081	0.031189	0.300927	0.018202	0.546117
FCER2	0.114882	0.000132	0.562931	3.97E-93	0.37077	3.06E-37
GFI1	0.627363	1.28E-121	0.719688	1.67E-176	0.365492	3.67E-36
HAVCR2	0.465557	2.29E-60	0.371185	2.51E-37	0.148284	7.64E-07
ICOS	0.631108	1.74E-123	0.67846	1.70E-149	0.406015	5.52E-45
IGSF6	0.554317	8.87E-90	0.46468	4.06E-60	0.197137	4.06E-11
IL10	0.459602	1.09E-58	0.350066	4.05E-33	0.175888	4.15E-09

(Continued on following page)

TABLE 2 (*Continued*) **Correlation analyses of IFNG, AC006369.1, and CCR7 with immune checkpoints in BRCA.**

Immune checkpoint	IFNG		AC006369.1		CCR7	
	r	p	r	p	r	p
IL1R1	0.049448	0.100874	0.097552	0.001185	0.039495	0.190162
IL1R2	0.056551	0.060567	0.023964	0.426762	0.010433	0.729378
IL2RA	0.421939	8.40E-49	0.330214	1.91E-29	0.182877	9.62E-10
IRF4	0.420006	2.50E-48	0.498811	2.21E-70	0.283082	9.40E-22
LAG3	0.769047	4.90E-216	0.377368	1.28E-38	0.188985	2.56E-10
MS4A1	0.219143	1.89E-13	0.667616	4.08E-143	0.423381	3.70E-49
PDCD1	0.72552	1.02E-180	0.667504	4.74E-143	0.343997	5.74E-32
SDC1	−0.04353	0.148748	−0.07891	0.008775	−0.05836	0.052782
SGPP2	0.160109	9.12E-08	0.035293	0.241745	0.016745	0.578712
SH2D1A	0.613515	6.18E-115	0.810235	1.76E-257	0.445128	9.67E-55
STAT5A	0.208098	3.02E-12	0.274147	1.89E-20	0.203145	9.96E-12
TIGIT	0.608306	1.66E-112	0.826834	3.57E-277	0.464217	5.50E-60
TNFRSF17	0.366472	2.32E-36	0.493874	7.97E-69	0.24518	1.50E-16
TNFRSF18	−0.02112	0.483628	−0.01387	0.64559	−0.01949	0.518006
TRAF6	0.081821	0.006575	0.076896	0.010663	0.025213	0.403071

Associations of IFNG, and co-expressed AC006369.1, and CCR7 with the TME components

IFNG was negatively connected with macrophages M0 and M2, mast cells resting, but was positively linked with macrophages M1, T cells CD4 memory resting and activated, T cells CD8, T cells follicular helper, T cells regulatory (Tregs), and NK cells resting (Figure 5A). This was indicative of the role of IFNG in regulating anti-tumor immunity. In Figure 5B, AC006369.1 presented the negative interactions with neutrophils, macrophages M0 and M2, NK cells activated, dendritic cells activated, and mast cells resting, with positive interactions with B cells naïve and memory, macrophages M1, T cells CD4 memory resting and activated, T cells CD8, Tregs, and NK cells resting. In addition, CCR7 exhibited the positive relationships with B cells naïve and memory, macrophages M1, T cells CD4 memory resting and activated, T cells CD8, and NK cells resting, with negative relationships with macrophages M0 and M2, NK cells activated, and mast cells resting (Figure 5C).

Interactions of IFNG, and co-expressed genes with immune checkpoints

As illustrated in Figure 6A; Table 2, IFNG, co-expressed genes (notably AC006369.1, and CCR7), and the prognostic model exhibited the positive connections with most immune checkpoint molecules. It was also noted the positive interactions of IFNG with CD274 (PD-1), LAG3, and PDCD1 (Figures 6B–D).

Genetic alterations and DNA methylation of IFNG, and co-expressed genes

Most IFNG, and co-expressed genes occurred frequent mutation across BRCA samples, such as CCR7 (6%), and IFNG (3%) (Figures 7A, B). In addition, frequent amplifications were found, which might contribute to their overexpression (Figure 7C). DNA methylation sites were also analyzed (Figure 7D). IFNG expression was positively connected with the beta value of cg01281450, with negative connections with the beta values of cg05224770, and cg26227465 (Figures 7E–G). Among the three CpGs, cg01281450 exhibited the lower beta value in BRCA *versus* controls, with lower value in high *versus* low IFNG expression tumors (Figure 7H). This indicated the contribution of cg05224770 hypomethylation to IFNG upregulation. Moreover, CCR7 expression exhibited the negative interactions with the beta values of cg07388018, cg13504059, cg17067993, cg07248223, cg16047279, cg23663547, cg26960939, and cg07479709, with positive interactions with the beta value of cg11729107 (Figures 7I–Q). Among the CpGs, cg07388018 owned the lower beta value in tumors with IFNG upregulation *versus* controls or tumors with IFNG downregulation (Figure 7R). Thus, hypomethylated cg07388018 possibly resulted in CCR7 overexpression.

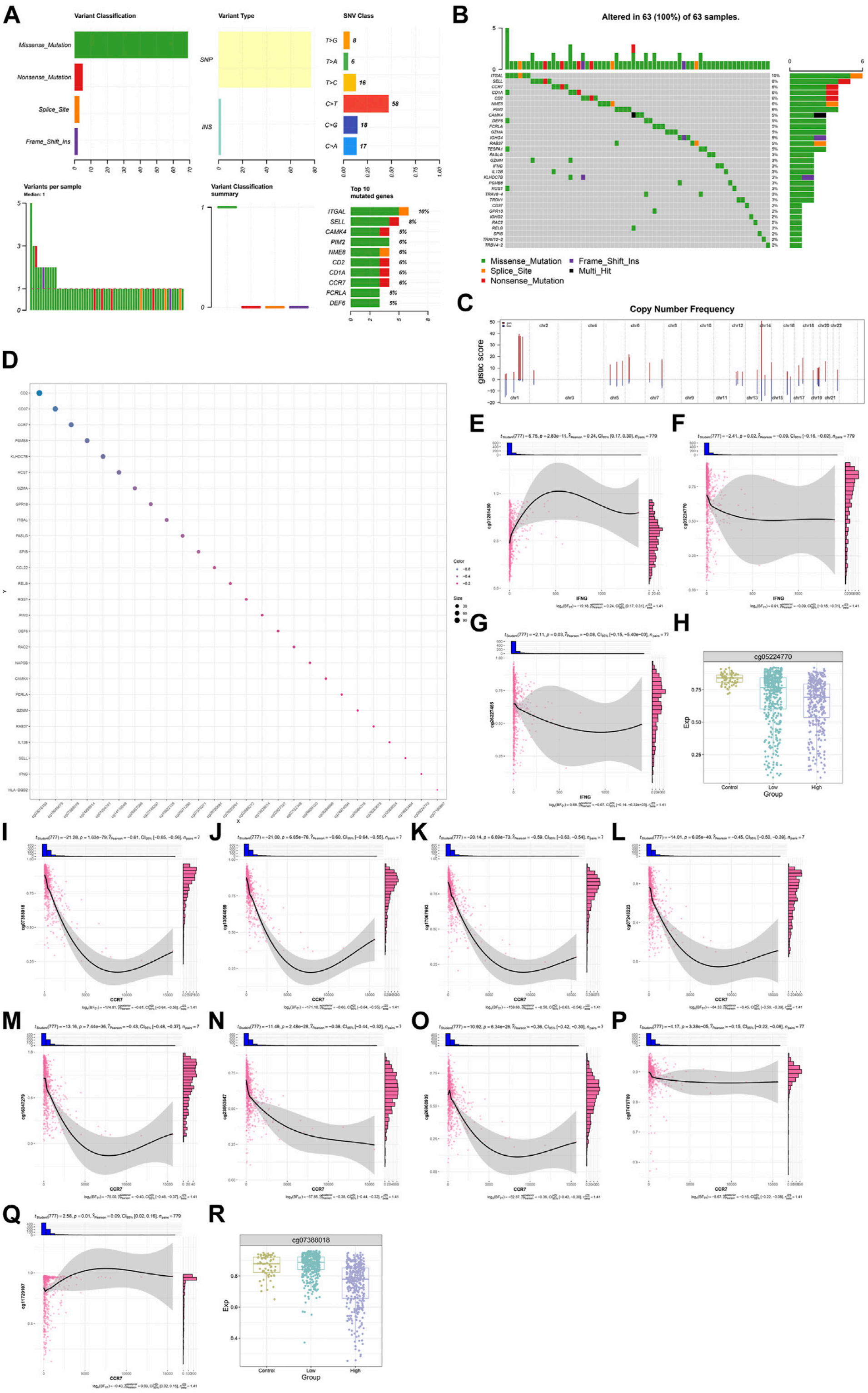


FIGURE 7
Genetic alterations and DNA methylation of IFNG, and co-expressed genes. **(A)** The summary of somatic mutations in BRCA. **(B)** The mutated frequency of IFNG, and co-expressed genes across BRCA. **(C)** Copy-number frequency of above genes. **(D)** Methylation sites of above genes. **(E–G)** Relationships of IFNG with **(E)** cg01281450, **(F)** cg05224770, and **(G)** cg26227465. **(H)** The beta level of cg05224770 in controls, BRCA with lowly or highly expressed IFNG. (I–Q) Associations of CCR7 with **(I)** cg07388018, **(J)** cg13504059, **(K)** cg17067993, **(L)** cg07248223, **(M)** cg16047279, **(N)** cg23663547, **(O)** cg26960939, **(P)** cg07479709, and **(Q)** cg11729107. **(R)** The beta level of cg07388018 across normal specimens, BRCA with down- or upregulated IFNG.

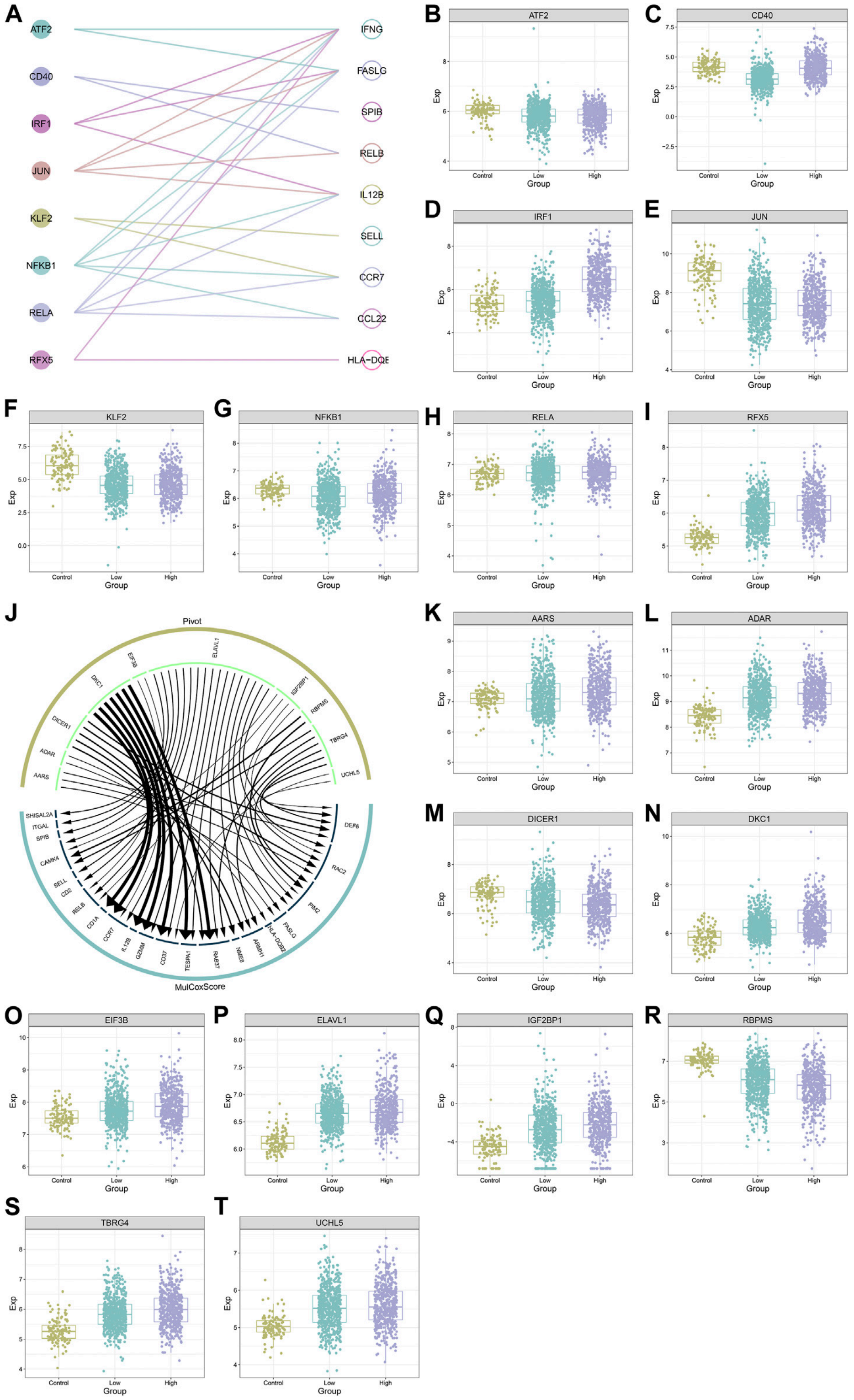


FIGURE 8
Transcription factors and RNA binding proteins that potentially modulate IFNG and co-expressed genes. **(A)** Interactions of IFNG and co-expressed genes with transcription factors. **(B–I)** The transcript level of **(B)** ATF2, **(C)** CD40, **(D)** IRF1, **(E)** JUN, **(F)** KLF2, **(G)** NFKB1, **(H)** RELA, and **(I)** RFX5 across controls, BRCA with down- or upregulated IFNG. **(J)** The interaction network of IFNG-co-expressed genes with RNA binding proteins. **(K–T)** The transcript level of **(K)** AARS, **(L)** ADAR, **(M)** DICER1, **(N)** DKC1, **(O)** EIF3B, **(P)** ELAVL1, **(Q)** IGF2BP1, **(R)** RBPMS, **(S)** TBRG4, and **(T)** UCHL5 in normal tissues, BRCA with down- or upregulated IFNG.

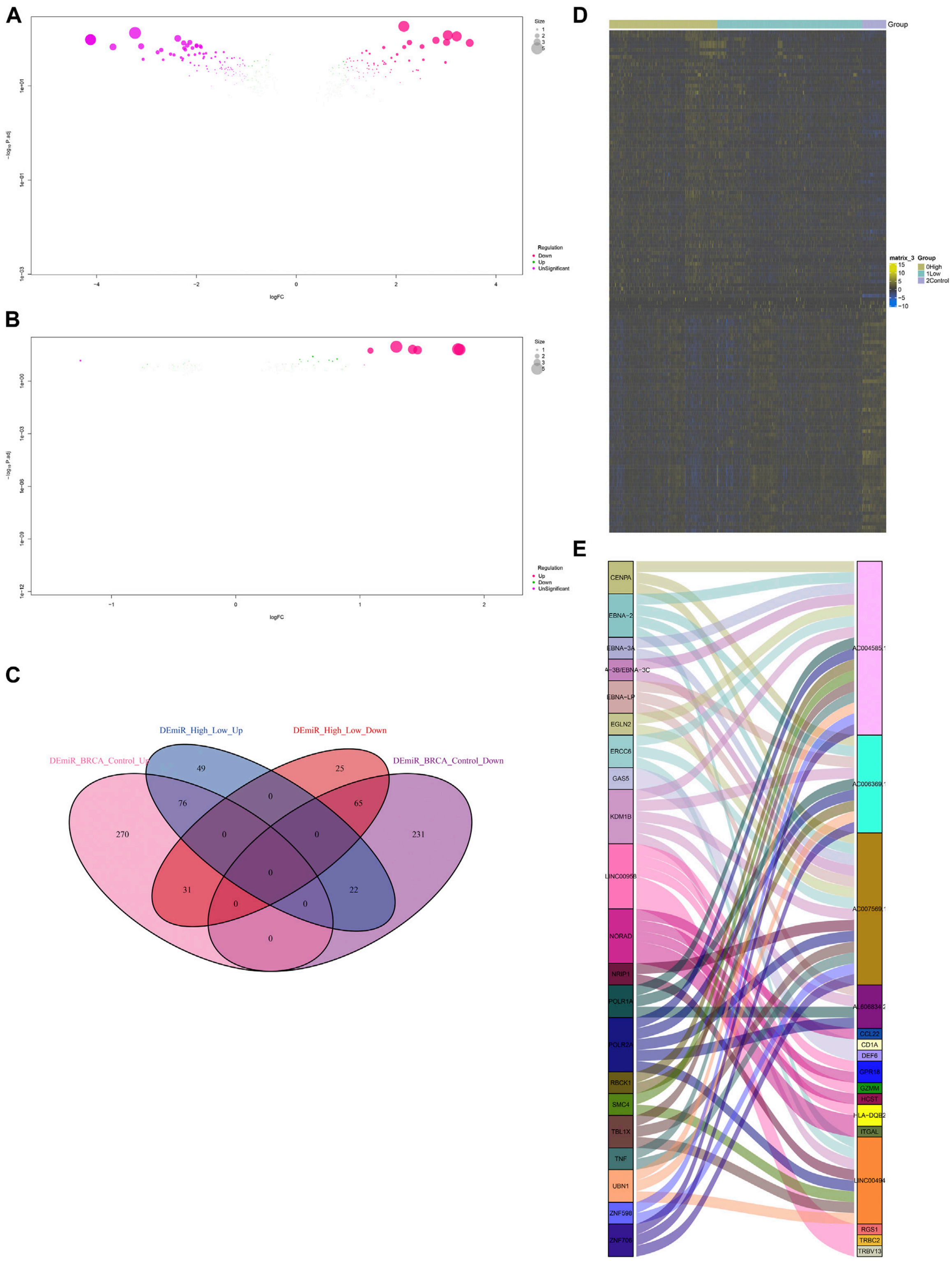


FIGURE 9
MiRNAs and lncRNAs that possibly regulate IFNG and co-expressed genes. **(A)** MiRNAs with aberrant expression in BRCA *versus* normal specimens. **(B)** MiRNAs with different expression between down- and upregulated IFNG BRCA. **(C)** Venn diagram illustrating the BRCA- and IFNG-relevant miRNAs. **(D)** The expression of above miRNAs across normal tissues, BRCA with down- or upregulated IFNG. **(E)** Correlations of IFNG-co-expressed genes with lncRNAs.

Transcription factors and RNA binding proteins that potentially modulate IFNG and co-expressed genes

Figure 8A illustrates eight transcription factors potentially modulating the transcription of IFNG and co-expressed genes, as follows: ATF2 (IFNG, FASLG), CD40 (SPIB, RELB), IRF1 (IL12B, FASLG, IFNG), JUN (FASLG, IL12B, IFNG, RELB), KLF2 (SELL, CCR7), NFKB1 (CCR7, IFNG, IL12B, CCL22, FASLG), RELA (FASLG, IL12B, CCL22, CCR7, IFNG), RFX5 (HLA-DQB2, IFNG). Additionally, these transcription factors exhibited the aberrant expression in BRCA *versus* controls (Figures 8B–I). The heterogeneity in their expression was also found between down- or upregulated IFNG tumors. Ten RNA-binding proteins post-transcriptionally modulated IFNG-co-expressed genes, following AARS (DEF6, RAC2, PIM2, FASLG, HLA-DQB2), DICER1 (ARMH1, RAC2, DEF6, NME8, RAB37), DKC1 (TESPA1, RAB37, CD37, GZMM, IL12B, CCR7, CD1A), EIF3B (CCR7, RELB), ELAVL1 (CD2, RAC2, SELL, ARMH1, CAMK4, DEF6, SPIB, FASLG, CD37, CCR7, TESPA1, RELB, ITGAL, GZMM, RAB37, PIM2, SHISAL2A), IGF2BP1 (PIM2, RELB, CAMK4, RAC2), RBPMS (PIM2, CAMK4), TBRG4 (DEF6, PIM2, ARMH1, CAMK4, RAB37, RAC2), and UCHL5 (DEF6, CD37) (Figure 8J). Except for DICER1, other RNA-binding proteins were upregulated in BRCA (Figures 8K–T).

MiRNAs and lncRNAs that possibly regulate IFNG and co-expressed genes

Non-coding RNA-mediated post-transcriptional mechanisms of IFNG and co-expressed genes were also probed. In Figure 9A, 695 miRNAs with aberrant expression were determined in BRCA relative to controls. Additionally, 268 miRNAs exhibited the different expression between lowly and highly expressed IFNG tumors (Figure 9B). Following the intersection, 141 BRCA- and IFNG-relevant miRNAs were selected, which were possibly associated with IFNG expression (Figures 9C, D; Supplementary Table S3). Several lncRNAs were then observed to be potentially interacted with IFNG-co-expressed genes (Figure 9E).

Discussion

IFNG presented the upregulation in BRCA, as priorly reported (Yaghoobi et al., 2018). Also, the upregulation was associated with favorable OS and RFS outcomes. Thus, IFNG might own the potential as a prognostic marker of BRCA. Two IFNG-co-expressed RNAs (AC006369.1, and CCR7) constituted a Cox regression model for BRCA prognostication. AC006369.1, and CCR7 were aberrantly expressed in BRCA, and in relation to survival outcomes. Similarly, Gu et al. identified AC006369.1 as an IFNG-relevant lncRNA that was connected with prognostic outcomes and the TME in uterine corpus endometrial carcinoma (Gu et al., 2022). Many studies have proven the essential function of CCR7 in BRCA. For instance, CXCL12 facilitates CCR7 ligand-driven BRCA cell invasion and migration towards lymphatic vessels (Hayasaka et al., 2022). Deng et al. reported that site-specific polyplex on downregulated CCR7 increases T cells for hindering lymphatic metastasis of BRCA (Deng et al., 2022). In addition, CCR7 chemokine receptor stimulation can induce rapid but transient dendritic cell migration towards draining lymph nodes, which is crucial for initiating protective immunity and maintaining immune homeostasis (Liu et al., 2019).

IFNG presented the negative connections with macrophages M0 and M2, mast cells resting, with the positive correlations to macrophages M1, T cells CD4 memory resting and activated, T cells CD8, T cells follicular helper, Tregs, and NK cells resting. The interactions of IFNG with such immune cells have been unveiled. For instance, tumor-associated macrophages accelerate metastases as well as hinder T cells. Non-etheless, macrophage polarization is capable of killing malignant cells. IFN-γ can reprogram CD206+ tumor-associated macrophages to inducible iNOS + macrophages in BRCA (Sun et al., 2021). Tregs maintain BRCA progression through manipulating IFN-γ-driven functional reprogramming of myeloid cells (Clark et al., 2020). IFN-γ impairs the cytotoxicity of NK cells via upregulation of PD-L1 on malignant cells as well as PD-1 on NK cells in trastuzumab-resistant HER2-positive BRCA (Zheng et al., 2021). IFN-γ-triggered intermediate monocytes hinder cancer metastasis through activating NK cells (Wang R. et al., 2022). The interactions of IFNG-co-expressed genes (especially AC006369.1 and CCR7) with the TME components were also investigated across BRCA.

Immunotherapy exhibits effective therapeutic potential for long-term cancer regression, but exerts a low response rate owing to insufficient immunogenicity of malignant cells (Tian et al., 2023). IFN-γ is an essential driver of PD1/PD-L1 expression in tumor and host cells. In addition, IFN-γ is capable of upregulating expression of other critical immune suppressive molecules within the TME. Mark Ayers et al. proposed an IFNG-relevant mRNA signature that can predict clinical response to anti-PD-1 therapy (Ayers et al., 2017). Nevertheless, the pleiotropic effects of IFN-γ on immunotherapy have been found, such as immunotherapeutic resistance. IFN-γ-driven adaptive resistance remains one barrier to the improvement in immunotherapy. In the Cucolo et al.'s study, IFN-γ-driven RIPK1 enhances malignant cell intrinsic as well as extrinsic resistance to immunotherapy (Cucolo et al., 2022). UBR5 facilitates tumor immune escape via elevating IFN-γ-driven PDL1 transcription in BRCA (Wu B. et al., 2022). This work also exhibited the close connections of IFNG with immune checkpoints in BRCA, proving the potential in improving immunotherapy.

The regulatory mechanisms of IFNG and co-expressed genes were further probed. It was found that somatic mutation frequencies of CCR7, and IFNG were separately 6%, and 3%. Frequent amplification also potentially led to their upregulation. Hypomethylated cg05224770 and cg07388018 might associate with IFNG and CCR7 upregulation. IFNG expression can be transcriptionally modulated by ATF2, IRF1, JUN, NFKB1, RELA, and RFX5. Among them, IRF1 has been proven as an IFNG-inducible gene (Qian et al., 2018). IFN-γ-induced IRF-1 attenuates BRCA cell specific growth (Armstrong et al., 2015). RNA-binding proteins (AARS, ADAR, DICER1, DKC1, EIF3B, ELAVL1, IGF2BP1, RBPMS, TBRG4, and UCHL5) and non-coding RNAs also post-transcriptionally affected IFNG and co-expressed genes. The interactions of IFN-γ with ADAR and DICER1 have been partly proven. ADAR (an interferon-inducible RNA-editing enzyme) mitigates IFN signaling in gastric carcinoma through down-regulating STAT1 and IRF9 by miR-302a (Jiang et al., 2020). DICER1 hinders the interferon response in murine embryonic stem cells (Gurung et al., 2021). Invasive micropapillary carcinoma is a rare histological subtype of BRCA with an aggressive phenotype and an undesirable prognosis (Verras et al., 2022a). Invasive micropapillary carcinoma has a high rate of lymphovascular invasion and lymph node metastasis, and has been reported in multiple organs (Verras et al., 2022b; Shi et al., 2022). However, so far, no studies have reported the role of inflammatory cell death-related IFNG and co-expressed RNAs (AC006369.1, and CCR7) in this subtype.

The limitations of our work require to be acknowledged. Despite the close connections of IFNG and co-expressed genes with the TME and immune checkpoint molecules, their roles in anti-tumor immunity need experimental verification. Moreover, further analyses are required for proving the regulatory mechanisms of IFNG and co-expressed genes in BRCA.

Conclusion

Altogether, this work characterized IFNG and its co-expressed RNAs (notably AC006369.1, and CCR7) as prognostic markers for BRCA individuals, and unveiled their potential as therapeutic targets for the improvement of immunotherapy. Despite this, in-depth experiments will be implemented for proving our conclusions in future 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.

Author contributions

XH conceived and designed the study. YD, ZL, and MP conducted most of the experiments and data analysis, and wrote the manuscript. HW and BN participated in collecting data and helped to draft the manuscript. All authors contributed to the article and approved the submitted version.

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.2023.1112251/full#supplementary-material

References

Armstrong, M. J., Stang, M. T., Liu, Y., Yan, J., Pizzoferrato, E., and Yim, J. H. (2015). IRF-1 inhibits NF-κB activity, suppresses TRAF2 and cIAP1 and induces breast cancer cell specific growth inhibition. *Cancer Biol. Ther.* 16 (7), 1029–1041. doi:10.1080/15384047.2015.1046646

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/jci91190

Bidard, F. C., Kaklamani, V. G., Neven, P., Streich, G., Montero, A. J., Forget, F., et al. (2022). Elacestrant (oral selective estrogen receptor degrader) versus standard endocrine therapy for estrogen receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: Results from the randomized phase III EMERALD trial. *J. Clin. Oncol.* 40 (28), 3246–3256. doi:10.1200/jco.22.00338

Clark, N. M., Martinez, L. M., Murdock, S., deLigio, J. T., Olex, A. L., Effi, C., et al. (2020). Regulatory T cells support breast cancer progression by opposing IFN-γ-Dependent functional reprogramming of myeloid cells. *Cell Rep.* 33 (10), 108482. doi:10.1016/j.celrep.2020.108482

Corti, C., Giachetti, P., Eggermont, A. M. M., Delaloge, S., and Curigliano, G. (2022). Therapeutic vaccines for breast cancer: Has the time finally come? *Eur. J. Cancer* 160, 150–174. doi:10.1016/j.ejca.2021.10.027

Cucolo, L., Chen, Q., Qiu, J., Yu, Y., Klapholz, M., Budinich, K. A., et al. (2022). The interferon-stimulated gene RIPK1 regulates cancer cell intrinsic and extrinsic resistance to immune checkpoint blockade. *Immunity* 55 (4), 671–685.e10. doi:10.1016/j.immuni.2022.03.007

Curigliano, G., Mueller, V., Borges, V., Hamilton, E., Hurvitz, S., Loi, S., et al. (2022). Tucatinib versus placebo added to trastuzumab and capecitabine for patients with pretreated HER2+ metastatic breast cancer with and without brain metastases (HER2CLIMB): Final overall survival analysis. *Ann. Oncol.* 33 (3), 321–329. doi:10.1016/j.annonc.2021.12.005

Deng, Y., Tan, C., Huang, S., Sun, H., Li, Z., Li, J., et al. (2022). Site-specific polyplex on CCR7 down-regulation and T cell elevation for lymphatic metastasis blocking on breast cancer. *Adv. Healthc. Mater* 11 (22), e2201166. doi:10.1002/adhm.202201166

Dörrie, J., Schuh, W., Keil, A., Bongards, E., Greil, J., Fey, G. H., et al. (1999). Regulation of CD95 expression and CD95-mediated cell death by interferon-gamma in acute lymphoblastic leukemia with chromosomal translocation t(4;11). *Leukemia* 13 (10), 1539–1547. doi:10.1038/sj.leu.2401479

Gadaleta, E., Thorn, G. J., Ross-Adams, H., Jones, L. J., and Chelala, C. (2022). Field cancerization in breast cancer. *J. Pathol.* 257 (4), 561–574. doi:10.1002/path.5902

Gu, C., Lin, C., Zhu, Z., Hu, L., Wang, F., Wang, X., et al. (2022). The IFN-γ-related long non-coding RNA signature predicts prognosis and indicates immune microenvironment infiltration in uterine corpus endometrial carcinoma. *Front. Oncol.* 12, 955979. doi:10.3389/fonc.2022.955979

Gurung, C., Fendereski, M., Sapkota, K., Guo, J., Huang, F., and Guo, Y. L. (2021). Dicer represses the interferon response and the double-stranded RNA-activated protein kinase pathway in mouse embryonic stem cells. *J. Biol. Chem.* 296, 100264. doi:10.1016/j.jbc.2021.100264

Hayasaka, H., Yoshida, J., Kuroda, Y., Nishiguchi, A., Matsusaki, M., Kishimoto, K., et al. (2022). CXCL12 promotes CCR7 ligand-mediated breast cancer cell invasion and migration toward lymphatic vessels. *Cancer Sci.* 113 (4), 1338–1351. doi:10.1111/cas.15293

Huober, J., Barrios, C. H., Niikura, N., Jarząb, M., Chang, Y. C., Huggins-Puhalla, S. L., et al. (2022). Atezolizumab with neoadjuvant anti-human epidermal growth factor receptor 2 therapy and chemotherapy in human epidermal growth factor receptor 2-positive early breast cancer: Primary results of the randomized phase III IMpassion050 trial. *J. Clin. Oncol.* 40 (25), 2946–2956. doi:10.1200/jco.21.02772

Jiang, L., Park, M. J., Cho, C. J., Lee, K., Jung, M. K., Pack, C. G., et al. (2020). ADAR1 suppresses interferon signaling in gastric cancer cells by MicroRNA-302a-mediated IRF9/STAT1 regulation. *Int. J. Mol. Sci.* 21 (17), 6195. doi:10.3390/ijms21176195

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

Liao, P., Wang, W., Wang, W., Kryczek, I., Li, X., Bian, Y., et al. (2022). CD8(+) T cells and fatty acids orchestrate tumor ferroptosis and immunity via ACSL4. *Cancer Cell* 40 (4), 365–378.e6. doi:10.1016/j.ccell.2022.02.003

Liu, J., Zhang, X., Chen, K., Cheng, Y., Liu, S., Xia, M., et al. (2019). CCR7 chemokine receptor-inducible lnc-dpf3 restrains dendritic cell migration by inhibiting HIF-1α-Mediated glycolysis. *Immunity* 50 (3), 600–615. doi:10.1016/j.immuni.2019.01.021

Loibl, S., Untch, M., Burchardi, N., Huober, J., Sinn, B. V., Blohmer, J. U., et al. (2019). A randomised phase II study investigating durvalumab in addition to an anthracycline taxane-based neoadjuvant therapy in early triple-negative breast cancer: Clinical results and biomarker analysis of GeparNuevo study. *Ann. Oncol.* 30 (8), 1279–1288. doi:10.1093/annonc/mdz158

Martin, M., Hegg, R., Kim, S. B., Schenker, M., Grecea, D., Garcia-Saenz, J. A., et al. (2022). Treatment with adjuvant abemaciclib plus endocrine therapy in patients with high-risk early breast cancer who received neoadjuvant chemotherapy: A prespecified analysis of the monarchE randomized clinical trial. *JAMA Oncol.* 8 (8), 1190–1194. doi:10.1001/jamaoncol.2022.1488

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

Mayer, E. L., Fesl, C., Hlauschek, D., Garcia-Estevez, L., Burstein, H. J., Zdenkowski, N., et al. (2022). Treatment exposure and discontinuation in the PALbociclib CoLlaborative adjuvant study of palbociclib with adjuvant endocrine therapy for hormone receptor-positive/human epidermal growth factor receptor 2-negative early breast cancer (PALLAS/AFT-05/ABCSG-42/BIG-14-03). *J. Clin. Oncol.* 40 (5), 449–458. doi:10.1200/jco.21.01918

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

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

Niu, X., Huang, B., Qiao, X., Liu, J., Chen, L., and Zhong, M. (2021). MicroRNA-1-3p suppresses malignant phenotypes of ameloblastoma through down-regulating lysosomal associated membrane protein 2-mediated autophagy. *Front. Med. (Lausanne)* 8, 670188. doi:10.3389/fmed.2021.670188

Qian, J., Wang, C., Wang, B., Yang, J., Wang, Y., Luo, F., et al. (2018). The IFN-γ/PD-L1 axis between T cells and tumor microenvironment: Hints for glioma anti-PD-1/PD-L1 therapy. *J. Neuroinflammation* 15 (1), 290. doi:10.1186/s12974-018-1330-2

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

Schmid, P., Rugo, H. S., Adams, S., Schneeweiss, A., Barrios, C. H., Iwata, H., et al. (2020). Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): Updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 21 (1), 44–59. doi:10.1016/s1470-2045(19)30689-8

Shepherd, J. H., Ballman, K., Polley, M. C., Campbell, J. D., Fan, C., Selitsky, S., et al. (2022). CALGB 40603 (alliance): Long-term outcomes and genomic correlates of response and survival after neoadjuvant chemotherapy with or without carboplatin and bevacizumab in triple-negative breast cancer. *J. Clin. Oncol.* 40 (12), 1323–1334. doi:10.1200/jco.21.01506

Shi, Q., Shao, K., Jia, H., Cao, B., Li, W., Dong, S., et al. (2022). Genomic alterations and evolution of cell clusters in metastatic invasive micropapillary carcinoma of the breast. *Nat. Commun.* 13 (1), 111. doi:10.1038/s41467-021-27794-4

Simpson, D. S., Pang, J., Weir, A., Kong, I. Y., Fritsch, M., Rashidi, M., et al. (2022). Interferon-γ primes macrophages for pathogen ligand-induced killing via a caspase-8 and mitochondrial cell death pathway. *Immunity* 55 (3), 423–441.e9. doi:10.1016/j.immuni.2022.01.003

Sun, L., Kees, T., Almeida, A. S., Liu, B., He, X. Y., Ng, D., et al. (2021). Activating a collaborative innate-adaptive immune response to control metastasis. *Cancer Cell* 39 (10), 1361–1374.e9. doi:10.1016/j.ccell.2021.08.005

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

Tian, Y., Xiao, H., Yang, Y., Zhang, P., Yuan, J., Zhang, W., et al. (2023). Crosstalk between 5-methylcytosine and N(6)-methyladenosine machinery defines disease progression, therapeutic response and pharmacogenomic landscape in hepatocellular carcinoma. *Mol. Cancer* 22 (1), 5. doi:10.1186/s12943-022-01706-6

Verras, G. I., Mulita, F., Tchabashvili, L., Grypari, I. M., Sourouni, S., Panagodimou, E., et al. (2022a). A rare case of invasive micropapillary carcinoma of the breast. *Prz. Menopauzalny* 21 (1), 73–80. doi:10.5114/pm.2022.113834

Verras, G. I., Tchabashvili, L., Mulita, F., Grypari, I. M., Sourouni, S., Panagodimou, E., et al. (2022b). Micropapillary breast carcinoma: From molecular pathogenesis to prognosis. *Breast Cancer* 14, 41–61. doi:10.2147/bctt.S346301

Wang, R., Bao, W., Pal, M., Liu, Y., Yazdanbakhsh, K., and Zhong, H. (2022a). Intermediate monocytes induced by IFN-γ inhibit cancer metastasis by promoting NK cell activation through FOXO1 and interleukin-27. *J. Immunother. Cancer* 10 (1), e003539. doi:10.1136/jitc-2021-003539

Wang, Z., Yao, J., Dong, T., and Niu, X. (2022b). Definition of a novel cuproptosis-relevant lncRNA signature for uncovering distinct survival, genomic alterations, and treatment implications in lung adenocarcinoma. *J. Immunol. Res.* 2022, 2756611. doi:10.1155/2022/2756611

Wei, T. T., Zhang, M. Y., Zheng, X. H., Xie, T. H., Wang, W., Zou, J., et al. (2022). Interferon-γ induces retinal pigment epithelial cell Ferroptosis by a JAK1-2/STAT1/SLC7A11 signaling pathway in Age-related Macular Degeneration. *Febs J.* 289 (7), 1968–1983. doi:10.1111/febs.16272

Witek Janusek, L., Tell, D., and Mathews, H. L. (2019). Mindfulness based stress reduction provides psychological benefit and restores immune function of women newly diagnosed with breast cancer: A randomized trial with active control. *Brain Behav. Immun.* 80, 358–373. doi:10.1016/j.bbi.2019.04.012

Wu, B., Song, M., Dong, Q., Xiang, G., Li, J., Ma, X., et al. (2022a). UBR5 promotes tumor immune evasion through enhancing IFN-γ-induced PDL1 transcription in triple negative breast cancer. *Theranostics* 12 (11), 5086–5102. doi:10.7150/thno.74989

Wu, H., Li, Y., Shi, G., Du, S., Wang, X., Ye, W., et al. (2022b). Hepatic interferon regulatory factor 8 expression suppresses hepatocellular carcinoma progression and enhances the response to anti-programmed cell death protein-1 therapy. *Hepatology* 76 (6), 1602–1616. doi:10.1002/hep.32316

Yaghoobi, H., Azizi, H., Oskooei, V. K., Taheri, M., and Ghafouri-Fard, S. (2018). Assessment of expression of interferon γ (IFN-G) gene and its antisense (IFNG-AS1) in breast cancer. *World J. Surg. Oncol.* 16 (1), 211. doi:10.1186/s12957-018-1508-1

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

Zheng, G., Guo, Z., Li, W., Xi, W., Zuo, B., Zhang, R., et al. (2021). Interaction between HLA-G and NK cell receptor KIR2DL4 orchestrates HER2-positive breast cancer resistance to trastuzumab. *Signal Transduct. Target Ther.* 6 (1), 236. doi:10.1038/s41392-021-00629-w

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 10 August 2023
DOI 10.3389/fgene.2023.1113137

EDITED BY
Xiang Xue,
University of New Mexico, United States

REVIEWED BY
Xisong Liang,
Central South University, China
Mojgan Alaeddini,
Tehran University of Medical Sciences, Iran
Gokul S,
Rutgers, The State University of New Jersey, United States

*CORRESPONDENCE
Qicheng Fu,
✉ fqc3329699@163.com

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

RECEIVED 01 December 2022
ACCEPTED 11 July 2023
PUBLISHED 10 August 2023

CITATION
Ai J, Tan Y, Liu B, Song Y, Wang Y, Xia X and Fu Q (2023), Systematic establishment and verification of an epithelial-mesenchymal transition gene signature for predicting prognosis of oral squamous cell carcinoma.
Front. Genet. 14:1113137.
doi: 10.3389/fgene.2023.1113137

Systematic establishment and verification of an epithelial-mesenchymal transition gene signature for predicting prognosis of oral squamous cell carcinoma

Jun Ai[1†], Yaqin Tan[1†], Bo Liu[2], Yuhong Song[3], Yanqin Wang[1], Xin Xia[1] and Qicheng Fu[3]*

[1]Department of Stomatology, Taihe Hospital, Hubei University of Medicine, Shiyan, Hubei, China, [2]Department of Urology, Taihe Hospital, Hubei University of Medicine, Shiyan, China, [3]Department of Stomatology, Shenzhen Longhua District Central Hospital, Shenzhen, China

Objective: Epithelial-mesenchymal transition (EMT) is linked to an unfavorable prognosis in oral squamous cell carcinoma (OSCC). Here, we aimed to develop an EMT gene signature for OSCC prognosis.

Methods: In TCGA dataset, prognosis-related EMT genes with $p < 0.05$ were screened in OSCC. An EMT gene signature was then conducted with LASSO method. The efficacy of this signature in predicting prognosis was externally verified in the GSE41613 dataset. Correlations between this signature and stromal/immune scores and immune cell infiltration were assessed by ESTIMATE and CIBERSORT algorithms. GSEA was applied for exploring significant signaling pathways activated in high- and low-risk phenotypes. Expression of each gene was validated in 40 paired OSCC and normal tissues via RT-qPCR.

Results: A prognostic 9-EMT gene signature was constructed in OSCC. High risk score predicted poorer clinical outcomes than low risk score. ROCs confirmed the well performance on predicting 1-, 3- and 5-year survival. Multivariate cox analysis revealed that this signature was independently predictive of OSCC prognosis. The well predictive efficacy was validated in the GSE41613 dataset. Furthermore, this signature was distinctly related to stromal/immune scores and immune cell infiltration in OSCC. Distinct pathways were activated in two subgroups. After validation, AREG, COL5A3, DKK1, GAS1, GPX7 and PLOD2 were distinctly upregulated and SFRP1 was downregulated in OSCC than normal tissues.

Conclusion: Our data identified and verified a robust EMT gene signature in OSCC, which provided a novel clinical tool for predicting prognosis and several targets against OSCC therapy.

Abbreviations: AUC, area under the curve; CI, confidence interval; EMT, epithelial-mesenchymal transition; ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; GEO, Gene Expression Omnibus; GSEA, Gene Set Enrichment Analysis; HR, hazard ratio; LASSO, least absolute shrinkage and selection operator; NES, normalized enrichment score; OSCC, oral squamous cell carcinoma; OS, overall survival; ROC, receiver operating characteristic; RT-qPCR, real-time quantitative polymerase chain reaction; TCGA, The Cancer Genome Atlas.

KEYWORDS

oral squamous cell carcinoma, epithelial-mesenchymal transition, gene signature, prognosis, immune microenvironment

Introduction

Oral squamous cell carcinoma (OSCC) represents the predominant type of head and neck squamous cell carcinoma (Park et al., 2019). Surgery, radiotherapy, as well as chemotherapy are the main therapeutic strategies of OSCC (Zhao et al., 2020). The 5-year survival rate is only 50% because of regional invasion as well as lymph node/distant metastases (Liu et al., 2020). Conventional prognostic factors, e.g., stage, are far from optimal (Omori et al., 2020). Although many researches have proposed prognostic markers for OSCC, most of them only focused on several well-studied markers (Almangush et al., 2017). Furthermore, these researches have been carried out in small cohorts, which is difficult to utilize these molecular markers for predicting OSCC prognosis in daily clinical practice (Ju et al., 2020).

Epithelial-mesenchymal transition (EMT) is a dynamic process in which epithelial cells acquire mesenchymal features (Ling et al., 2021), leading to the upregulation of migratory and invasive capacities of tumor cells (Qiao et al., 2020). OSCC primarily contains epithelial dysplasia, loss of epithelial differentiation as well as acquisition of mesenchymal phenotype (Bai et al., 2020). It has been confirmed that EMT process is in relation to OSCC invasiveness and metastasis (Peng et al., 2020). However, it remains vacant concerning the EMT gene signatures and their prognostic value in OSCC. Because of the easy accessibility of gene expression profiles from public databases, exploration of the prognostic gene signatures has been given wide attention (Wu et al., 2019). Based upon the critical role of EMT process in OSCC progression, it is of significance to establish an EMT gene signature for OSCC prognosis. Thus, our research set out to further understand the underlying clinical utility of EMT genes as prognostic markers and to build up individualized clinical outcome evaluation for OSCC.

Materials and methods

Data collection

Normalized transcriptome data and clinical information of OSCC were retrieved from the Cancer Genome Atlas (TCGA) database via the UCSC Xena (https://tcga.xenahubs.net) on 11 March 2020. Moreover, the GSE41613 dataset was downloaded from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database (Lohavanichbutr et al., 2013). The specific clinical information was listed in Supplementary Table S1. Before surgery, all patients did not receive radiotherapy or chemotherapy. Primary tumor site of all patients was the same. The "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION" gene set containing 200 genes (Supplementary Table S2) was obtained from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp).

Data preprocessing

In the TCGA dataset, 328 OSCC patients with complete clinical information were regarded as the discovery set. In the GSE41613 dataset, raw microarray data of 97 patients that possessed complete survival information were pre-corrected, transformed by log2 and normalized, followed by gene annotation. This dataset was utilized as the validation set.

Establishment of an EMT gene model

Prognosis-related EMT genes were firstly screened in OSCC. Univariate cox analyses were employed to screen EMT genes with $p < 0.05$ in the TCGA dataset. By applying glmnet package, an optimal prognostic model was built with least absolute shrinkage and selection operator (LASSO) Cox regression analysis based on the prognosis-related EMT genes (Friedman et al., 2010). The optimal value of penalty parameter λ was determined via ten-fold cross-validation. The risk score of OSCC samples was computed according to regression coefficient as well as expression value of each gene in this model. The formula of the risk score was as follows: risk score $= \sum_{i=1}^{n} (LASSO\ coefficient\ of\ gene\ i * expression\ of\ gene\ i)$. OSCC patients in discovery and validation datasets were separated into high- and low-risk subgroups according to the median value of risk score, respectively. High-risk subgroup was defined as patients who had risk scores greater than the median while low-risk subgroup was defined as patients who had risks scores less than the median. To compare the difference in overall survival (OS) time between two subgroups, Kaplan–Meier curves were depicted by survival package and difference was determined with log-rank tests. Time-dependent receiver operating characteristic (ROC) curves of 1-year, 3-year and 5-year OS time were conducted for calculating the area under the curve (AUC) values to assess the predictive efficiency of the gene signature and other clinical features (age, gender, grade and stage) by applying survivalROC package (Heagerty and Zheng, 2005). Furthermore, univariate cox analyses were presented for evaluating the relationships of OSCC prognosis with the gene signature and clinical features. To validate whether the gene signature could be independently predictive of patients' prognosis, multivariate cox analyses were performed based on prognosis-related factors with $p < 0.05$. Hazard ratio (HR) as well as 95% confidence interval (CI) was computed. Factors with HR > 1 were risk factors and those with HR < 1 were protective factors.

Independence of the EMT gene signature from other clinical features

For determining whether the gene signature was independent of other clinical features, age, gender, grade, and stage were separated into high- and low-risk subgroups in the discovery dataset. OSCC subjects were stratified into age >65 and <65 subgroups, female and male subgroups, grade I-II/III-IV subgroups, stage I-II/III-IV subgroups. Kaplan-Meier OS analysis was carried out in each subgroup and difference was evaluated by log-rank test.

Assessment of stromal score, immune score, and tumor purity

Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) package (https://sourceforge.net/projects/estimateproject) (Yoshihara et al., 2013) was employed for inferring stromal score, immune score, and tumor purity in each OSCC sample from the TCGA dataset based on gene expression profiles.

Characterization of immune cell compositions in OSCC tissues

By applying CIBERSORT package (https://cibersort.stanford.edu/) (Newman et al., 2015), the infiltration levels of different immune cells were inferred in OSCC tissue specimens from the TCGA dataset according to gene expression profiling. Only specimens with $p < 0.05$ could be retained for further analyses. LM22 leukocyte gene signature matrix was used as the reference, which contained 547 genes that distinguished 22 human hematopoietic cells as follows: 7 kinds of T cell types, naïve and memory B cells, plasma cells, NK cells as well as myeloid subsets.

Pathway enrichment analysis

Gene Set Enrichment Analysis (GSEA) software was utilized for exploring the activated signaling pathways through comparing high- and low-risk subgroups from the TCGA dataset (Subramanian et al., 2005). The KEGG gene set (c2. cp.kegg.v7.1. symbols) was used as reference. 1,000 gene-set permutations were carried out. The terms with normalized enrichment score $|NES| > 1.5$ and FDR <0.05 were chosen as distinct pathways activated in high- or low-risk phenotypes, which were used for multiple GSEA gene sets.

Somatic mutation analysis

Mutation Annotation Format (MAF) of OSCC samples was downloaded from TCGA database. These specimens were equally separated into high- and low-risk subgroups. The waterfall plots of two subgroups were depicted for illustrating the different mutation events using the Maftools package (Mayakonda et al., 2018).

Prognostic values of genes in the prognostic gene signature

Univariate cox analyses were applied to evaluate the prognostic value of each gene in the prognosis-related gene signature for OSCC patients from the TCGA dataset. Moreover, their expression was visualized in OSCC and normal tissue specimens. Spearman correlation between genes in this signature was evaluated in OSCC samples. The protein expression of genes from the prognostic gene signature in OSCC tissues was assessed through the Human Protein Atlas (https://www.proteinatlas.org/) (Colwill and Gräslund, 2011).

TABLE 1 Primer sequences for RT-qPCR.

Target genes	Primer sequences
AREG	5′-GTGGTGCTGTCGCTCTTGATA-3′ (F)
	5′-CCCCAGAAAATGGTTCACGCT-3′ (R)
COL5A3	5′-TGACCGGGCATTCAGAATTGG-3′ (F)
	5′-CGGGCACCCCTTTCATCAT-3′ (R)
DKK1	5′-CCTTGAACTCGGTTCTCAATTCC-3′ (F)
	5′-CAATGGTCTGGTACTTATTCCCG-3′ (R)
GAS1	5′-ATGCCGCACCGTCATTGAG-3′ (F)
	5′-TCATCGTAGTAGTCGTCCAGG-3′ (R)
GPX7	5′-CCCACCACTTTAACGTGCTC-3′ (F)
	5′-GGCAAAGCTCTCAATCTCCTT-3′ (R)
PLOD2	5′-CATGGACACAGGATAATGGCTG-3′ (F)
	5′-AGGGGTTGGTTGCTCAATAAAAA-3′ (R)
SFRP1	5′-ACGTGGGCTACAAGAAGATGG-3′ (F)
	5′-CAGCGACACGGGTAGATGG-3′ (R)
GAPDH	5′-CTGGGCTACACTGAGCACC-3′ (F)
	5′-AAGTGGTCGTTGAGGGCAATG-3′ (R)

OSCC tissue specimens

Totally, 40 paired OSCC and adjacent normal frozen tissues were collected from patients who experienced operation in the Taihe Hospital, Hubei University of Medicine between January 2020 and January 2021. All subjects did not receive chemoradiotherapy before operation. Diagnosis and staging were performed by experienced pathologists according to the American Joint Committee on Cancer Staging System. The study protocol gained the approval of the Ethics Committee of Taihe Hospital, Hubei University of Medicine (KY 2020-024), with written informed consent acquired from each subject. In addition, the research followed the Declaration of Helsinki.

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

Each gene in this prognostic model was verified in OSCC by RT-qPCR. Total RNA was isolated from OSCC tissues utilizing TRIzol reagent (Beyotime, China), which was reverse transcribed into cDNA utilizing primers and SuperScriptIII reverse transcriptase. RT-qPCR was carried out through Prime Script RT Reagent Kit and 7500 Real-Time PCR System (Applied Biosystems, United States). GAPDH served as the reference control. These amplification procedures included: denaturation at 95 °C lasting 5 min, 40 cycles of denaturation at 95 °C lasting 15 s, annealing at 55 °C lasting 30 s, and extension at 60 °C lasting 1 min. Table 1 listed the primer sequences of target genes. The relative expression was determined with $2^{-\Delta\Delta CT}$ method.

TABLE 2 Prognosis-related EMT genes in OSCC by univariate cox regression analysis.

Genes	HR	HR.95L	HR.95H	P
ANPEP	0.822134	0.681012	0.992499	0.041518
AREG	1.154353	1.03309	1.289849	0.011249
COL5A3	0.835201	0.710865	0.981286	0.028548
DKK1	1.161784	1.048932	1.286777	0.004024
FMOD	0.826339	0.715891	0.953827	0.009168
GAS1	0.782505	0.659839	0.927975	0.004814
GPX7	0.818143	0.693531	0.965146	0.017277
PLOD2	1.193821	1.007232	1.414976	0.041046
SFRP1	0.883559	0.799458	0.976508	0.015276
TNFRSF11B	0.638404	0.462455	0.881295	0.00637
VEGFA	1.221469	1.006439	1.482442	0.042874

Abbreviations: HR, hazard ratio; HR.95L, 95% lower confidence interval; HR.95H, 95% upper confidence interval.

Immunohistochemistry

From The Human Protein Atlas (https://www.proteinatlas.org/), immunohistochemistry staining of genes in the EMT gene signature in OSCC and normal oral tissues from 10 OSCC patients. Staining, intensity, quantity and location were also obtained. The used antibodies are as follows: AREG (HPA008720), COL5A3 (HPA048256), GAS1 (HPA066902), PLOD2 (CAB025898) and SFRP1 (CAB008116).

Statistical analysis

Statistical analysis was carried out by R packages (version 3.5.2) and GraphPad Prism software (version 8.0.1). Comparisons between two subgroups were presented via Student's t-test or Wilcoxon rank-sum test. Kaplan-Meier survival curves were conducted, and survival difference was analyzed through log-rank test. The predictive efficacy was estimated with ROC curves. Spearman's test was executed for correlation analysis. $p < 0.05$ was considered significant.

Results

Establishment of an EMT gene signature for predicting OSCC prognosis

To screen prognosis-related EMT genes in OSCC, univariate cox regression analysis was employed in the TCGA dataset. As a result, 11 genes including ANPEP, AREG, COL5A3, DKK1, FMOD, GAS1, GPX7, PLOD2, SFRP1, TNFRSF11B, VEGFA were significantly associated with OSCC patients' prognosis (Table 2). These prognostic genes were further assessed by LASSO analysis. Totally, 9 genes were screened for establishing the LASSO model (Figure 1A). The regression coefficient of each gene was calculated, as shown in Figure 1B. The risk score of each patient was then determined, as follows: AREG expression * 0.0471791751815675 + COL5A3 expression * (−0.0543433112313582) + DKK1 expression * 0.0619741505236927 + GAS1 expression * (−0.118881742831249) + GPX7 expression * (−0.142215922759831) + PLOD2 expression * 0.259497264609766 + SFRP1 expression * (−0.0579599175758466) + TNFRSF11B expression * (−0.197512366515929) + VEGFA expression * (0.053308700735239). We separated all patients in the discovery dataset into two subgroups according to the median value of risk score (Figure 1C). Survival status was further compared in the two subgroups. There were more patients with dead status for high-risk in comparison to low-risk subgroups (Figure 1D). Heat map depicted the different expressions of the genes (SFRP1, TNFRSF11B, PLOD2, GPX7, COL5A3, GAS1, VEGFA, AREG and DKK1) in this prognostic model between high- and low-risk subgroups (Figure 1E). Our further analysis demonstrated that high-risk patients exhibited worse survival time in comparison to low-risk subjects ($p = 6.615e-05$; Figure 1F). These data indicated that the risk score could be employed for predicting OSCC patients' clinical outcomes. We further assessed the predictive efficacy of the risk score for OSCC prognosis by ROCs. The AUCs under 1-year, 3-year and 5-year OS were separately 0.669, 0.715 and 0.622, confirming the well predictive performance for clinical outcomes (Figure 1G). Also, we compared the predictive value of the risk score with other clinical features. Our data demonstrated that the risk score displayed the highest AUC value (0.715) for OS time among age (AUC = 0.575), gender (AUC = 0.487), grade (AUC = 0.557) and stage (AUC = 0.625; Figure 1H), indicating that this signature was more advantageous in comparison to other clinical features regarding prediction of survival time.

External verification of prognostic potential of this EMT gene signature in OSCC

The GSE41613 cohort was employed to externally validate the predictive efficacy of this EMT gene signature in OSCC patients' prognosis. With the same formula, we calculated the risk scores of OSCC patients. Consistently, we separated OSCC subjects into two subgroups according to the median value of risk score (Figure 2A). Low-risk patients exhibited more optimistic survival status than high-risk individuals (Figure 2B). The expression of the genes in this model was visualized in each OSCC sample (Figure 2C). The survival difference between subgroups was further validated in the GSE41613 dataset. As

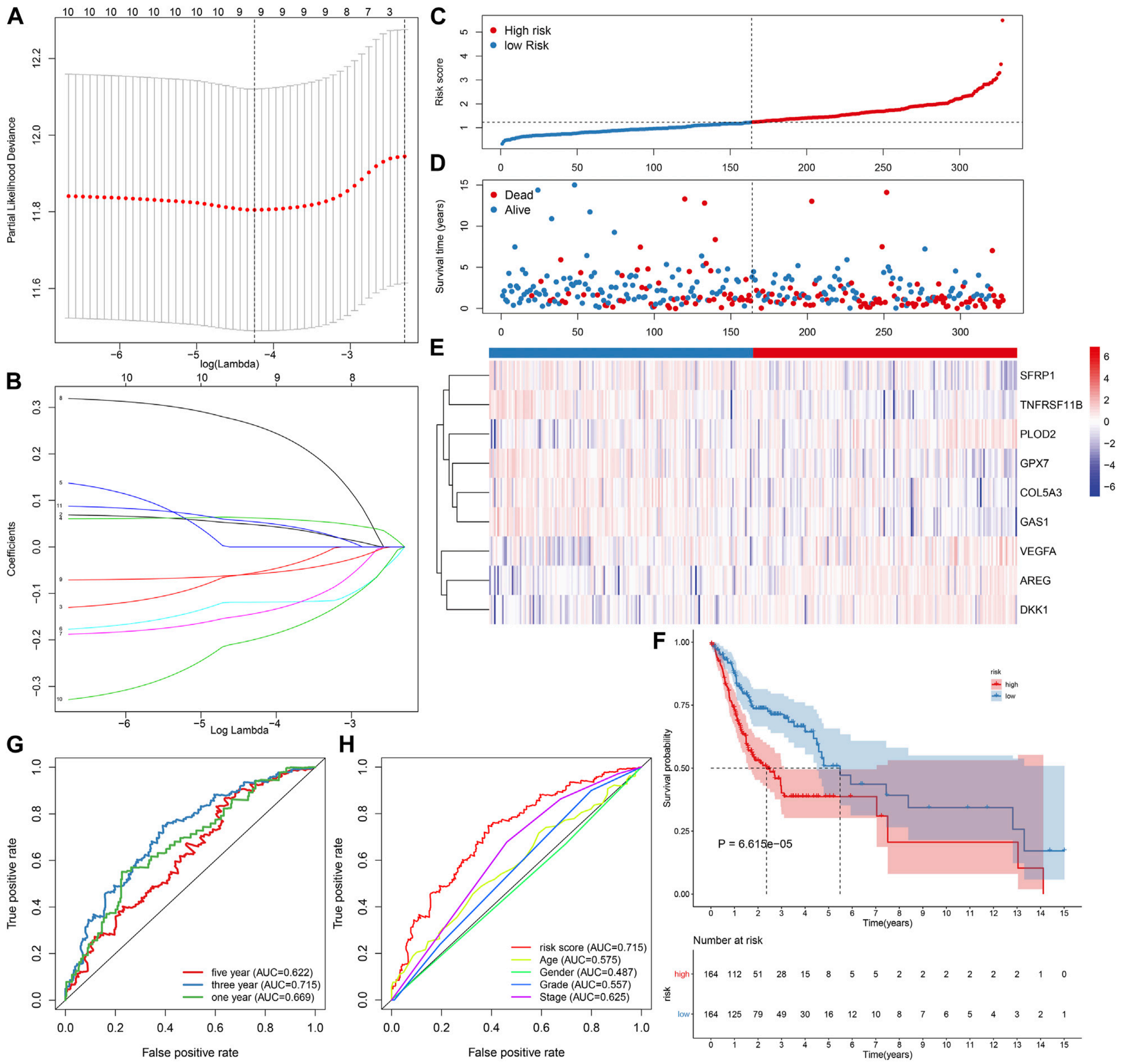


FIGURE 1
Establishment of an EMT gene signature for predicting OSCC patients' clinical outcomes in the TCGA-discovery dataset. **(A)** Distribution of partial likelihood deviances corresponding to lambda values. **(B)** Determination of regression coefficients of genes in the LASSO model. **(C)** Distribution of risk scores and determination of high-/low-risk subgroups. Vertical dotted line indicates the median value of risk score. **(D)** Distribution of survival status of high- and low-risk patients. Red dot indicates dead while blue dot is indicative of alive. **(E)** Heat map for expression pattern of these genes in this model in high- (red) and low-risk (blue) subgroups. Red indicates upregulation and blue indicates downregulation. **(F)** OS analysis for high-/low-risk patients. The difference between subgroups was compared with log-rank test. **(G)** ROCs under 1-year, 3-year and 5-year OS for the risk score. **(H)** Comparison of the AUCs among the risk score and other clinical features.

expected, high-risk patients displayed shorter OS time than low-risk patients ($p = 7.869e\text{-}04$; Figure 2D). ROCs were conducted for evaluation of the predictive efficacy of the risk score. In Figure 2E, AUCs under 1-, 3- and 5-year OS were separately 0.800, 0.778 and 0.729, confirming that this risk score could predict OS time of OSCC patients. Univariate analysis revealed that age, stage, risk score displayed distinct associations to OSCC patients' prognosis in the TCGA dataset (Figure 2F). As confirmed by multivariate analysis, age, stage as well as risk score were independently related to prognosis (Figure 2G).

Subgroup analysis identifies the sensitivity of the EMT gene signature to predict OSCC prognosis

The predictive potency of this EMT prognostic model was further evaluated among different subgroups from the discovery cohort. Kaplan-Meier OS analysis showed that high-risk patients were predicted to have worse clinical outcomes compared with low-risk patients in age >65 ($p = 0.012$; Figure 3A), age <65 ($p = 0.001$;

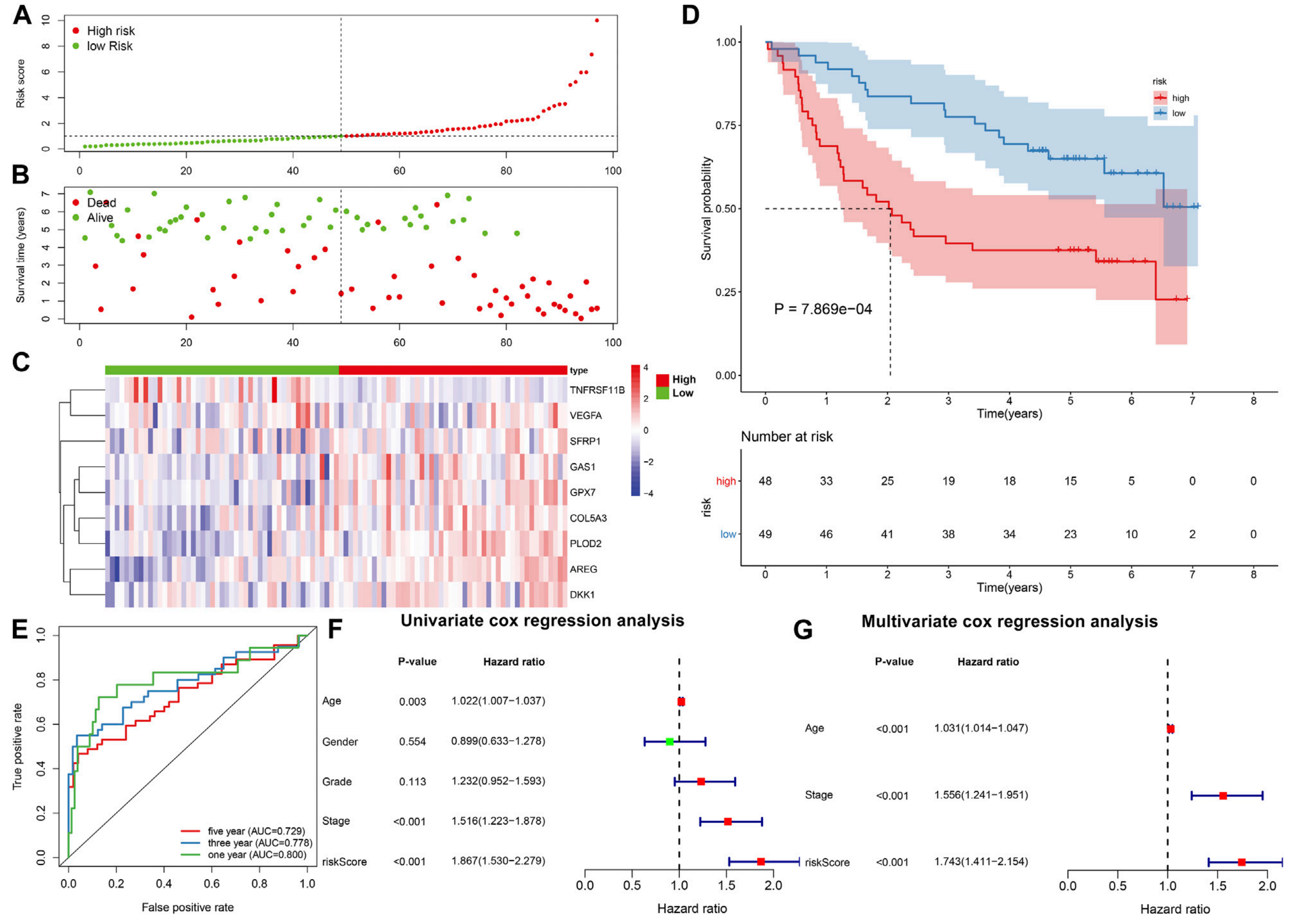


FIGURE 2
External verification of prognostic value of the EMT gene model in OSCC. **(A)** Distribution of risk scores and identification of high-/low-risk subgroups in the GSE41613 dataset. Vertical dotted line represents the median value of risk score. **(B)** Distribution of survival status in high-/low-risk patients. Red dot represents dead while blue dot represents alive. **(C)** Heat map for expression pattern of genes in this model in high- (red) and low-risk (blue) subgroups. Red indicates upregulation and blue indicates downregulation. **(D)** OS analysis for high-/low-risk patients. *p*-value was calculated with log-rank test. **(E)** ROCs under 1-year, 3-year and 5-year OS based on the risk score. **(F,G)** Univariate and multivariate analyses of the relationships of OSCC prognosis with risk score and other clinical features.

Figure 3B), female (p = 0.064; Figure 3C), male (p < 0.001; Figure 3D), grade I-II (p = 0.001; Figure 3E), grade III-IV (p = 0.010; Figure 3F), stage I-II (p = 0.112; Figure 3G) as well as stage III-IV (p = 0.002; Figure 3H) subgroups.

Association between the EMT gene signature and immune microenvironment in OSCC

ESTIMATE algorithm was employed to assess stromal score, immune score, and tumor purity of OSCC samples from the TCGA dataset. Our data showed that higher stromal (p < 0.001) and immune scores (p < 0.001) were found in high-risk samples than low-risk samples (Figure 4A). Also, there was lower tumor purity in high-risk samples compared with low-risk samples (p < 0.001). The infiltration levels of 22 kinds of immune cells of OSCC specimens were determined by applying CIBERSORT algorithm. There were lowered infiltration levels of naïve B cells (p < 0.001), T follicular helper cells (p < 0.01), Tregs (p < 0.001), T gamma delta cells (p < 0.05) and resting mast cells (p < 0.001) in high-risk specimens compared with low-risk specimens (Figure 4B). Meanwhile, higher infiltration levels of CD4 memory activated T cells (p < 0.05), resting NK cells (p < 0.05), activated dendritic cells (p < 0.05), activated mast cells (p < 0.001) and eosinophils (p < 0.01) were examined in high-risk compared with low-risk subgroups.

Assessment of the EMT gene signature-related signaling pathways and somatic mutation in OSCC

GSEA was applied to explore signaling pathways associated with the EMT gene signature. As a result, basal transcription factors (NES = 2.04, NOM p = 0.002 and FDR = 0.008), base excision repair (NES =

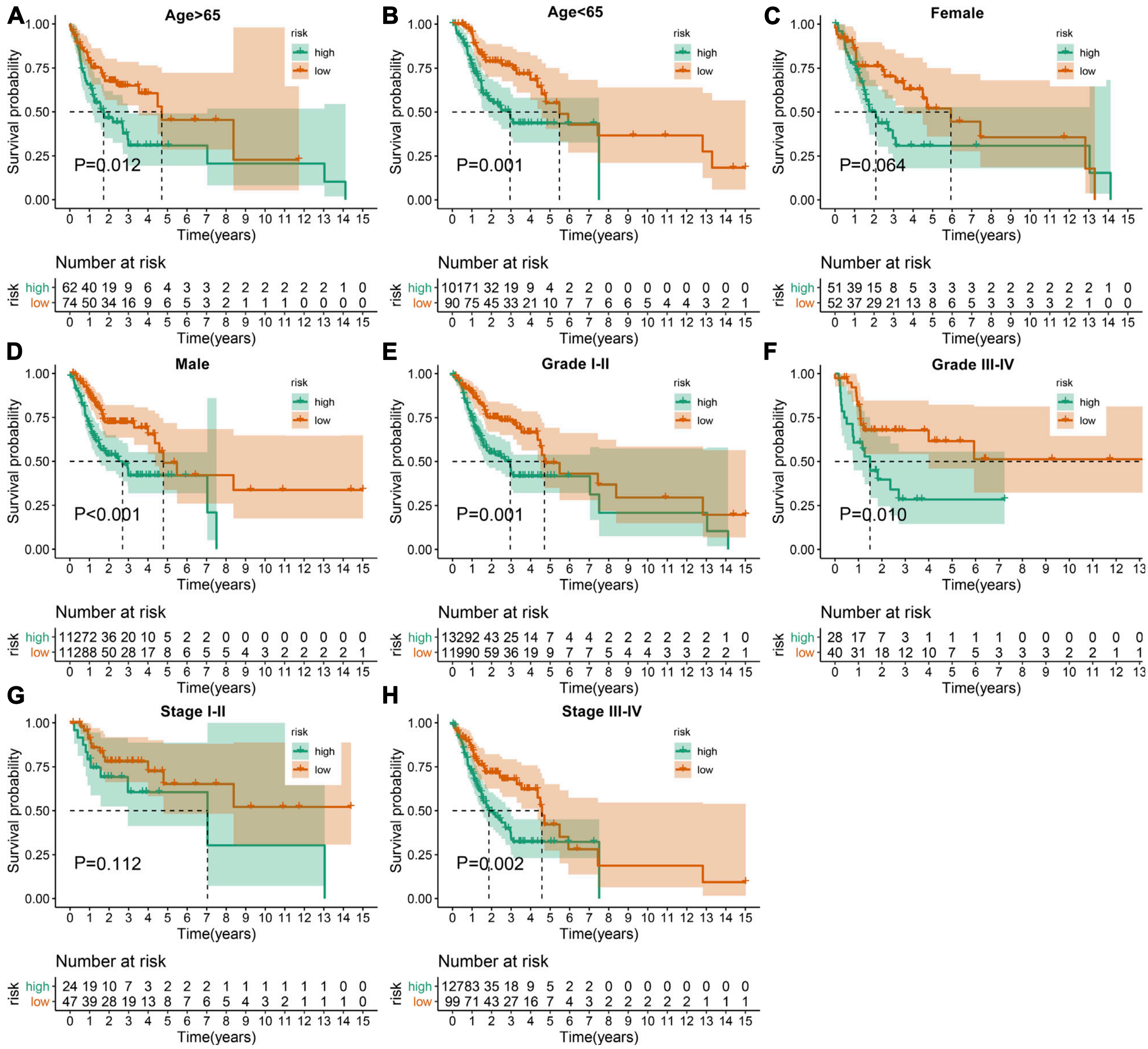


FIGURE 3
Assessment of the sensitivity of the EMT gene signature to predict OSCC prognosis by subgroup analysis. Kaplan-Meier OS analysis of high- and low-risk patients in **(A)** age >65, **(B)** age <65, **(C)** female, **(D)** male, **(E)** grade I-II, **(F)** grade III-IV, **(G)** stage I-II as well as **(H)** stage III-IV subgroups. *p* values were determined by log-rank test.

1.82, NOM $p = 0.006$ and FDR = 0.047), cell cycle (NES = 1.96, NOM $p = 0.006$ and FDR = 0.017), nucleotide excision repair (NES = 2.08, NOM $p = 0.002$ and FDR = 0.006) and spliceosome (NES = 2.13, NOM $p < 0.001$ and FDR = 0.004) were distinctly upregulated in high-risk OSCC samples (Figure 5A). Moreover, calcium signaling pathway (NES = −2.21, NOM $p < 0.001$ and FDR <0.001), cytokine-cytokine receptor interaction (NES = −1.99, NOM $p = 0.002$ and FDR = 0.008), ECM receptor interaction (NES = −1.98, NOM $p = 0.006$ and FDR = 0.009), MAPK signaling pathway (NES = −1.94, NOM $p < 0.001$ and FDR = 0.011) and VEGF signaling pathway (NES = −2.18, NOM $p < 0.001$ and FDR <0.001) were activated in low-risk samples (Figure 5B). The somatic mutation was further assessed in high- and low-risk OSCC samples. Our data showed the first 20 mutated genes across OSCC samples. We found that higher frequent genetic mutations occurred in high-risk subgroup (Figure 5C) than low-risk subgroup (Figure 5D), especially TP53, FAT1, and CDKN2A.

Genes in the EMT gene signature are associated with OSCC prognosis

Prognostic value of each gene in this EMT gene model was evaluated for OSCC samples from the TCGA dataset. Our univariate cox regression analysis demonstrated that high expression of AREG ($p < 0.001$, HR = 1.85, 95%CI: 1.31-2.61; Figure 6A), DKK1 ($p < 0.001$, HR = 2.05, 95%CI: 1.47-2.87; Figure 6B), PLOD2 ($p = 0.01$, HR = 1.58, 95%CI: 1.13-2.22; Figure 6C) and VEGFA ($p = 0.019$, HR = 1.51, 95%CI: 1.03-2.21;

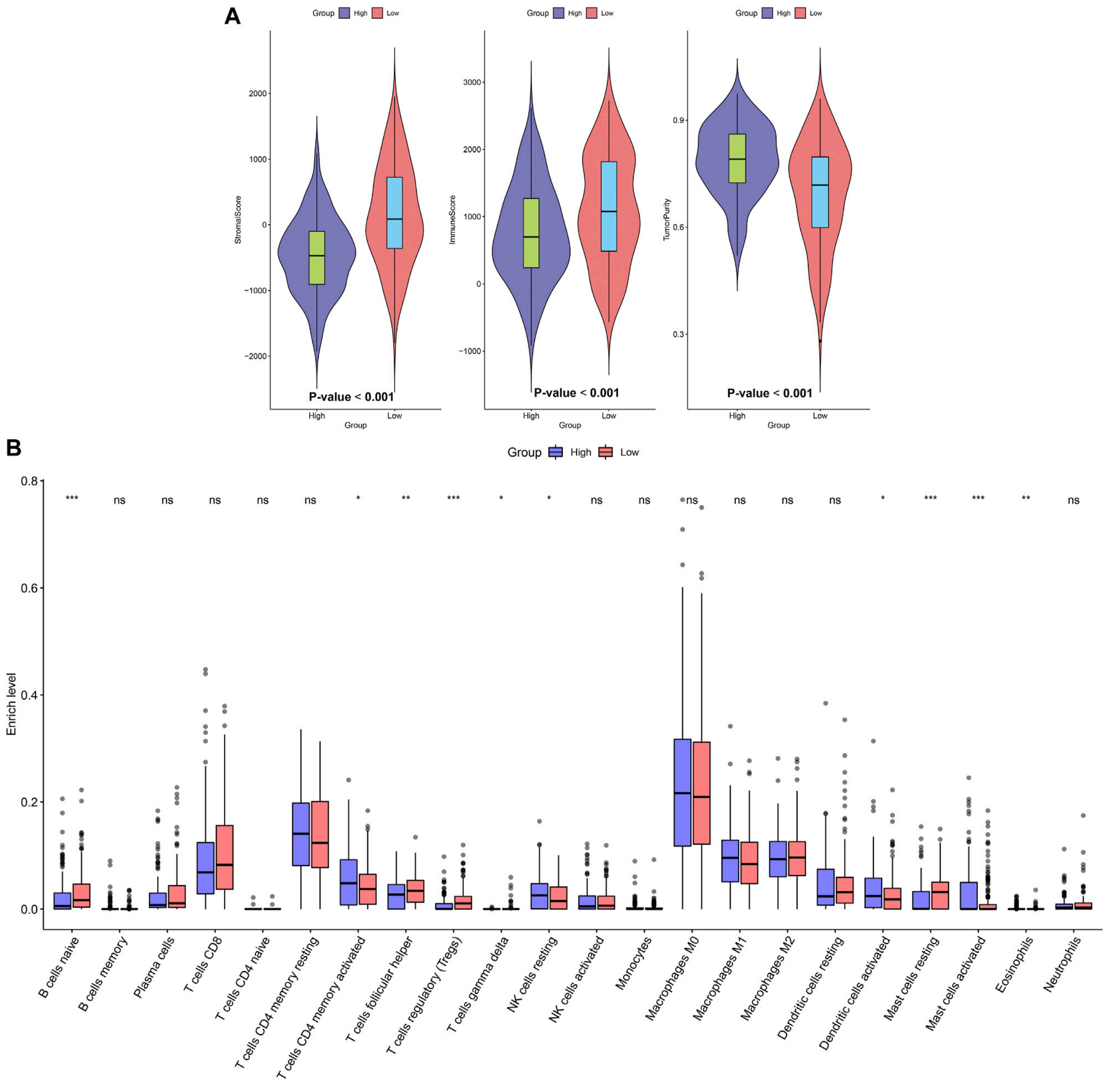


FIGURE 4
Association between the EMT gene signature and immune microenvironment in OSCC. **(A)** Distributions of stromal score, immune score, and tumor purity in high- and low-risk subgroups. **(B)** Assessment of the infiltration levels of immune cells in high- and low-risk subgroups. *p* values were assessed by Wilcoxon rank-sum test. Ns: not significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Figure 6D) was indicative of poorer prognosis of OSCC patients than their low expression. Furthermore, high expression of COL5A3 ($p = 0.036$, HR = 0.66, 95%CI: 0.46-0.94; Figure 6E), GAS1 ($p = 0.006$, HR = 0.61, 95%CI: 0.44-0.86; Figure 6F), GPX7 ($p = 0.004$, HR = 0.58, 95%CI: 0.41-0.82; Figure 6G), SFRP1 ($p = 0.002$, HR = 0.58, 95%CI: 0.42-0.82; Figure 6H) and TNFRSF11B ($p < 0.001$, HR = 0.55, 95%CI: 0.39-0.77; Figure 6I) displayed significant associations with prolonged survival time compared with their low expression. The prognostic implications of above genes were also confirmed in the GSE41613 cohort (Figures 7A–I).

Abnormal expression of genes in the EMT gene signature for OSCC

The expression of genes in the EMT gene signature was compared between OSCC and normal tissues. Our data showed that AREG (Figure 8A), COL5A3 (Figure 8B), DKK1 (Figure 8C), GAS1 (Figure 8D), GPX7 (Figure 8E) and PLOD2 (Figure 8F) were significantly upregulated in OSCC than normal tissues (all $p < 0.05$). Furthermore, lower SFRP1 expression was found in OSCC compared to normal specimens ($p < 0.05$;

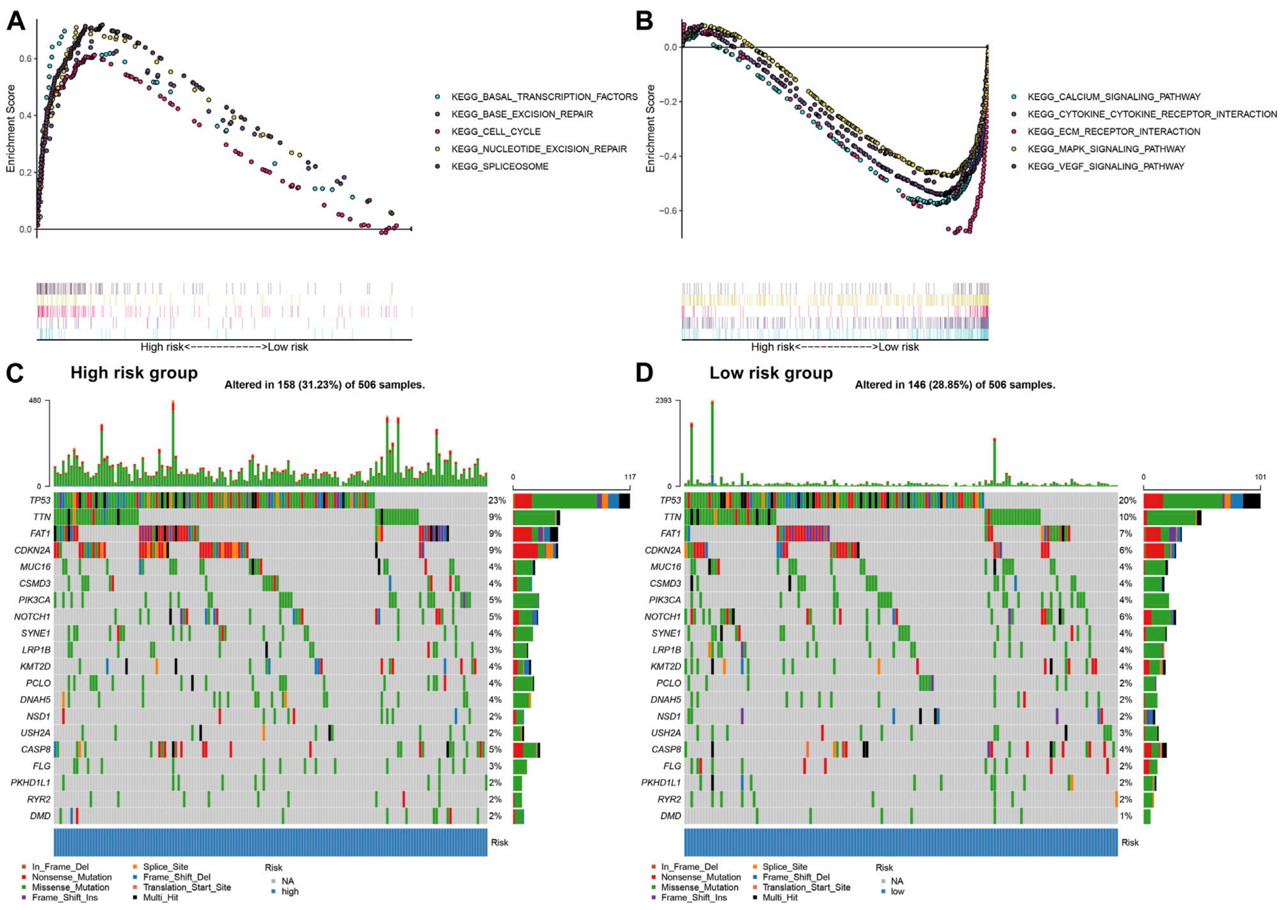


FIGURE 5
Assessment of the EMT gene signature-related signaling pathways and somatic mutation in OSCC. **(A)** Signaling pathways activated in high-risk subgroup by GSEA. **(B)** Signaling pathways activated in low-risk subgroup. **(C)** The waterfall plots for the first 20 mutated gene signatures in high-risk subgroup. **(D)** The waterfall plots for the first 20 mutated gene signatures in low-risk subgroup.

Figure 8G). Our correlation analyses demonstrated that COL5A3 exhibited significant correlations to GAS1, GPX7, PLOD2, SFRP1 and TNFRSF11B in OSCC samples (Figure 8H). GAS1 exhibited significant correlations to GPX7, PLOD2, SFRP1 and TNFRSF11B. GPX7 was distinctly associated with PLOD2 and TNFRSF11B. These data indicated that there were distinct correlations between the genes in the EMT gene signature.

Validation of gene expression in this EMT gene model

This study further confirmed gene expression in the EMT gene signature between 40 paired OSCC and normal specimens by RT-qPCR. Consistently, our data confirmed that AREG (Figure 9A), COL5A3 (Figure 9B), DKK1 (Figure 9C), GAS1 (Figure 9D), GPX7 (Figure 9E) and PLOD2 (Figure 9F) were distinctly highly expressed in OSCC compared with normal tissues (all $p < 0.0001$). Also, SFRP1 exhibited lower expression in OSCC than normal specimens ($p < 0.0001$; Figure 9G). Abnormal expression of AREG, COL5A3, GAS1, PLOD2 and SFRP1 was also confirmed in OSCC tissues by immunohistochemistry (Figure 9H). We also evaluated the difference in genes from the EMT gene signature across distinct pathological stages, as shown in Figures 10A–G. Among them, COL5A3, PLOD2 and SFRP1 were differentially expressed among pathological stages, indicative of their potential relationships with disease progression.

Discussion

OSCC represents a progressive malignancy with high heterogeneity (Panarese et al., 2019). Hence, it is of urgency to acquire robust prognostic markers for improving prognosis evaluation and individualized therapy (Zhu et al., 2020). As previous studies, several prognostic signatures have been established for OSCC. For instance, Cao et al. established a 3-mRNA signature (CLEC3B, C6 and CLCN1) in OSCC prognosis (Cao et al., 2019). Hou and colleagues developed an autophagy gene model for speculation of clinical outcomes of OSCC (Hou et al., 2020). Wu and colleagues established an independent

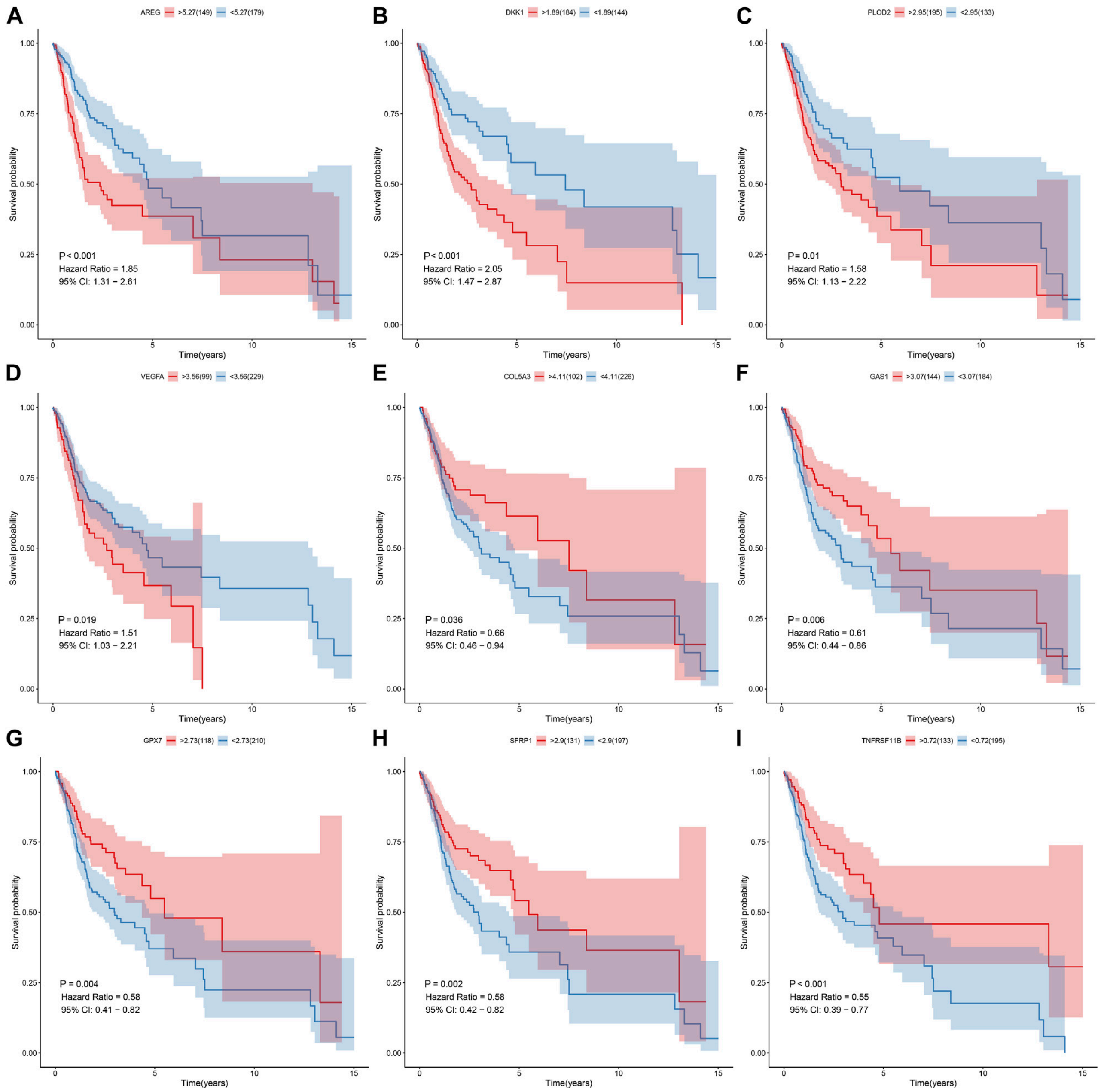


FIGURE 6
Univariate cox regression analysis for the association between each gene in the EMT gene signature and OSCC prognosis in the TCGA cohort. The survival difference was evaluated between high and low expression of **(A)** AREG; **(B)** DKK1 **(C)** PLOD2; **(D)** VEGFA **(E)** COL5A3; **(F)** GAS1 **(G)** GPX7; **(H)** SFRP1 **(I)** TNFRSF11B subgroups.

transcriptional model according to 5 metabolism pathways concerning OSCC prognosis (Wu et al., 2020). Huang et al. constructed a 7-metabolic gene signature for OSCC (Huang et al., 2021). However, the above gene signatures have not been validated in multiple datasets. Furthermore, so far, no gene signature has been applied in clinical practice. Although many molecular markers and gene signatures have been conducted in OSCC, systematic analyses of expression profiles of EMT genes have not been still performed. In this study, we conducted an EMT gene signature for OSCC prognosis by LASSO method. After external verification, our model robustly and stably predicted patient survival.

The tumor microenvironment contains tumor-associated fibroblasts, immune cells as well as extracellular matrix (Chen et al., 2021). The relationships between tumor microenvironment and tumor cells play key roles in modulating malignant biological behaviors like metastasis and recurrence as well as clinical outcomes of OSCC

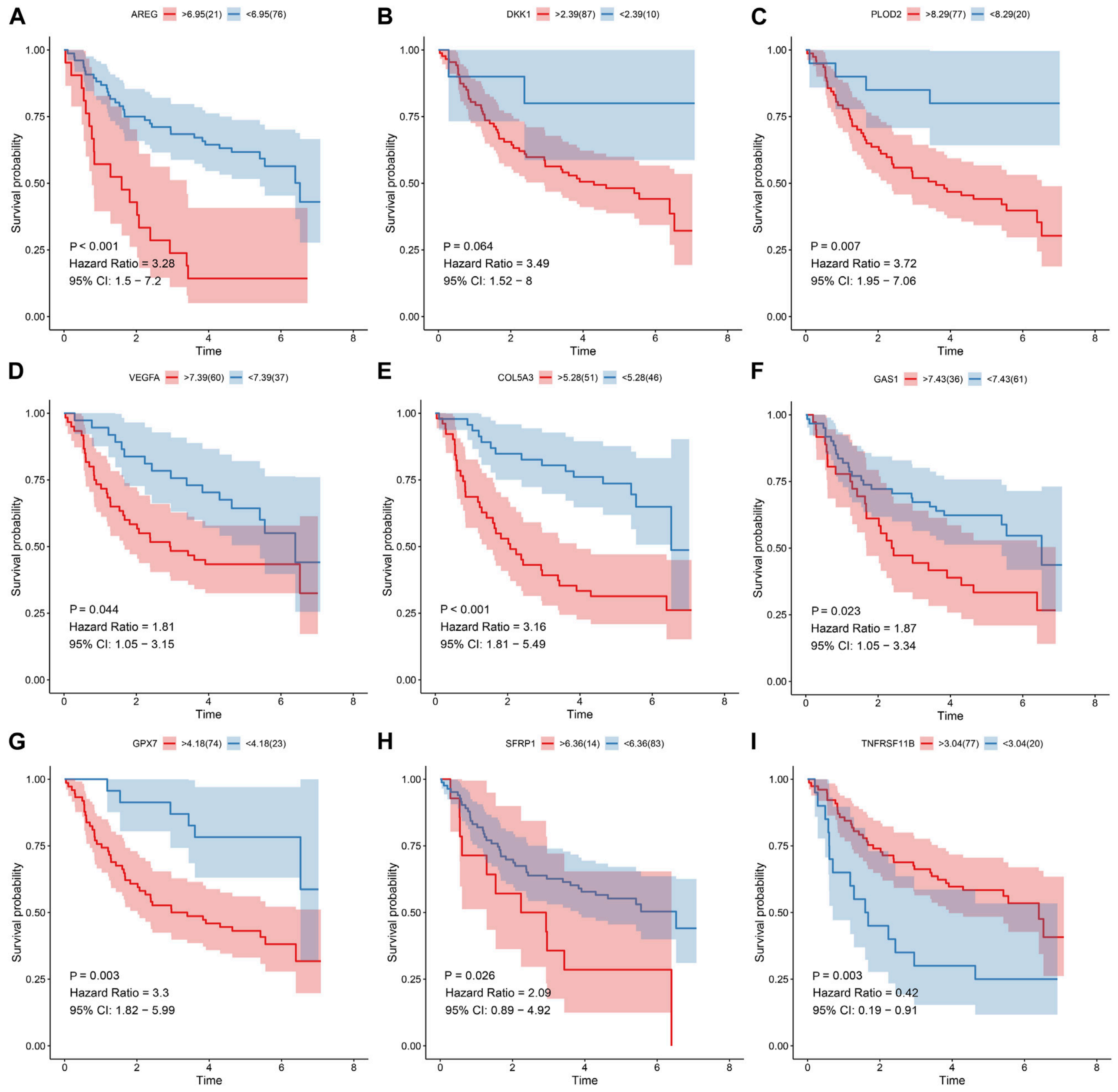


FIGURE 7
Univariate cox regression analysis for the association between each gene in the EMT gene signature and OSCC prognosis in the GSE41613 cohort. The survival difference was evaluated between high and low expression of **(A)** AREG; **(B)** DKK1 **(C)** PLOD2; **(D)** VEGFA **(E)** COL5A3; **(F)** GAS1 **(G)** GPX7; **(H)** SFRP1 **(I)** TNFRSF11B subgroups.

(Zhou et al., 2020). It has been found that OSCC is highly related to immune infiltration and immune infiltrates are reliable prognostic factors for OSCC (Zhou et al., 2020). For instance, high infiltration of CD103$^+$ T and dendritic cells is indicative of prolonged survival outcomes of OSCC (Xiao et al., 2019). Activation of myeloid derived suppressor cells accelerates the malignant progression of OSCC (Pang et al., 2020). Activation of T helper cells in sentinel node indicates unfavorable clinical outcomes in OSCC (Kågedal et al., 2020). Therefore, the variations of immune cell subpopulations in the tumor microenvironments may affect the prognosis of OSCC. Here, our data showed that higher immune or stromal scores were detected in high-than low-risk subgroups. Furthermore, there were lowered infiltration levels of naïve B cells, T follicular helper cells, Tregs, T gamma delta cells and resting mast cells in high-risk than low-risk subgroups. Also, higher infiltration levels of CD4 memory activated T cells, resting NK cells, activated dendritic cells, activated mast cells and eosinophils were examined in high-compared with low-risk subgroups. Thus, this EMT gene

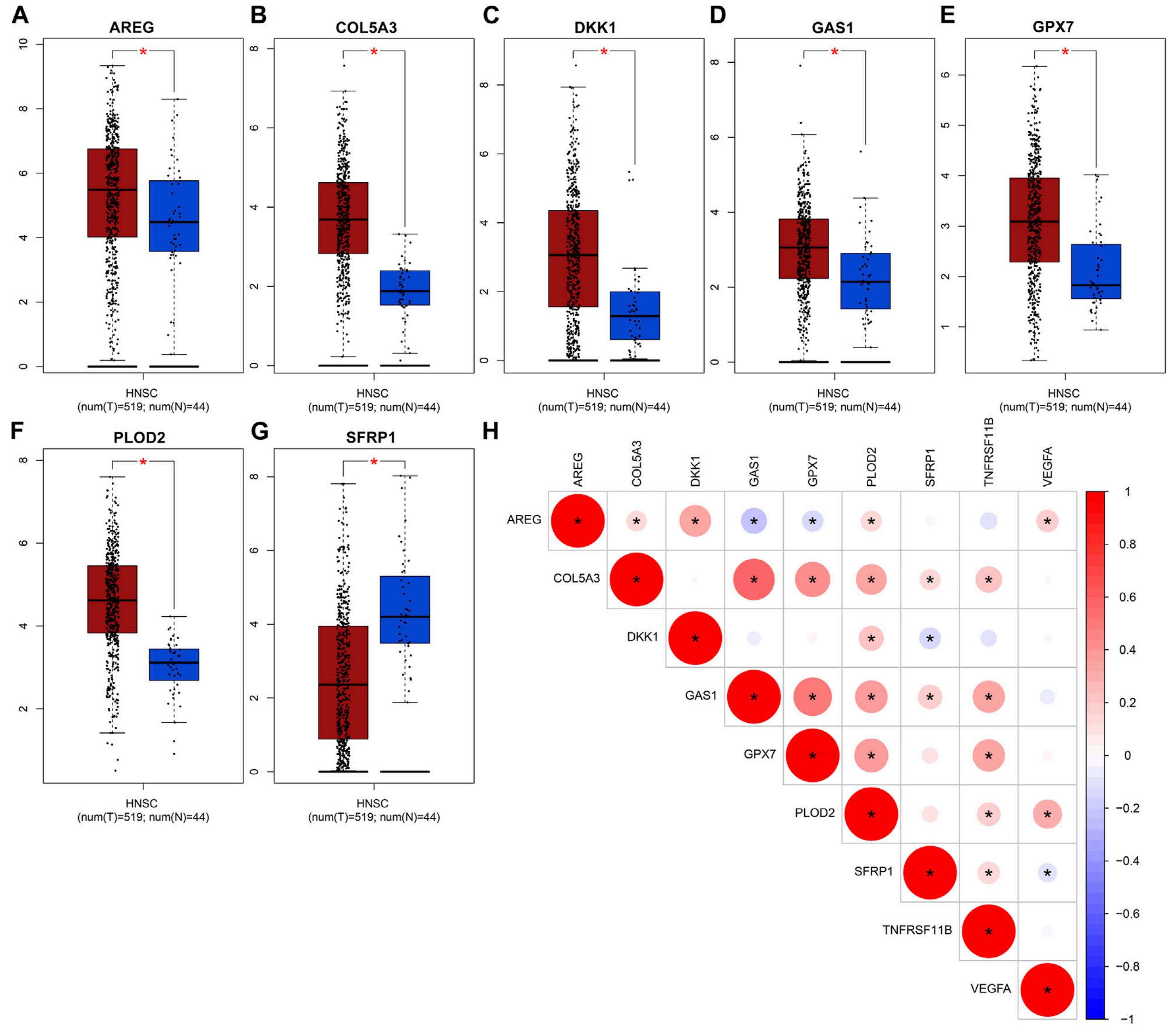


FIGURE 8
Abnormal expression of genes in the EMT gene signature for OSCC. Box plots for expression of **(A)** AREG, **(B)** COL5A3, **(C)** DKK1, **(D)** GAS1, **(E)** GPX7, **(F)** PLOD2 as well as **(G)** SFRP1 in OSCC and normal tissues. **(H)** Correlation analysis between the genes in the EMT gene signature. Red demonstrates positive correlation as well as blue demonstrates negative correlation. $^*p < 0.05$. The bigger the circle, the stronger the correlation.

signature might be distinctly linked to tumor microenvironment of OSCC.

Our further analysis found that basal transcription factors, base excision repair, cell cycle, nucleotide excision repair as well as spliceosome were activated in high-risk OSCC samples. Consistently, we found that more frequent somatic mutation occurred in high-risk OSCC samples. Calcium signaling pathway, cytokine-cytokine receptor interaction, ECM receptor interaction, MAPK signaling pathway and VEGF signaling pathway were activated in low-risk samples. Previously, calcium-dependent and cell cycle pathways may mediate OSCC progression (Jia et al., 2020). MAPK (Jin et al., 2020) and VEGF pathways (Lien et al., 2020) enhance OSCC progression. These data indicated that the genes in this signature might participate in OSCC pathogenesis by above pathways.

Among the genes in this prognostic signature, AREG, COL5A3, DKK1, GAS1, GPX7 and PLOD2 were distinctly upregulated and SFRP1 was downregulated in OSCC than normal tissues. High expression of AREG, DKK1, PLOD2 and VEGFA was indicative of poorer prognosis of OSCC patients while high expression of COL5A3, GAS1, GPX7, SFRP1 and TNFRSF11B were significantly associated with prolonged survival time. Previously, AREG upregulation has been found in OSCC and it can increase drug resistance against vincristine (Hsieh et al., 2019). Also, AREG expression can independently predict OSCC prognosis (Gao et al., 2016). DKK1 is highly expressed in OSCC and induces proliferation and migration of

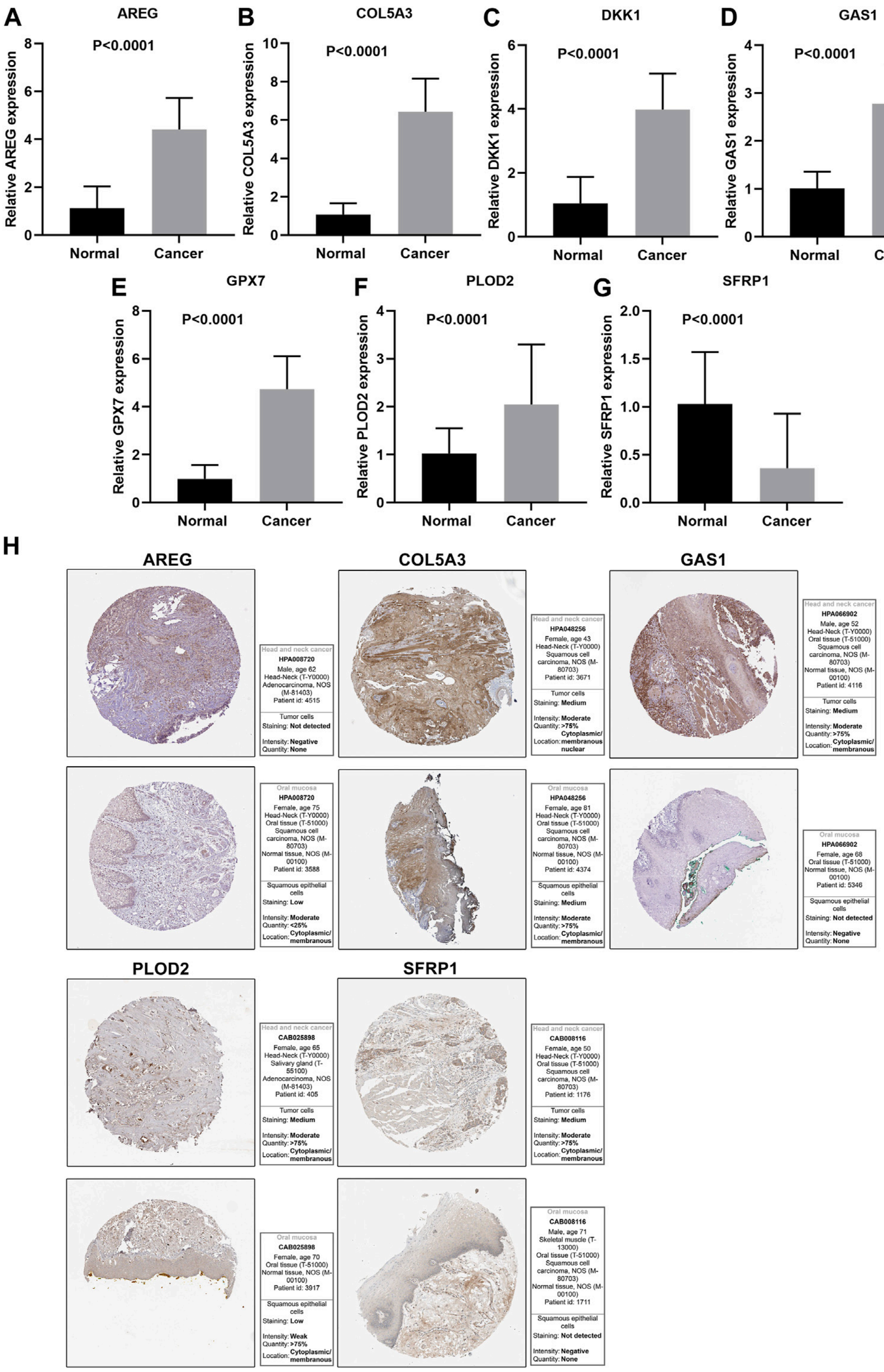


FIGURE 9
Validation of expression of genes in the EMT gene signature. RT-qPCR for detecting expressions of **(A)** AREG, **(B)** COL5A3, **(C)** DKK1, **(D)** GAS1, **(E)** GPX7, **(F)** PLOD2 and **(G)** SFRP1 in 40 paired OSCC and normal tissue specimens. **(H)** Immunohistochemistry for expression of AREG, COL5A3, GAS1, PLOD2 and SFRP1 in OSCC tissues.

OSCC cells (Wang et al., 2018). RT-qPCR confirmed the abnormal expression of the genes in OSCC. Combining previous research, the genes in this signature might be potential therapy targets against OSCC. More experiments will be presented for validating their biological functions and clinical implications in OSCC.

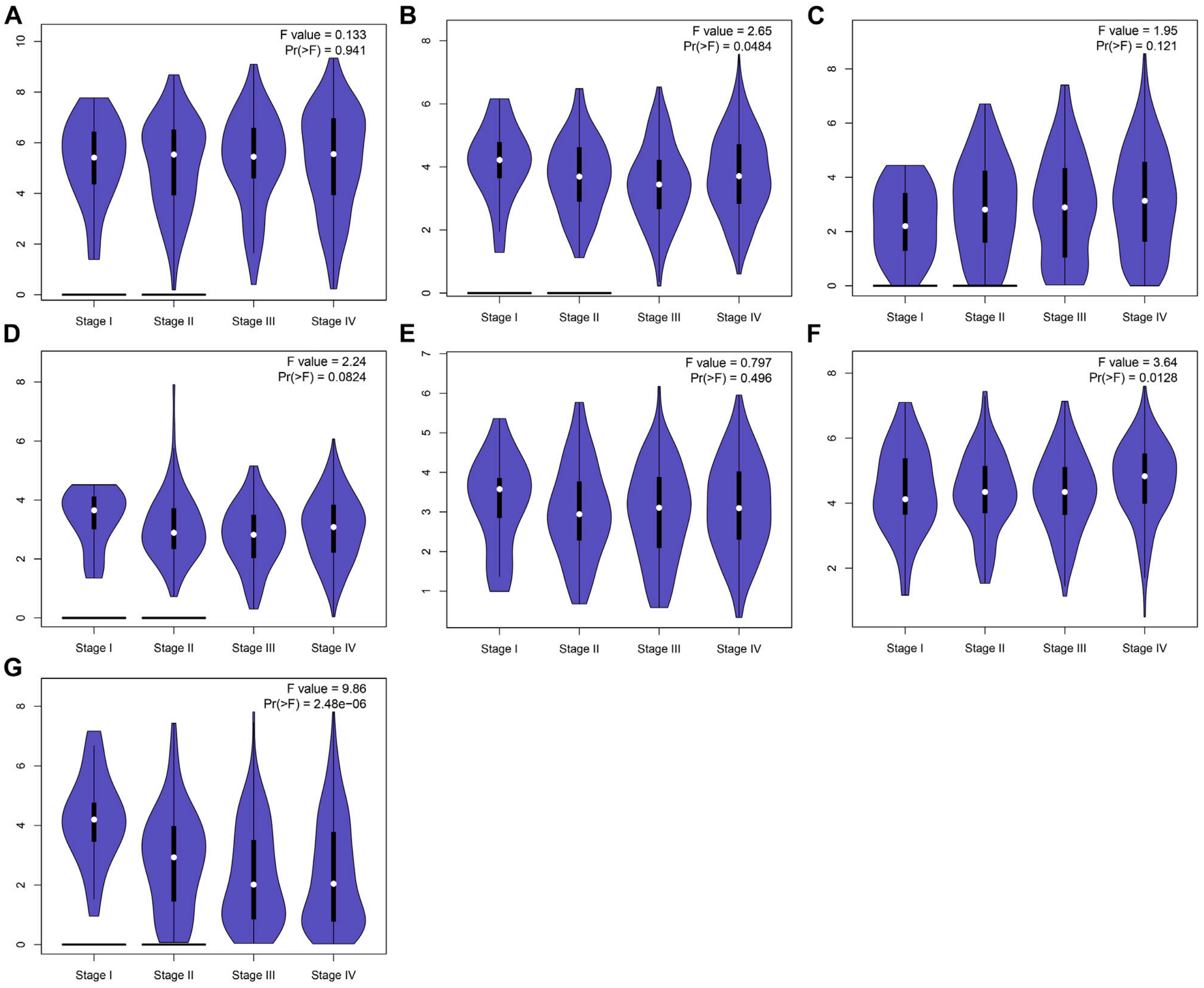


FIGURE 10
Expression of genes in the EMT gene signature across diverse pathological stages. **(A)** AREG, **(B)** COL5A3, **(C)** DKK1, **(D)** GAS1, **(E)** GPX7, **(F)** PLOD2 and **(G)** SFRP1.

Conclusion

Collectively, based on gene expression profiling, we screened prognosis-related EMT genes and established a 9-EMT gene signature. These data showed that this signature could be utilized to predict clinical outcomes of OSCC subjects, thereby contributing to individual therapy and shedding a novel insight into EMT targeted therapy. Nevertheless, the clinical utility of this signature requires to be verified in a large and prospective OSCC cohort.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Taihe Hospital, Hubei University of Medicine (KY2020-024). The patients/participants provided their written informed consent to participate in this study.

Author contributions

QF conceived and designed the study. JA, YT, and YS conducted most of the experiments and data analysis, and wrote the manuscript. BL and YW participated in collecting data and helped to draft the manuscript. XX participated in the writing of the first draft and proofread the final draft. All authors contributed to the article and approved the submitted version.

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.2023.1113137/full#supplementary-material

References

Almangush, A., Heikkinen, I., Mäkitie, A. A., Coletta, R. D., Läärä, E., Leivo, I., et al. (2017). Prognostic biomarkers for oral tongue squamous cell carcinoma: A systematic review and meta-analysis. *Br. J. Cancer* 117 (6), 856–866. doi:10.1038/bjc.2017.244

Bai, Y., Sha, J., and Kanno, T. (2020). The role of carcinogenesis-related biomarkers in the wnt pathway and their effects on epithelial-mesenchymal transition (EMT) in oral squamous cell carcinoma. *Cancers (Basel)* 12 (3), 555. doi:10.3390/cancers12030555

Cao, R., Wu, Q., Li, Q., Yao, M., and Zhou, H. (2019). A 3-mRNA-based prognostic signature of survival in oral squamous cell carcinoma. *PeerJ* 7, e7360. doi:10.7717/peerj.7360

Chen, L., Niu, X., Qiao, X., Liu, S., Ma, H., Shi, X., et al. (2021). Characterization of interplay between autophagy and ferroptosis and their synergistical roles on manipulating immunological tumor microenvironment in squamous cell carcinomas. *Front. Immunol.* 12, 739039. doi:10.3389/fimmu.2021.739039

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

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Gao, J., Ulekleiv, C. H., and Halstensen, T. S. (2016). Epidermal growth factor (EGF) receptor-ligand based molecular staging predicts prognosis in head and neck squamous cell carcinoma partly due to deregulated EGF- induced amphiregulin expression. *J. Exp. Clin. Cancer Res.* 35 (1), 151. doi:10.1186/s13046-016-0422-z

Heagerty, P. J., and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* 61 (1), 92–105. doi:10.1111/j.0006-341X.2005.030814.x

Hou, C., Cai, H., Zhu, Y., Huang, S., Song, F., and Hou, J. (2020). Development and validation of autophagy-related gene signature and nomogram for predicting survival in oral squamous cell carcinoma. *Front. Oncol.* 10, 558596. doi:10.3389/fonc.2020.558596

Hsieh, M. J., Chen, Y. H., Lee, I. N., Huang, C., Ku, Y. J., and Chen, J. C. (2019). Secreted amphiregulin promotes vincristine resistance in oral squamous cell carcinoma. *Int. J. Oncol.* 55 (4), 949–959. doi:10.3892/ijo.2019.4866

Huang, Z. D., Yao, Y. Y., Chen, T. Y., Zhao, Y. F., Zhang, C., and Niu, Y. M. (2021). Construction of prognostic risk prediction model of oral squamous cell carcinoma based on nine survival-associated metabolic genes. *Front. Physiol.* 12, 609770. doi:10.3389/fphys.2021.609770

Jia, L., Wang, T., Ding, G., Kuai, X., Wang, X., Wang, B., et al. (2020). Trop2 inhibition of P16 expression and the cell cycle promotes intracellular calcium release in OSCC. *Int. J. Biol. Macromol.* 164, 2409–2417. doi:10.1016/j.ijbiomac.2020.07.234

Jin, T., Guo, Y., Huang, Z., Zhang, Q., Huang, Z., Zhang, Y., et al. (2020). Vitamin D inhibits the proliferation of Oral Squamous Cell Carcinoma by suppressing lncRNA LUCAT1 through the MAPK pathway. *J. Cancer* 11 (20), 5971–5981. doi:10.7150/jca.45389

Ju, W. T., Ma, H. L., Zhao, T. C., Liang, S. Y., Zhu, D. W., Wang, L. Z., et al. (2020). Stathmin guides personalized therapy in oral squamous cell carcinoma. *Cancer Sci.* 111 (4), 1303–1313. doi:10.1111/cas.14323

Kågedal, Å., Hjalmarsson, E., Farrajota Neves da Silva, P., Piersiala, K., Georén, S. K., Margolin, G., et al. (2020). Activation of T helper cells in sentinel node predicts poor prognosis in oral squamous cell carcinoma. *Sci. Rep.* 10 (1), 22352. doi:10.1038/s41598-020-79273-3

Lien, M. Y., Chang, A. C., Tsai, H. C., Tsai, M. H., Hua, C. H., Cheng, S. P., et al. (2020). Monocyte chemoattractant protein 1 promotes VEGF-A expression in OSCC by activating ILK and MEK1/2 signaling and downregulating miR-29c. *Front. Oncol.* 10, 592415. doi:10.3389/fonc.2020.592415

Ling, Z., Cheng, B., and Tao, X. (2021). Epithelial-to-mesenchymal transition in oral squamous cell carcinoma: Challenges and opportunities. *Int. J. Cancer* 148 (7), 1548–1561. doi:10.1002/ijc.33352

Liu, L., Wu, Y., Li, Q., Liang, J., He, Q., Zhao, L., et al. (2020). METTL3 promotes tumorigenesis and metastasis through BMI1 m(6)A methylation in oral squamous cell carcinoma. *Mol. Ther.* 28 (10), 2177–2190. doi:10.1016/j.ymthe.2020.06.024

Lohavanichbutr, P., Méndez, E., Holsinger, F. C., Rue, T. C., Zhang, Y., Houck, J., et al. (2013). A 13-gene signature prognostic of HPV-negative OSCC: Discovery and external validation. *Clin. Cancer Res.* 19 (5), 1197–1203. doi:10.1158/1078-0432.Ccr-12-2647

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

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

Omori, H., Nishio, M., Masuda, M., Miyachi, Y., Ueda, F., Nakano, T., et al. (2020). YAP1 is a potent driver of the onset and progression of oral squamous cell carcinoma. *Sci. Adv.* 6 (12), eaay3324. doi:10.1126/sciadv.aay3324

Panarese, I., Aquino, G., Ronchi, A., Longo, F., Montella, M., Cozzolino, I., et al. (2019). Oral and oropharyngeal squamous cell carcinoma: Prognostic and predictive parameters in the etiopathogenetic route. *Expert Rev. Anticancer Ther.* 19 (2), 105–119. doi:10.1080/14737140.2019.1561288

Pang, X., Fan, H. Y., Tang, Y. L., Wang, S. S., Cao, M. X., Wang, H. F., et al. (2020). Myeloid derived suppressor cells contribute to the malignant progression of oral squamous cell carcinoma. *PLoS One* 15 (2), e0229089. doi:10.1371/journal.pone.0229089

Park, J., Zhang, X., Lee, S. K., Song, N. Y., Son, S. H., Kim, K. R., et al. (2019). CCL28-induced RARβ expression inhibits oral squamous cell carcinoma bone invasion. *J. Clin. Investig.* 129 (12), 5381–5399. doi:10.1172/jci125336

Peng, Q. S., Cheng, Y. N., Zhang, W. B., Fan, H., Mao, Q. H., and Xu, P. (2020). circRNA_0000140 suppresses oral squamous cell carcinoma growth and metastasis by targeting miR-31 to inhibit Hippo signaling pathway. *Cell. Death Dis.* 11 (2), 112. doi:10.1038/s41419-020-2273-y

Qiao, X., Niu, X., Shi, J., Chen, L., Wang, X., Liu, J., et al. (2020). Wnt5a regulates ameloblastoma cell migration by modulating mitochondrial and cytoskeletal dynamics. *J. Cancer* 11 (18), 5490–5502. doi:10.7150/jca.46547

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

Wang, Z., Wang, J., Chen, Z., Wang, K., and Shi, L. (2018). MicroRNA-1-3p inhibits the proliferation and migration of oral squamous cell carcinoma cells by targeting DKK1. *Biochem. Cell. Biol.* 96 (3), 355–364. doi:10.1139/bcb-2017-0015

Wu, H. T., Chen, W. T., Li, G. W., Shen, J. X., Ye, Q. Q., Zhang, M. L., et al. (2019). Analysis of the differentially expressed genes induced by cisplatin resistance in oral squamous cell carcinomas and their interaction. *Front. Genet.* 10, 1328. doi:10.3389/fgene.2019.01328

Wu, X., Yao, Y., Li, Z., Ge, H., Wang, D., and Wang, Y. (2020). Identification of a transcriptional prognostic signature from five metabolic pathways in oral squamous cell carcinoma. *Front. Oncol.* 10, 572919. doi:10.3389/fonc.2020.572919

Xiao, Y., Li, H., Mao, L., Yang, Q. C., Fu, L. Q., Wu, C. C., et al. (2019). CD103(+) T and dendritic cells indicate a favorable prognosis in oral cancer. *J. Dent. Res.* 98 (13), 1480–1487. doi:10.1177/0022034519882618

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

Zhao, W., Cui, Y., Liu, L., Qi, X., Liu, J., Ma, S., et al. (2020). Splicing factor derived circular RNA circUHRF1 accelerates oral squamous cell carcinoma tumorigenesis via feedback loop. *Cell. Death Differ.* 27 (3), 919–933. doi:10.1038/s41418-019-0423-5

Zhou, C., Diao, P., Wu, Y., Wei, Z., Jiang, L., Zhang, W., et al. (2020). Development and validation of a seven-immune-feature-based prognostic score for oral squamous cell carcinoma after curative resection. *Int. J. Cancer* 146 (4), 1152–1163. doi:10.1002/ijc.32571

Zhu, X., Li, T., Niu, X., Chen, L., and Ge, C. (2020). Identification of UBE2T as an independent prognostic biomarker for gallbladder cancer. *Oncol. Lett.* 20 (4), 44. doi:10.3892/ol.2020.11903

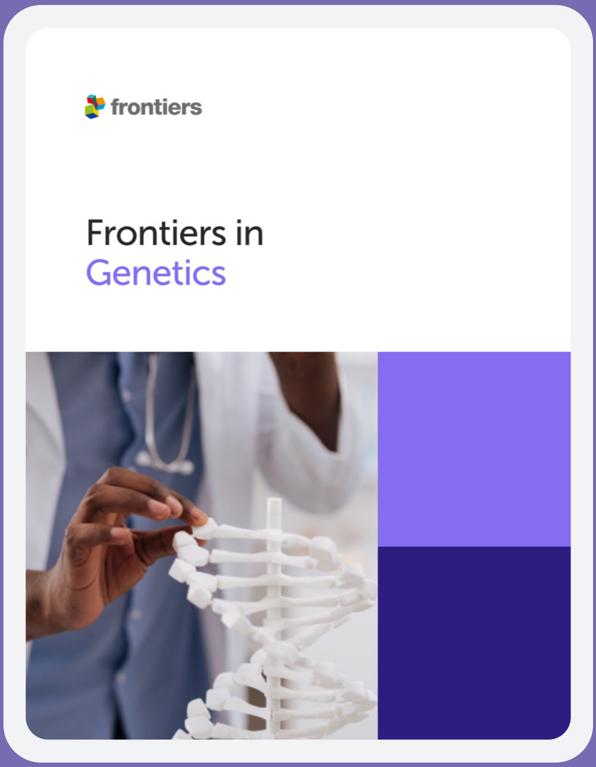
Frontiers in
Genetics

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

Frontiers in
Genetics

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact