Am J Cancer Res 2022;12(7):3051-3066
www.ajcr.us /ISSN:2156-6976/ajcr0141503

Original Article

Constructing novel molecular subtypes and an 11-gene signature based on pyroptosis signaling for lung adenocarcinoma

Lu Li[1], Qing He[1], Zhenchao Tao[1], Rui Zhang[2], Yayun Cui[1], Liting Qian[1]

[1]*The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230031, Anhui, China;* [2]*Department of Oncology, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei 230601, Anhui, China*

Received January 4, 2022; Accepted April 5, 2022; Epub July 15, 2022; Published July 30, 2022

Abstract: Pyroptosis plays important roles in various cancers. In this study, we focused on lung adenocarcinoma (LUAD) and aimed to develop new molecular subtypes based on pyroptosis signaling. Pyroptosis-related genes were used as a basis to classify molecular subtypes through unsupervised consensus clustering. Gene set enrichment analysis was performed to characterize tumor microenvironment (TME) and functional pathways. Univariate Cox regression and least absolute shrinkage and selection operator (LASSO) analysis were conducted to identify prognostic genes for establishing a prognostic model. Three molecular subtypes were established with distinct overall survival, TME and enriched pathways. C3 subtype had the longest survival and the highest immune infiltration. 11 prognostic genes were screened to build a prognostic signature for predicting LUAD prognosis. This study emphasized the important role of pyroptosis in LUAD development. Pyroptosis was considered to play critical roles in regulating TME. Moreover, the 11-gene signature could serve as an indicator for predicting LUAD prognosis, and was potential targets for developing targeted drugs.

Keywords: Lung adenocarcinoma, pyroptosis, molecular subtypes, tumor microenvironment, prognostic genes

Introduction

Lung adenocarcinoma (LUAD) is the most common histological type in non-small cell lung cancer (NSCLC), contributing to about 40% of lung cancer [1]. Lung cancer is a frequently diagnosed cancer worldwide, with 2,206,771 newly reported cases in 2020 [2]. According to global cancer statistics, 1,796,144 lung cancer deaths occurred, consisting of 18% of all cancer deaths in 2020 [2]. Over half of lung cancer patients would die within one year, mostly due the fact that patients were already at the advanced stage when diagnosed, and the 5-year overall survival (OS) is lower than 18% [1]. Although therapeutic drugs have been developed, drug resistance will reduce treatment efficiency. Numerous studies have illustrated that mutated oncogenes and dysregulated signaling pathways lead to tumor cell progression, which encourages the studies on novel therapeutic drugs for lung cancer [3, 4].

Programmed cell death is a physiological process of cell proliferation for maintaining homeostasis, and it is also considered as an antitumor mechanism. Tumor cells have the ability to evade cell death, which impairs the homeostasis of cell proliferation and cell death, and leads to malignant transformation. Therefore, targeting cell death pathway is an effective strategy against tumor cell growth [5]. Pyroptosis is one of the mechanisms leading to cell death, and will result in the release of intracellular proinflammatory contents [6]. Evidence supports that the cleavage of gasdermin D (GSDMD) activates caspase-1/4/5/11, thereby causing pyroptosis [7-9]. In oesophageal squamous cell carcinoma, GSDMD overexpression can switch apoptosis into pyroptosis [10]. Inhibition of GSDMD expression delays pyroptosis and accelerates tumor cell proliferation in gastric cancer through promoting the transition from S to G2 phase [11]. Therefore, pyroptosis is a potential therapeutic target for inhibiting

tumor cell growth through promoting pyroptosis. In NSCLC, Wang et al. discovered that simvastatin may function against tumor cells through inducing pyroptosis [12]. Pseudoalteromonas haloplanktis TAC125 has been demonstrated to have antiproliferative activity by inducing pyroptosis in LUAD cells [13].

To further understand the role of pyroptosis in LUAD, pyroptosis-related genes were selected for classifying molecular subtypes. We explored the association between tumor microenvironment (TME) and molecular subtypes, and discovered differentially enriched pathways among the subtypes. Furthermore, based on differentially expressed genes (DEGs) among the subtypes, we established a prognostic signature that could serve as an indicator to predict prognosis for LUAD patients. In addition, these prognostic genes may be new drug targets in LUAD treatment.

Materials and methods

Data source

TCGA-LUAD dataset containing gene expression profiles, single nucleotide variation (SNV) and copy number variation (CNV) data were obtained from The Cancer Genome Atlas (TCGA) database. After removing samples without survival time or survival status, a total of 472 samples (413 tumor samples and 59 normal samples) containing expression data remained. GSE31210 and GSE50081 datasets were obtained from Gene Expression Omnibus (GEO) database. Probes were converted to gene symbol. One probe corresponding to multiple genes was removed, whereas the median value was selected when multiple probes corresponding to one gene. Finally, 226 samples and 21,655 genes from GSE31210 dataset remained, and 181 samples and 21,655 genes from GSE50081 dataset remained. Pyroptosis-related genes in pyroptosis pathway were obtained from Molecular Signatures Database (MSigDB, v7.4) [14].

Assessment of gene alternations on pyroptosis-related genes

To delineate CNVs of 27 pyroptosis-related genes in TCGA-LUAD dataset, copy numbers were classified into three groups by segment_mean value, including amplification (Segment_Mean > 0.2), diploid (-0.2 < Segment_Mean < 0.2), deletion (Segment_Mean < -0.2). Expression of 27 pyroptosis-related genes corresponding to different CNVs was analyzed. For analyzing SNVs of 27 genes, SNV data in TCGA-LUAD dataset were included and mutect2 algorithm was applied to detect gene mutations [15].

Unsupervised consensus clustering

To classify molecular subtypes based on pyroptosis-related genes, expression profiles of 27 genes were normalized through scale function (center = T and scale = F). ConsensusClusterPlus R package was employed to conduct unsupervised consensus clustering [16] in TCGA-LUAD dataset. Parameters of clusterAlg = "pam", distance = "maximum" and pItem = 0.8 were set, and cumulative distribution function (CDF) and CDF delta area were used to determine the optimal cluster number k (2 to 10). The effectiveness of molecular subtyping was verified by GSE31210 and GSE50081 datasets.

Analysis on immune infiltration

CIBERSORT was introduced for assessing immune cell distribution in complex tissues from their gene expression profiles [17]. For characterizing the degree of immune infiltration and stromal infiltration, Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) was applied to calculate immune score and stromal score [18]. CIBERSORT and ESTIMATE both rely on gene set enrichment analysis (GSEA) according to a series of gene signatures [19], and have been widely used to characterize TME for tumor tissues.

Gene set enrichment analysis

GSEA is a powerful analytic tool for interpreting gene expression profiles based on gene sets with different biological function [19]. It has been popularly applied for identify meaningful biological pathways. Single-sample gene set enrichment analysis (ssGSEA) is a method based on GSEA but uses a novel algorithm for calculating enrichment score for each sample [20]. We conducted ssGSEA in GSVA R package to calculate enrichment score of "c2.cp.kegg.v7.0.symbols.gmt" pathways for each

sample in TCGA-LUAD dataset grouping by molecular subtypes [21].

Construction and validation of a prognostic model

Limma R package was employed to identify differentially expressed genes (DEGs) between two subtypes [22]. DEGs were screened with conditions of false discovery rate (FDR) < 0.05 and |log2(fold change)| > 1. Univariate Cox regression analysis was conducted to identify DEGs that were significantly associated with overall survival in TCGA-LUAD dataset, with DEGs with P < 0.01 being considered as prognostic genes. Then least absolute shrinkage and selection operator (LASSO) regression analysis was performed to shrink the number of prognostic genes and simplify the prognostic model [23]. By increasing the lambda value, coefficients of prognostic genes close to zero, and the optimal lambda value was selected according to partial likelihood deviance. Finally, the prognostic model was defined as: risk score = Σ (gene expression i) * (gene coefficient i), where i represented prognostic genes.

TCGA-LUAD dataset served as a training dataset. GSE31210 and GSE50081 datasets were independent validation datasets. Risk score was converted to z-score, and z-score = 0 was the cut-off to divide samples into high-risk and low-risk groups. Kaplan-Meier survival analysis was performed to evaluate prognosis of the two groups. Receiver operating characteristic curve was used to assess the efficiency of the prognostic model for predicting 1-year, 3-year and 5-year overall survival.

Functional analysis related to risk score

In TCGA-LUAD dataset, GSEA was performed to analyze hallmark pathways "h.all.v7.4.symbols.gmt" from MSigDB for high-risk and low-risk groups. Norminal *P* value (NP) < 0.01 was selected to screen enriched pathways. To identify genes associated with risk score, Pearson correlation analysis in Hmisc R package (https://hbiostat.org/R/Hmisc/) was conducted. Pearson |R| > 0.4 and P < 0.05 were determined to screen genes significantly associated with risk score. WebGestaltR package was implemented to annotate Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [24].

Statistical analysis

All statistical analysis was performed in R (v4.1.1) software. Parameters not showing were default. Statistical methods were described in the corresponding sections. P < 0.05 was considered as significant. ns, no significance. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Results

Using pyroptosis-related genes to construct molecular subtypes

As pyroptosis has both suppressive and promotive roles in tumor development and the release of many pro-inflammatory factors, we first assessed whether there was a difference of pyroptosis-related genes between tumor and normal tissues in LUAD patients. A total of 27 genes in pyroptosis pathway were included. Based on CNV degree, three groups including amplification, diploid and deletion were classified. We found that 26 pyroptosis-related genes had significant copy number variations (CNVs) apart from CHMP2A (Supplementary Table 1). The expression level among the three groups showed a significant difference with relatively higher expression in amplification group (Supplementary Figure 1). In the single nucleotide variations, TP53 was the most mutated, accounting for a proportion of 48% in TCGA-LUAD samples (Supplementary Figure 2). Missense mutations consisted of majority of samples, followed by nonsense mutations.

Given that these pyroptosis-related genes were greatly altered, we then built a molecular subtyping system based on the gene expression. By using unsupervised consensus clustering, we set cluster number k from 2 to 10, and obtained consensus CDF curve of different cluster numbers (**Figure 1A**). According to the CDF and consensus matrix, cluster number k = 3 was selected to classify samples into three molecular subtypes (**Figure 1B**). Kaplan-Meier survival analysis showed differential overall survival (OS) among three subtypes (C1, C2 and C3) in all three datasets including TCGA-LUAD (P = 0.035, **Figure 1C**), GSE31210 (P = 0.0039, **Figure 1D**) and GSE50081 (P = 0.036, **Figure 1E**). C3 subtype had a better OS than C1 and C2 subtypes. The distinct OS among the three subtypes suggested that this

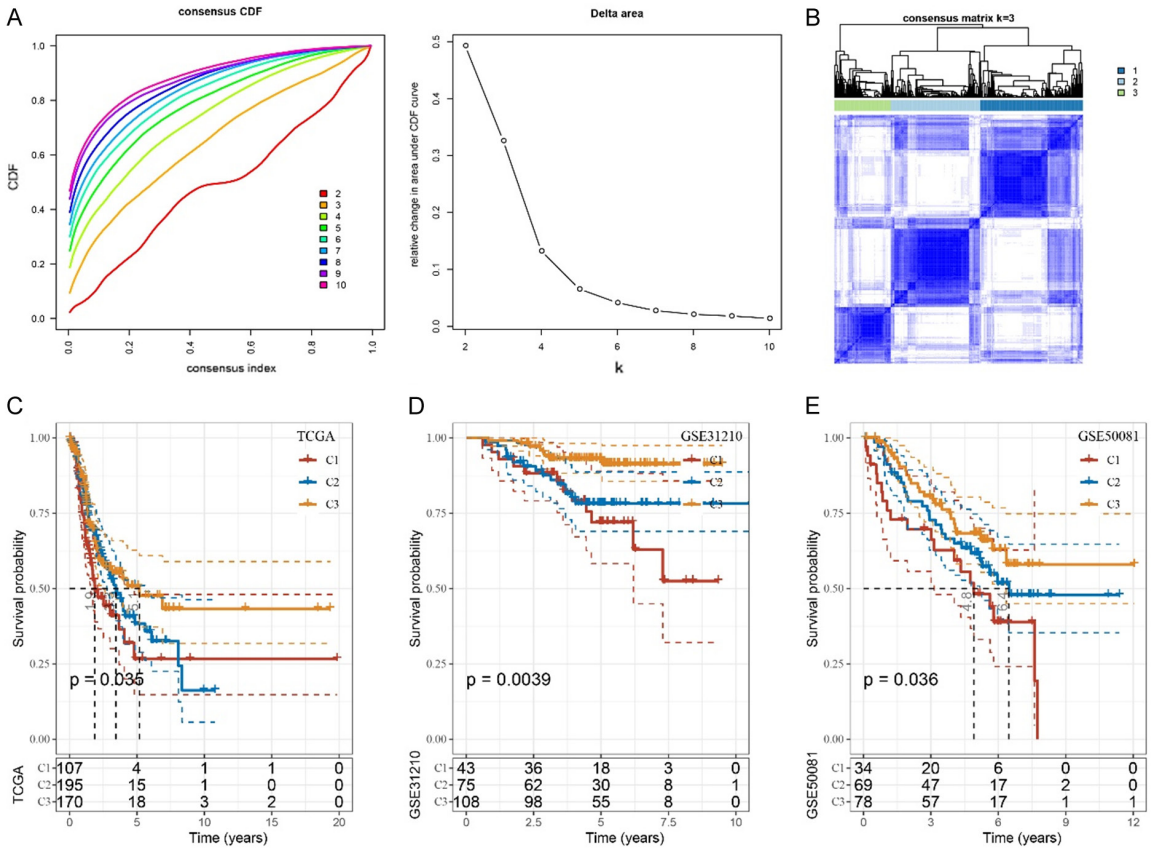


Figure 1. Constructing molecular subtypes for LUAD based on pyroptosis-related genes. A. Consensus CDF curve and delta area when cluster number k = 2 to 10 in TCGA-LUAD dataset. B. Consensus matrix when k = 3 by unsupervised consensus clustering in TCGA-LUAD dataset. C-E. Kaplan-Meier survival curves of three molecular subtypes in TCGA-LUAD, GSE31210 and GSE50081 datasets. Log-rank test was conducted. CDF, cumulative distribution function.

molecular subtyping based on pyroptosis-related genes was effective.

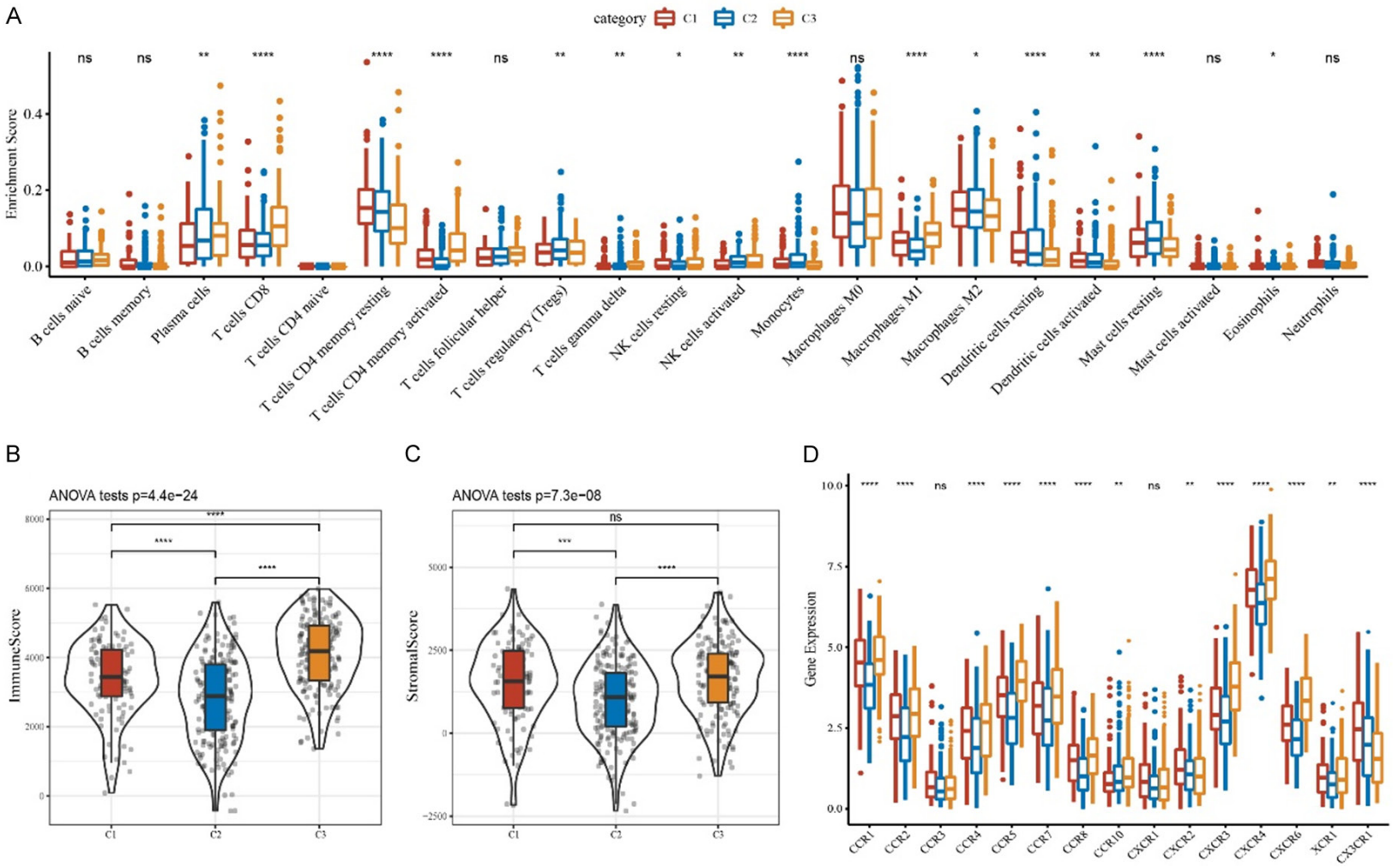
Differential TME among the three subtypes

Evidence supported that pyroptosis can promote the infiltration of immune cells and the generation of inflammatory TME [25, 26]. Therefore, we evaluated the infiltration of various immune cells through CIBERSORT. 15 of 22 immune cells were found to be differentially distributed among the three subtypes (P < 0.05, **Figure 2A**). Notably, CD8 T cells, activated memory CD4 T cells and M1 macrophages were more enriched in C3 subtype compared to other two subtypes. Simultaneously, C3 subtype had higher immune score and stromal score, indicating a high immune and stromal infiltration (**Figure 2B** and **2C**).

Chemokines and chemokine receptors are important mediators for assisting the infiltration of immune cells. Among 18 chemokine receptors and 41 chemokines, 16 out of 18 and 36 out of 41 presented significantly differential expression among the three subtypes (P < 0.05, **Figure 2D** and **2E**). Especially, a majority of chemokines and chemokine receptors were the most enriched in C3 subtype, which contributed to immune-infiltrated tumor microenvironment. In addition, we also assessed the expression of 47 immune checkpoints [27], and found that C2 subtype had the lowest expression level (**Figure 2F**). Apart from TNFSF18, other checkpoints were all differentially expressed among the three subtypes, suggesting that pyroptosis played a certain role in modulating tumor microenvironment.

Different pathways enriched in the three subtypes

To identify if there were different enriched pathways among the three subtypes, we firstly ana-


A
category C1 C2 C3
Enrichment Score
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
B
ANOVA tests p=4.4e−24
ImmuneScore
C
ANOVA tests p=7.3e−08
StromalScore
D
Gene Expression
CCR1
CCR2
CCR3
CCR4
CCR5
CCR7
CCR8
CCR10
CXCR1
CXCR2
CXCR3
CXCR4
CXCR6
XCR1
CX3CR1

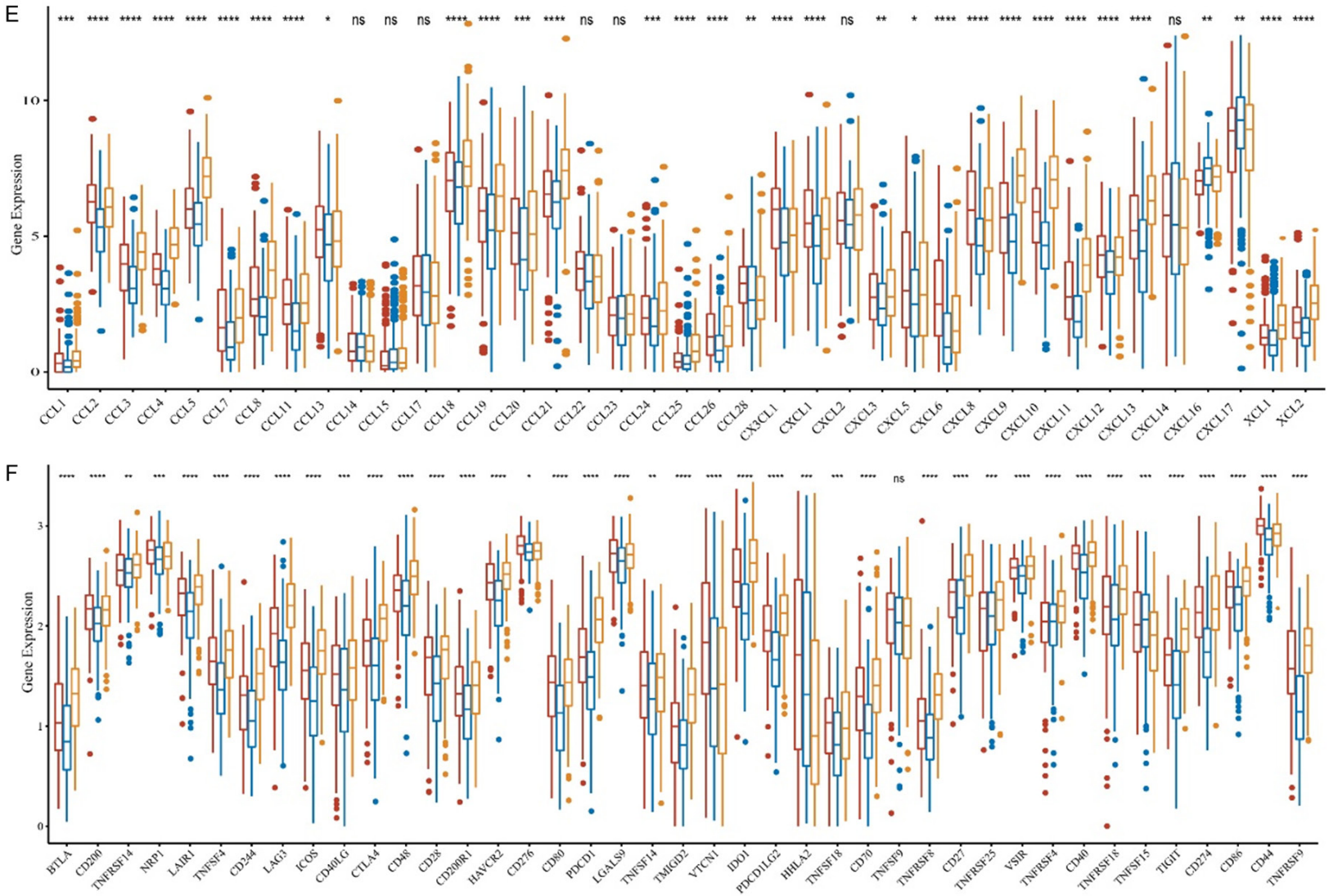


Figure 2. Characterization of TME in three molecular subtypes in TCGA-LUAD dataset. (A) Enrichment score of 22 immune cells through CIBERSORT. (B, C) Immune score (B) and stromal score (C) of three subtypes calculated by ESTIMATE. (D, E) Expression of chemokine receptors (D) and chemokines (E) in three subtypes. (F) Expression of 41 immune checkpoints in three subtypes. ANOVA was conducted among three groups. ns, no significance. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

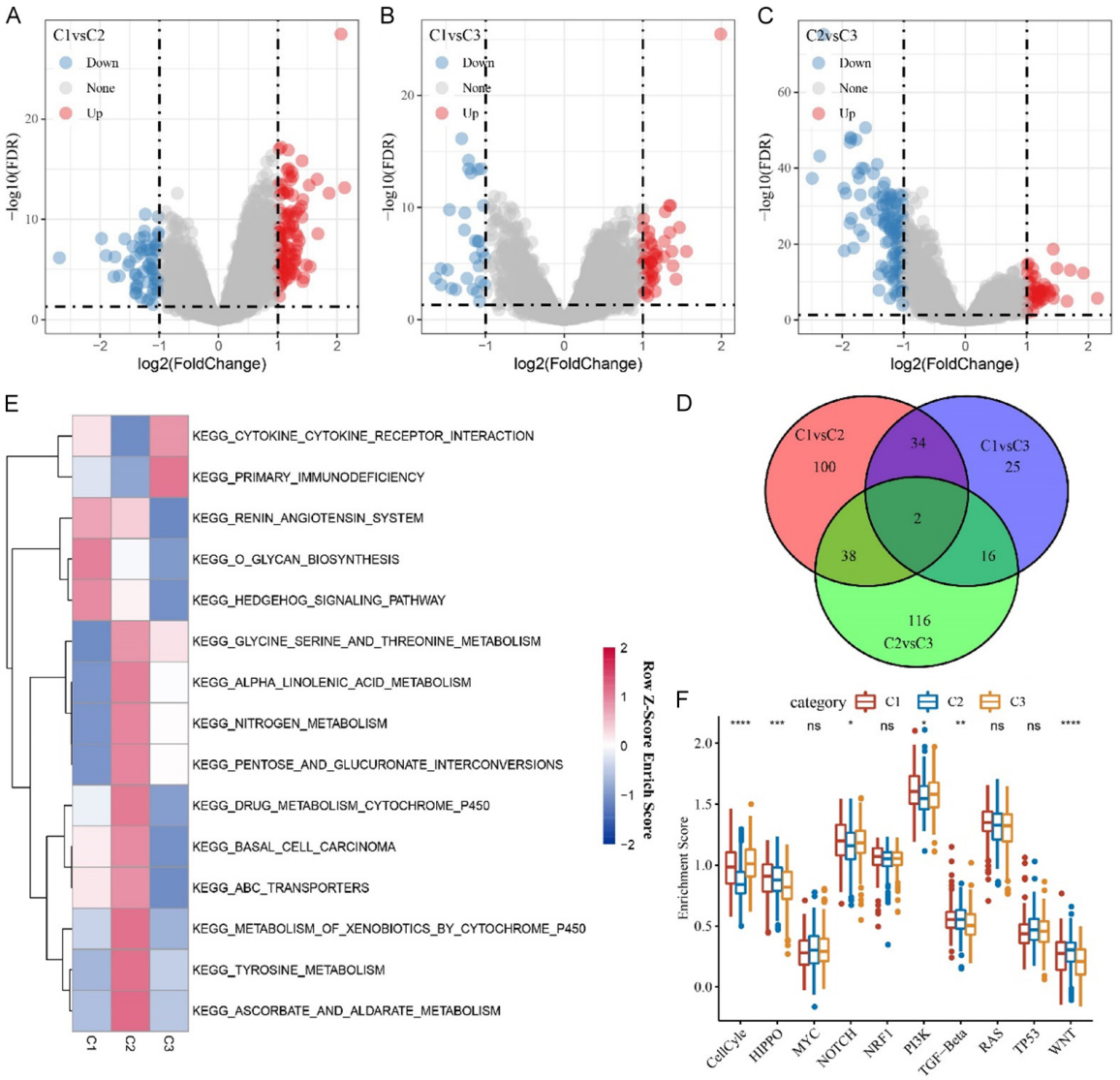


Figure 3. Enriched pathways of three subtypes in TCGA-LUAD dataset. (A-C) Identification of DEGs in C1 vs. C2 (A), C1 vs. C3 (B) and C2 vs. C3 (C). Blue indicates downregulated genes and red indicates upregulated genes. Horizontal dashed line indicates FDR < 0.05. Vertical dashed line indicates |fold change| > 2. (D) Venn plot of common DEGs between different subtypes. (E) Enrichment analysis of KEGG pathways in "c2.cp.kegg.v7.0.symbols.gmt" for three subtypes through ssGSEA. (F) Enrichment of oncogenic pathways in three subtypes. ANOVA was conducted. FDR, false discovery rate. ns, no significance. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

lyzed the DEGs between each of two groups (C1 vs. C2, C2 vs. C3 and C1 vs. C3) through limma R package [22]. By comparing C1 with C2 (C1 vs. C2), we screened 112 upregulated genes and 62 downregulated genes (**Figure 3A**). In C1 vs. C3, 47 upregulated and 30 downregulated genes were screened (**Figure 3B**). In C2 vs. C3, we identified 48 upregulated and 124 downregulated genes (**Figure 3C**). Venn plot of DEGs showed that only 2 same DEGs were identified in all the three subtypes (**Figure 3D**), indicating that the distinctly altered genes may lead to different enriched pathways.

Then ssGSEA was applied to investigate the enrichment of pathways in "c2.cp.kegg.v7.0.symbols.gmt" for each sample. The three subtypes manifested differential enrichment of the top 15 enriched pathways (**Figure 3E**). O-glycan biosynthesis, hedgehog ignaling pathway and renin angiotensin system were the most enriched pathways in C1 subtype, but showed a significantly lower enrichment in C2 and C3 subtypes. Metabolism-related pathways such as cytochrome P450, tyrosine metabolism and ascorbate and aldarate metabolism were highly enriched in C2 subtype, but

they were all downregulated in C1 and C3 subtypes. In addition, only cytokine-cytokine receptor interaction and primary immunodeficiency were obviously enriched in C3 subtype. Furthermore, we also compared the enrichment of 10 oncogenic pathways in the three subtypes. Six pathways of them were found to be differentially enriched among the three subtypes, including cell cycle, Hippo signaling, Notch signaling, PI3K signaling, TGF-beta signaling and Wnt signaling pathways (**Figure 3F**). Especially, cell cycle pathway was significantly downregulated in C2 subtype.

Establishing a prognostic model based on DEGs among three subtypes

Based on 90 DEGs (34 DEGs identified in both C1 vs. C2 and C1 vs. C3, 38 DEGs in C1 vs. C2 and C2 vs. C3, 16 DEGs in C1 vs. C2 and C1 vs. C3, and 2 DEGs in all three subtypes) among three subtypes (**Figure 4D**), we established a prognostic model. Univariate Cox regression analysis was performed to identify the genes significantly associated with overall survival, and 51 DEGs were screened with $P < 0.01$. To shrink the number of genes in the model, we applied LASSO Cox regression analysis to deduct gene numbers by increasing lambda value. The coefficients close to zero with the increasing lambda value (**Figure 4A**). According to the confidence interval of each lambda, the model was the optimal when lambda = 0.0404 (**Figure 4B**). Finally, 11 genes, including *IL1A*, *CLDN1*, *ANLN*, *PKIB*, *GJB3*, *MUC16*, *TNS4*, *DKK1*, *CPS1*, *HAS3* and *CYP4B1,* were retained. The prognostic model was defined as: risk score = 0.083*IL1A + 0.152*CLDN1 + 0.051*ANLN + 0.036*PKIB + 0.006*GJB3 + 0.041*MUC16 + 0.004*TNS4 + 0.028*DKK1 + 0.095*CPS1 - 0.084*HAS3 - 0.022*CYP4B1.

We calculated the risk score for each sample in TCGA-LUAD dataset, and converted risk score to z-score. Samples were classified into high-risk and low-risk groups according to the cut-off of z-score = 0. Samples with dead status were more in high-risk group (**Figure 4C**). The 11 genes were significantly differentially expressed in high-risk and low-risk groups. Apart from HAS3 and CYP4B1, other genes were all relatively higher-expressed in high-risk group (**Figure 4C**). ROC analysis showed that 1-year, 3-year and 5-year survival predicted by the prognostic model had favorable AUC score, with 0.70, 0.65 and 0.60, respectively (**Figure 4D**). Kaplan-Meier survival curve demonstrated that two groups had obviously different OS ($P < 0.0001$, **Figure 4E**), and risk score was a risk factor for survival (HR = 1.63, 95% CI = 1.42-1.88). To verify the effectiveness and robustness of the prognostic model, we examined it in the other two independent datasets (GSE31210 and GSE50081). The results indicated that the samples could be clearly classified into high-risk and low-risk groups with distinct prognosis ($P = 0.00077$ and $P < 0.0001$, Supplementary Figure 3).

Risk score was associated with clinical features

In the above section, we demonstrated that the 11-gene prognostic model was effective to classify LUAD samples into high-risk and low-risk groups with distinct OS. Next, we examined the robustness in samples with different clinical features including different ages, genders and stages. Except for T3 and T4, stage III and IV, samples with other clinical features could be all clearly divided into the two groups ($P < 0.05$, **Figure 5A**). Furthermore, to evaluate the independence of risk score as a risk factor, univariate and multivariate Cox regression analysis was conducted. Univariate analysis showed that T stage, N stage, stage and risk score (risk type) were risk factors for LUAD, with the risk score presenting the highest HR = 2.09 (95% CI = 1.55-2.80, $P < 0.0001$, **Figure 5B**). Multiple analysis revealed that only risk score was the independent risk factor with HR = 2.01 (95% CI = 1.39-2.89, $P < 0.0001$, **Figure 5C**). The results proved that the 11-gene prognostic model was reliable in predicting LUAD prognosis.

Functional pathways related to high-risk and low-risk score

The enriched pathways in high-risk and low-risk groups were assessed using GESA. Hallmark pathways in "h.all.v7.4.symbols.gmt" were included, and *NP* < 0.01 was set to screen enriched pathways. In low-risk group, metabolism-related pathways were significantly enriched, including taurine and hypotaurine metabolism (NES = 1.6, P = 0.034), alpha linolenic acid metabolism (NES = 1.6, P = 0.038), and valine leucine and isoleucine degradation (NES = 1.7, P = 0.025) (**Figure 6A**). In high-risk group, tumor-related pathways were enriched,

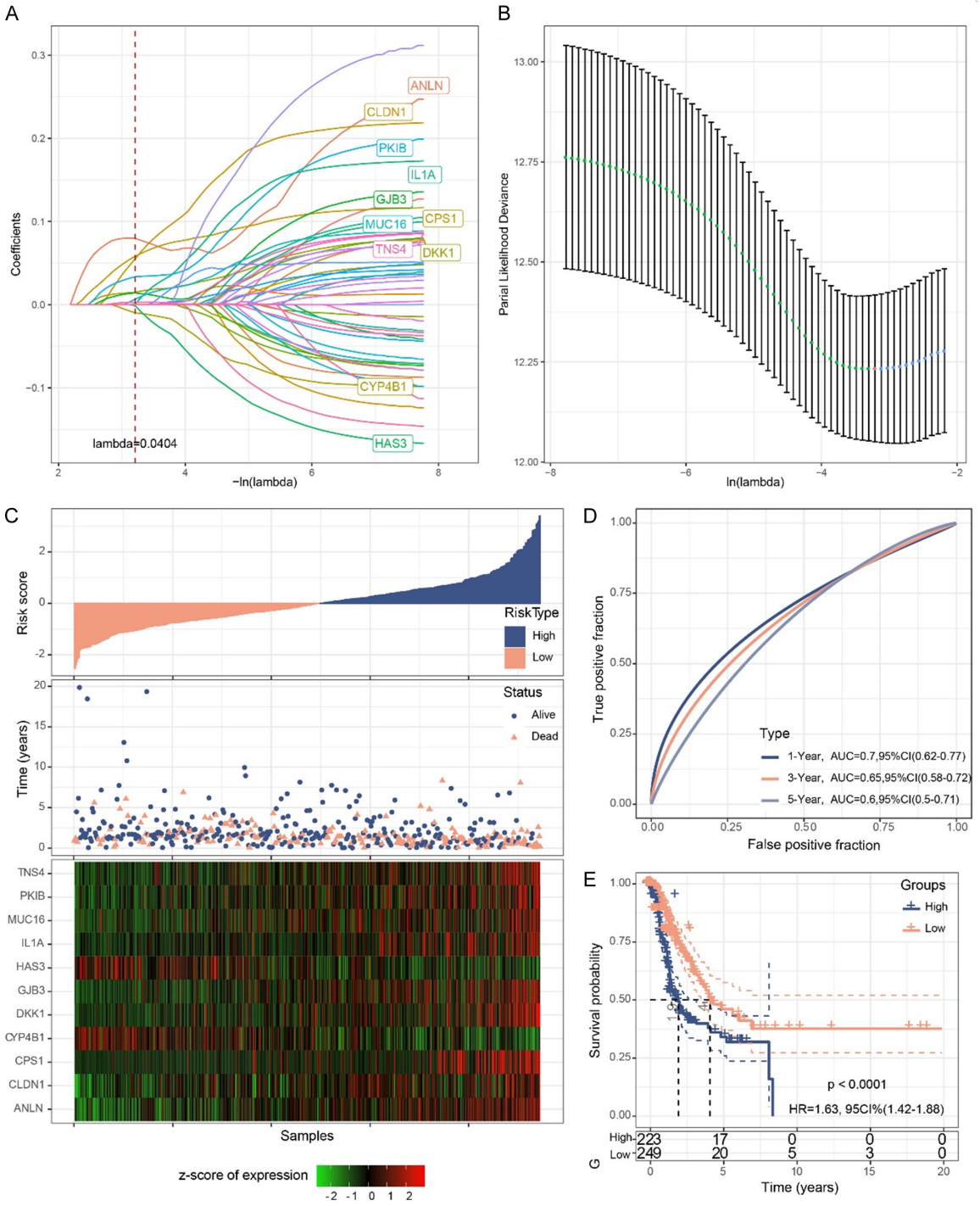
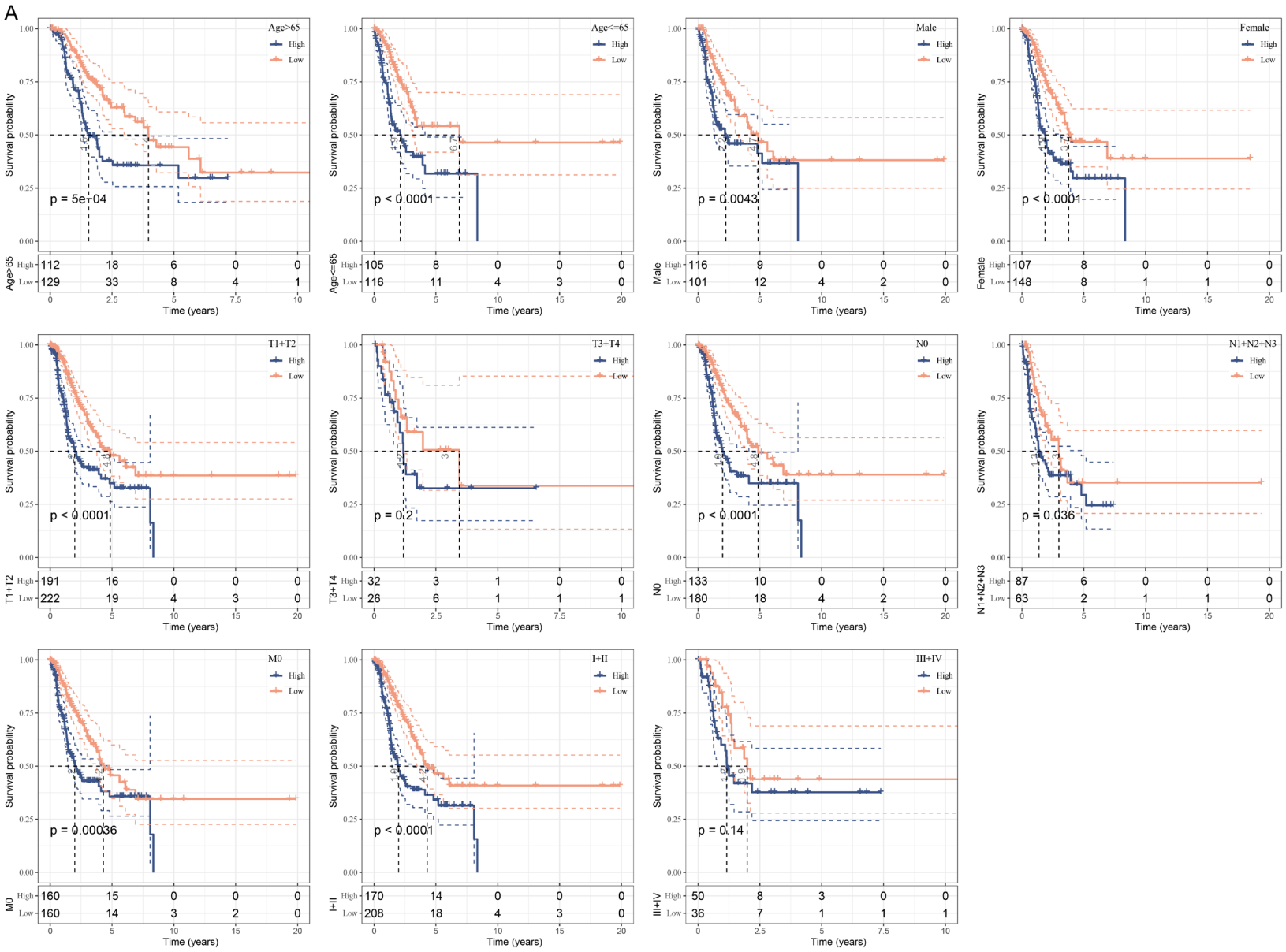


Figure 4. Construction of the 11-gene prognostic model in TCGA-LUAD dataset. A. The trajectory of gene coefficients changing with lambda value. Dashed line indicates lambda = 0.0404. B. Partial likelihood deviance of different lambda value. Red dot indicates lambda = 0.0404. C. The distribution of each sample and expression of 11 prognostic genes ranking by risk score from low to high. Of risk type, dark blue indicates high-risk group and orange indicates low-risk group. Of survival status, dark blue indicates alive samples and orange indicates dead samples. Of gene expression, relatively low to high expression was indicated from green to red. D. ROC curve of the prediction for 1-year, 3-year and 5-year by the prognostic model. E. Kaplan-Meier survival curve of high-risk and low-risk group in TCGA-LUAD dataset. Log-rank test was conducted. AUC, area under ROC curve. HR, hazard ratio. CI, confidence interval.


A
Age>65
High
Low
Survival probability
p = 5e-04
Age>65
High 112 18 6 0 0
Low 129 33 8 4 1
Time (years)
Age<=65
p < 0.0001
Age<=65
High 105 8 0 0 0
Low 116 11 4 3 0
Male
p = 0.0043
Male
High 116 9 0 0 0
Low 101 12 4 2 0
Female
p < 0.0001
Female
High 107 8 0 0 0
Low 148 8 1 1 0
T1+T2
p < 0.0001
T1+T2
High 191 16 0 0 0
Low 222 19 4 3 0
T3+T4
p = 0.2
T3+T4
High 32 3 1 0 0
Low 26 6 1 1 1
N0
p < 0.0001
N0
High 133 10 0 0 0
Low 180 18 4 2 0
N1+N2+N3
p = 0.036
N1+N2+N3
High 87 6 0 0 0
Low 63 2 1 1 0
M0
p = 0.00036
M0
High 160 15 0 0 0
Low 160 14 3 2 0
I+II
p < 0.0001
I+II
High 170 14 0 0 0
Low 208 18 4 3 0
III+IV
p = 0.14
III+IV
High 50 8 3 0 0
Low 36 7 1 1 1

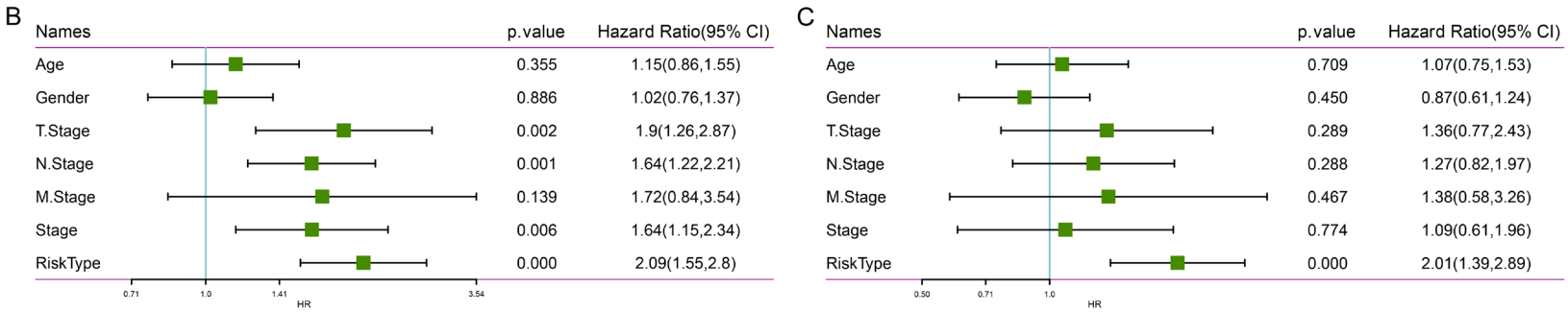
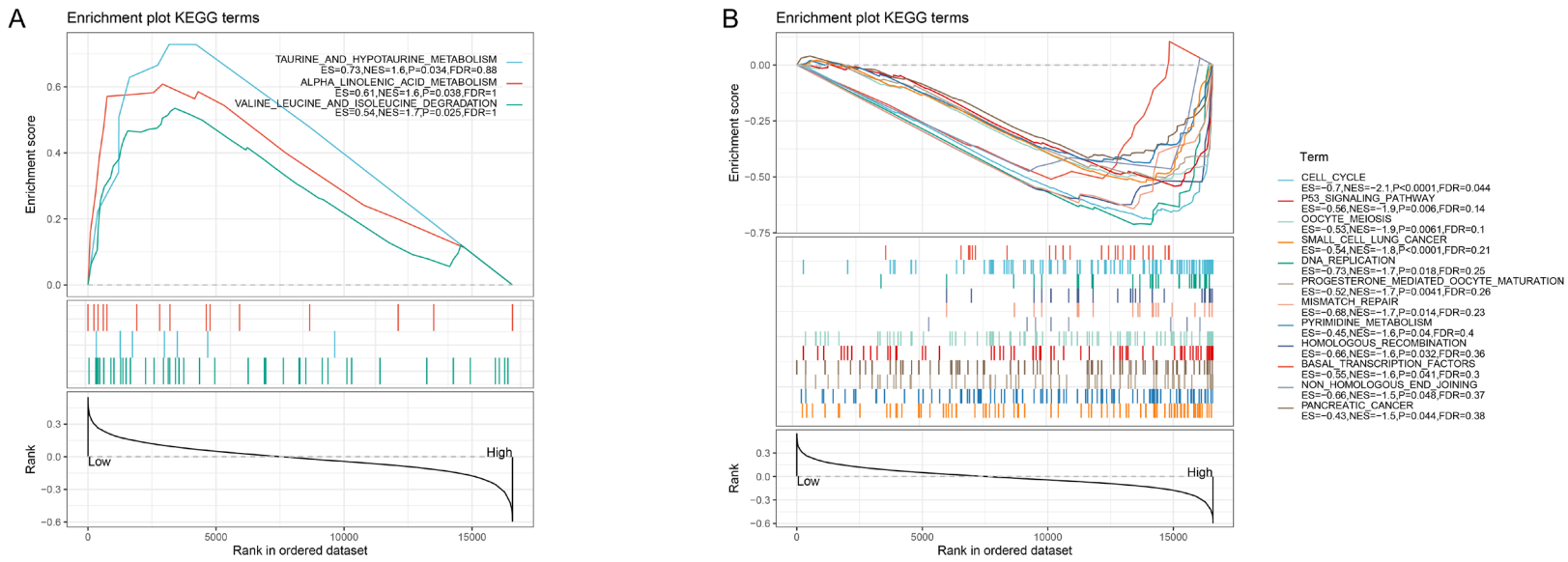


Figure 5. The relation between risk score and clinical features. (A) Kaplan-Meier survival plots of high-risk and low-risk groups grouping by different clinical features. Log-rank test was conducted. (B, C) Univariate (B) and multivariate (C) Cox regression analysis on clinical features and risk score. Log-rank test was conducted. HR, hazard ratio. CI, confidence interval.



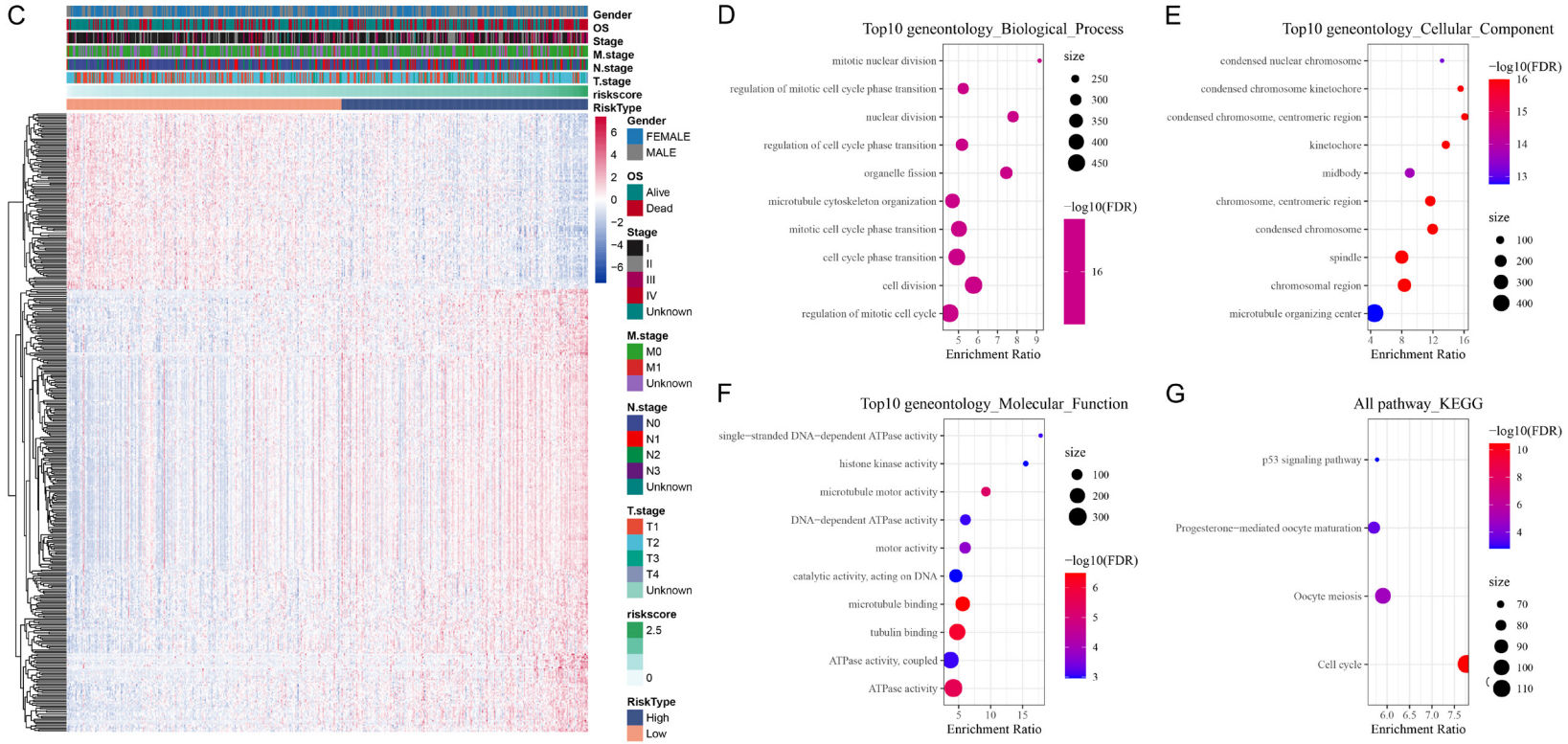


Figure 6. Functional analysis on pathways related to risk score in TCGA-LUAD dataset. (A, B) GSEA of low-risk (A) and high-risk (B) groups on "h.all.v7.4.symbols.gmt" hallmark pathways. Horizontal axis represents samples. Enriched pathways were labeled with different colors. (C) A heatmap of gene expression significantly associated with risk score based on Pearson correlation analysis. Horizontal axis represents samples labeled with different groups. Vertical axis of each line represents each gene. (D-F) GO function analysis through WebGestaltR on biological process (D), cellular component (E) and molecular function (F). The top 10 enriched terms were listed. (G) Four KEGG pathways associated with risk score. Size indicates gene counts. ES, enrichment score. NES, normalized enrichment score. FDR, false discovery rate.

for example, cell cycle (NES = -2.1, P < 0.0001), P53 signaling (NES = -1.9, P = 0.006), small cell lung cancer (NES = -1.8, P < 0.0001), and DNA replication (NES = -1.7, P = 0.018) (**Figure 6B**).

In addition, according to Pearson correlation analysis, 300 genes positively correlated with risk score and 119 genes negatively correlated with risk score in TCGA-LUAD dataset (|R| > 0.4 and P < 0.05). A heatmap showing the relation of gene expression and risk score demonstrated that high-risk and low-risk groups had obvious different expression patterns of these genes (**Figure 6C**). Furthermore, we performed KEGG and GO analysis on these genes to identify enriched pathways and GO terms. The top 10 enriched terms of biological process, cellular component and molecular function were visualized (P < 0.05, **Figure 6D-F**). Cell cycle-related terms, such as nuclear division, cell cycle phase transition and cell division, were enriched. Only four KEGG pathways containing two oncogenic pathways (P53 signaling pathway and cell cycle) were significantly enriched, (**Figure 6G**).

Discussion

In this study, we used consensus clustering to construct three molecular subtypes based on the expression of pyroptosis-related genes. The three subtypes displayed significantly differential prognosis and TME, suggesting that pyroptosis played an important role in immune modulation and tumor development in LUAD. Similar to the complexity of TME in tumorigenesis, pyroptosis is also considered to play dual roles (promotive and suppressive) in tumor cell development. In one hand, during the alternation of immune microenvironment, pyroptosis presents tumor-promoting effects through activating inflammasome and the release of cytokines [28], on the other hand, pyroptosis can be induced to eliminate tumor cells by using chemotherapeutic drugs [29].

Nod-like receptor protein 3 (NLRP3) signaling and caspase-1 signaling are two key components of pyroptosis signaling. NLRP3 is a type of inflammsomes, whose expression in macrophages is associated with survival and metastasis in the mouse model with breast cancer [28]. By knocking out NLRP3 and caspase-1 in THP1 monocytes, Pachathundikandi et al. found that pyroptotic cell death was inhibited along with suppressed interleukin-1β (IL-1β) expression [30]. In addition, Gao et al. proposed that GSDMD-mediated pyroptosis may help escape the innate immune response through downregulating the expression of IL-1β in human NSCLC tissues [31]. Inflammasomes can stimulate caspase-1-mediated pyroptosis and the release of cytokines especially IL-1β and IL-18, thereby creating a chronic inflammatory microenvironment [32].

Strong correlation between pyroptotic signaling and TME modulation has also been discovered in various cancers including in esophagus cancer [33], hepatocellular carcinoma [34], and pancreatic carcinoma [35]. Daley et al. revealed that NLRP3 could accelerate the accumulation of immunosuppressive macrophages in pancreatic carcinoma [35]. M2 macrophages express a high level of tumor-promoting NLRP3 and IL-1β, which can actuate CD4 T cells differentiating into T helper 2 cells (Th2 cells), Th17 cells, and regulatory T cells but inhibit the polarization of Th1 cells and the activation of cytotoxic CD8 T cells [35]. This supported the observation in our study that C3 subtype with the most favorable prognosis had lower enriched M2 macrophages and regulatory T cells, and higher enriched CD8 T cells. Simultaneously, pyroptosis-direct granzyme (Gzm) induces pyroptosis through activating caspases to cleave GSDM family [26]. It has been shown that Gzm-A and Gzm-B released from cytotoxic T lymphocytes (CTLs) induce pyroptotic cell death and enhance anti-tumor immune response through activating GSDMD cleavage [26, 36]. (CD8 T cells) in C3 subtype demonstrated that pyroptosis-related genes played a critical role in activating immune response.

In addition to immune cell modulation, pyroptosis is also demonstrated to enhance anti-tumor response in the treatment of combining PD-1/PD-L1 inhibitors with chemotherapy [37]. Hou et al. have found that chemotherapeutic drugs can induce pyroptosis mediated by gasdermin C (GSDMC)/Caspas-8 and nuclear PDL1 in breast cancer, which indicates that breast cancer patients with PD-L1+ or GSDMC+ may benefit much from chemotherapeutic drugs [37]. Therefore, PD-L1 is considered to control pyroptosis leading to tumor necrosis [38]. In the current study, we observed significantly high expression of PDCD1 (PD-1) and CD274 (PD-

L1) in C3 subtype. To some extent, the superior prognosis of C3 subtype may partially result from PD-L1-mediated pyroptosis, and C3 subtype may also exhibit more favorable outcome if treated by combined therapy.

Among the three subtypes, functional pathways were distinctly enriched. Especially, oncogenic pathways were differentially accumulated, such as cell cycle, Hippo signaling, TGF-β signaling and WNT signaling pathways. MST1 is one of vital components in Hippo pathway, and its overexpression can increase cellular reactive oxygen species (ROS) and activate caspase-1 to induce pyroptosis [39]. Orning et al. have demonstrated that blocking TGF-β activated kinase-1 (TAK1) or IKK kinases elicits non-inflammatory caspase-8 to cleave GSDMD and leads to pyroptosis-mediated cell death [40]. Wellenstein et al. have revealed that suppressing WNT secretion can inhibit IL-1β release from macrophages, and hinder metastasis in breast cancer cells [41]. The differential enrichment of these pathways in the three subtypes further supported close associations between pyroptosis and these oncogenic pathways.

Based on DEGs among the three subtypes, we identified 11 prognostic genes that could serve as biomarkers to predict prognosis for LUAD patients. Tumor-related pathways such as cell cycle and p53 signaling were enriched in high-risk group, which may contribute to its unfavorable prognosis. Distinct expression patterns between the two groups ranked by risk score demonstrated that these 11 genes may play important roles in regulating pathways and gene expression. Furthermore, the risk score showed a stronger performance as an independent risk factor compared with clinical features.

Conclusions

In conclusion, this study showed that pyroptosis was a critical factor in tumor development in LUAD. According to the expression of pyroptosis-related genes, we developed three molecular subtypes with different OS, TME and enriched pathways. Pyroptosis as one of important cell death process plays essential roles in modulating immune microenvironment. Notably, the 11-gene signature related to pyroptosis could be an indicator to predict LUAD prognosis. The 11 prognostic genes may be new directions for revealing pyroptosis mechanism in tumorigenesis, or new targets for exploiting therapeutic drugs.

Acknowledgements

This study was funded by [Major Program of China Medicine Education Association] under grant number [2020KTS008], [Hefei Science and Technology Bureau Fund] under grant number [J2020Y01].

Disclosure of conflict of interest

None.

Address correspondence to: Liting Qian, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230031, Anhui, China. Tel: +86-0551-63602184; E-mail: money2004@sina.com

References

[1] Zappa C and Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res 2016; 5: 288-300.

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[3] Denisenko TV, Budkevich IN and Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma. Cell Death Dis 2018; 9: 117.

[4] Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, Alvarez JJS, Lu B, Lim JQ, Takano A, Nahar R, Lee YY, Phua CZJ, Chua KP, Suteja L, Chen PJ, Chang MM, Koh TPT, Ong BH, Anantham D, Hsu AAL, Gogna A, Too CW, Aung ZW, Lee YF, Wang L, Lim TKH, Wilm A, Choi PS, Ng PY, Toh CK, Lim WT, Ma S, Lim B, Liu J, Tam WL, Skanderup AJ, Yeong JPS, Tan EH, Creasy CL, Tan DSW, Hillmer AM and Zhai W. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet 2020; 52: 177-186.

[5] Cerella C, Teiten MH, Radogna F, Dicato M and Diederich M. From nature to bedside: pro-survival and cell death mechanisms as therapeutic targets in cancer treatment. Biotechnol Adv 2014; 32: 1111-1122.

[6] Fang Y, Tian S, Pan Y, Li W, Wang Q, Tang Y, Yu T, Wu X, Shi Y, Ma P and Shu Y. Pyroptosis: a new frontier in cancer. Biomed Pharmacother 2020; 121: 109595.

[7] Shi J, Zhao Y, Wang K, Shi X, Wang Y, Huang H, Zhuang Y, Cai T, Wang F and Shao F. Cleavage

of GSDMD by inflammatory caspases determines pyroptotic cell death. Nature 2015; 526: 660-665.
[8] Liu X, Zhang Z, Ruan J, Pan Y, Magupalli VG, Wu H and Lieberman J. Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores. Nature 2016; 535: 153-158.
[9] Ramos-Junior ES and Morandini AC. Gasdermin: a new player to the inflammasome game. Biomed J 2017; 40: 313-316.
[10] Wu M, Wang Y, Yang D, Gong Y, Rao F, Liu R, Danna Y, Li J, Fan J, Chen J, Zhang W and Zhan Q. A PLK1 kinase inhibitor enhances the chemosensitivity of cisplatin by inducing pyroptosis in oesophageal squamous cell carcinoma. EBioMedicine 2019; 41: 244-255.
[11] Wang WJ, Chen D, Jiang MZ, Xu B, Li XW, Chu Y, Zhang YJ, Mao R, Liang J and Fan DM. Downregulation of gasdermin D promotes gastric cancer proliferation by regulating cell cycle-related proteins. J Dig Dis 2018; 19: 74-83.
[12] Wang F, Liu W, Ning J, Wang J, Lang Y, Jin X, Zhu K, Wang X, Li X, Yang F, Ma J and Xu S. Simvastatin suppresses proliferation and migration in non-small cell lung cancer via pyroptosis. Int J Biol Sci 2018; 14: 406-417.
[13] Sannino F, Sansone C, Galasso C, Kildgaard S, Tedesco P, Fani R, Marino G, de Pascale D, Ianora A, Parrilli E, Larsen TO, Romano G and Tutino ML. Pseudoalteromonas haloplanktis TAC125 produces 4-hydroxybenzoic acid that induces pyroptosis in human A459 lung adenocarcinoma cells. Sci Rep 2018; 8: 1190.
[14] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP and Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015; 1: 417-425.
[15] Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES and Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013; 31: 213-219.
[16] Wilkerson MD and Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010; 26: 1572-1573.
[17] Chen B, Khodadoust MS, Liu CL, Newman AM and Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol 2018; 1711: 243-259.
[18] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB and Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.
[19] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.
[20] Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, Michel K, Mermel C, Silver SJ, Weir BA, Reiling JH, Sheng Q, Gupta PB, Wadlow RC, Le H, Hoersch S, Wittner BS, Ramaswamy S, Livingston DM, Sabatini DM, Meyerson M, Thomas RK, Lander ES, Mesirov JP, Root DE, Gilliland DG, Jacks T and Hahn WC. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009; 462: 108-112.
[21] Hänzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013; 14: 7.
[22] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.
[23] Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33: 1-22.
[24] Liao Y, Wang J, Jaehnig EJ, Shi Z and Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res 2019; 47: W199-W205.
[25] Zhang Z, Zhang Y, Xia S, Kong Q, Li S, Liu X, Junqueira C, Meza-Sosa KF, Mok TMY, Ansara J, Sengupta S, Yao Y, Wu H and Lieberman J. Gasdermin E suppresses tumour growth by activating anti-tumour immunity. Nature 2020; 579: 415-420.
[26] Zhou Z, He H, Wang K, Shi X, Wang Y, Su Y, Wang Y, Li D, Liu W, Zhang Y, Shen L, Han W, Shen L, Ding J and Shao F. Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells. Science 2020; 368: eaaz7548.
[27] Danilova L, Ho WJ, Zhu Q, Vithayathil T, De Jesus-Acosta A, Azad NS, Laheru DA, Fertig EJ, Anders R, Jaffee EM and Yarchoan M. 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 2019; 7: 886-895.
[28] Weichand B, Popp R, Dziumbla S, Mora J, Strack E, Elwakeel E, Frank AC, Scholich K,

Pierre S, Syed SN, Olesch C, Ringleb J, Ören B, Döring C, Savai R, Jung M, von Knethen A, Levkau B, Fleming I, Weigert A and Brüne B. S1PR1 on tumor-associated macrophages promotes lymphangiogenesis and metastasis via NLRP3/IL-1β. J Exp Med 2017; 214: 2695-2713.

[29] Okondo MC, Johnson DC, Sridharan R, Go EB, Chui AJ, Wang MS, Poplawski SE, Wu W, Liu Y, Lai JH, Sanford DG, Arciprete MO, Golub TR, Bachovchin WW and Bachovchin DA. DPP8 and DPP9 inhibition induces pro-caspase-1-dependent monocyte and macrophage pyroptosis. Nat Chem Biol 2017; 13: 46-53.

[30] Pachathundikandi SK, Blaser N, Bruns H and Backert S. Helicobacter pylori avoids the critical activation of nlrp3 inflammasome-mediated production of oncogenic mature il-1β in human immune cells. Cancers (Basel) 2020; 12: 803.

[31] Gao J, Qiu X, Xi G, Liu H, Zhang F, Lv T and Song Y. Downregulation of GSDMD attenuates tumor proliferation via the intrinsic mitochondrial apoptotic pathway and inhibition of EGFR/Akt signaling and predicts a good prognosis in non-small cell lung cancer. Oncol Rep 2018; 40: 1971-1984.

[32] Li L, Jiang M, Qi L, Wu Y, Song D, Gan J, Li Y and Bai Y. Pyroptosis, a new bridge to tumor immunity. Cancer Sci 2021; 112: 3979-3994.

[33] Barber G, Anand A, Katarzyna O, Phelan JJ, Heeran AB, Flis E, Clarke NE, Watson JA, Strangmann J, Flood B, O'Neill H, O'Toole D, MacCarthy F, Ravi N, Reynolds JV, Kay EW, Quante M, O'Sullivan J and Creagh EM. Characterizing caspase-1 involvement during esophageal disease progression. Cancer Immunol Immunother 2020; 69: 2635-2649.

[34] Luan J and Ju D. Inflammasome: a double-edged sword in liver diseases. Front Immunol 2018; 9: 2201.

[35] Daley D, Mani VR, Mohan N, Akkad N, Pandian G, Savadkar S, Lee KB, Torres-Hernandez A, Aykut B, Diskin B, Wang W, Farooq MS, Mahmud AI, Werba G, Morales EJ, Lall S, Wadowski BJ, Rubin AG, Berman ME, Narayanan R, Hundeyin M and Miller G. NLRP3 signaling drives macrophage-induced adaptive immune suppression in pancreatic carcinoma. J Exp Med 2017; 214: 1711-1724.

[36] Xi G, Gao J, Wan B, Zhan P, Xu W, Lv T and Song Y. GSDMD is required for effector CD8(+) T cell responses to lung cancer cells. Int Immunopharmacol 2019; 74: 105713.

[37] Hou J, Zhao R, Xia W, Chang CW, You Y, Hsu JM, Nie L, Chen Y, Wang YC, Liu C, Wang WJ, Wu Y, Ke B, Hsu JL, Huang K, Ye Z, Yang Y, Xia X, Li Y, Li CW, Shao B, Tainer JA and Hung MC. PD-L1-mediated gasdermin C expression switches apoptosis to pyroptosis in cancer cells and facilitates tumour necrosis. Nat Cell Biol 2020; 22: 1264-1275.

[38] Blasco MT and Gomis RR. PD-L1 controls cancer pyroptosis. Nat Cell Biol 2020; 22: 1157-1159.

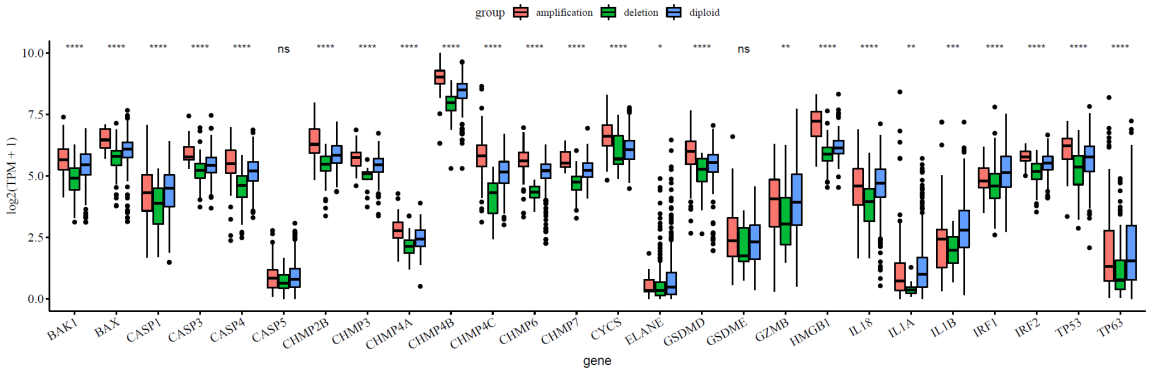
[39] Cui J, Zhou Z, Yang H, Jiao F, Li N, Gao Y, Wang L, Chen J and Quan M. MST1 suppresses pancreatic cancer progression via ROS-induced pyroptosis. Mol Cancer Res 2019; 17: 1316-1325.

[40] Orning P, Weng D, Starheim K, Ratner D, Best Z, Lee B, Brooks A, Xia S, Wu H, Kelliher MA, Berger SB, Gough PJ, Bertin J, Proulx MM, Goguen JD, Kayagaki N, Fitzgerald KA and Lien E. Pathogen blockade of TAK1 triggers caspase-8-dependent cleavage of gasdermin D and cell death. Science 2018; 362: 1064-1069.

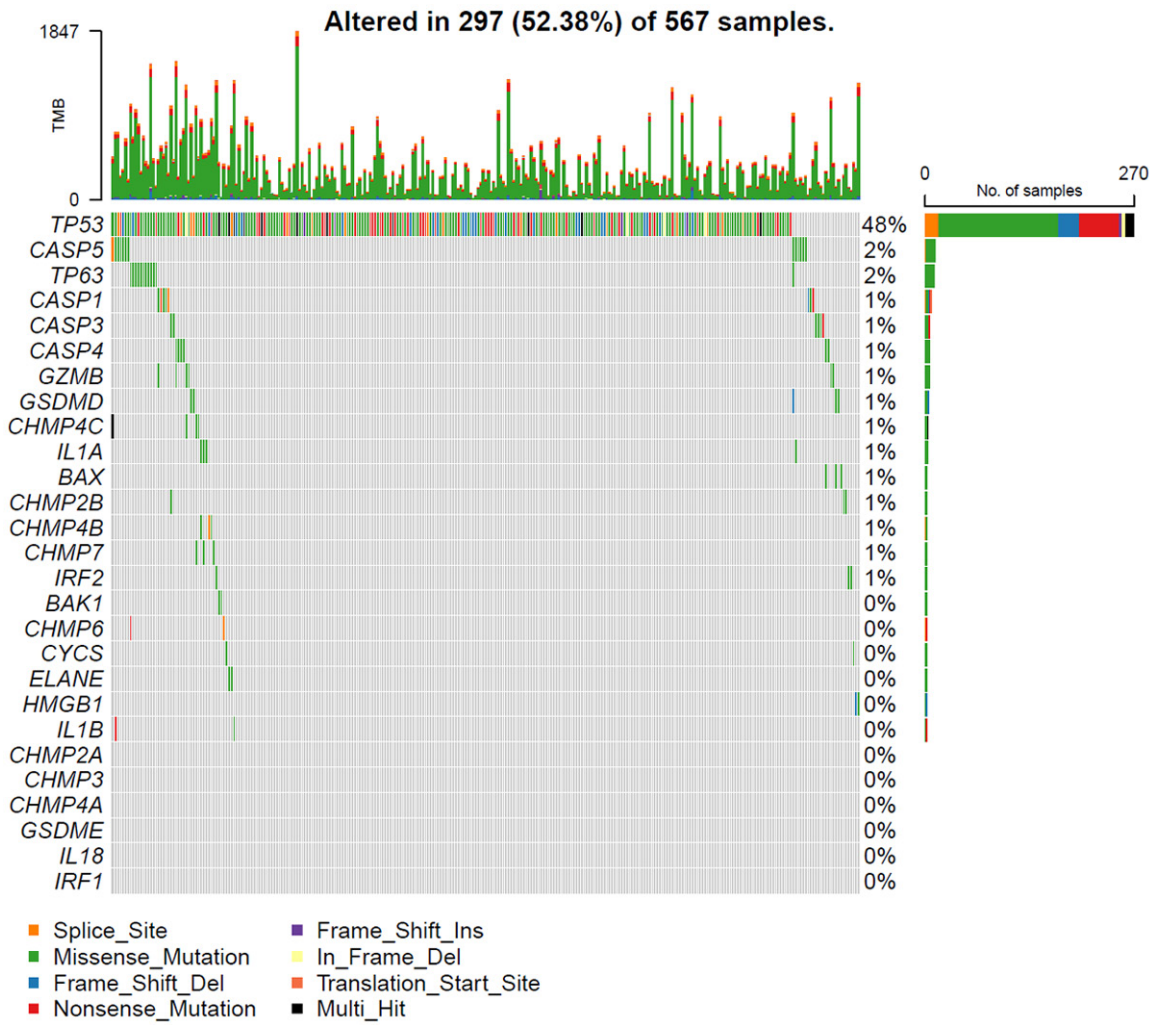
[41] Wellenstein MD, Coffelt SB, Duits DEM, van Miltenburg MH, Slagter M, de Rink I, Henneman L, Kas SM, Prekovic S, Hau CS, Vrijland K, Drenth AP, de Korte-Grimmerink R, Schut E, van der Heijden I, Zwart W, Wessels LFA, Schumacher TN, Jonkers J and de Visser KE. Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. Nature 2019; 572: 538-542.

Supplementary Table S1. The distribution of CNVs in 27 pyroptosis-related genes

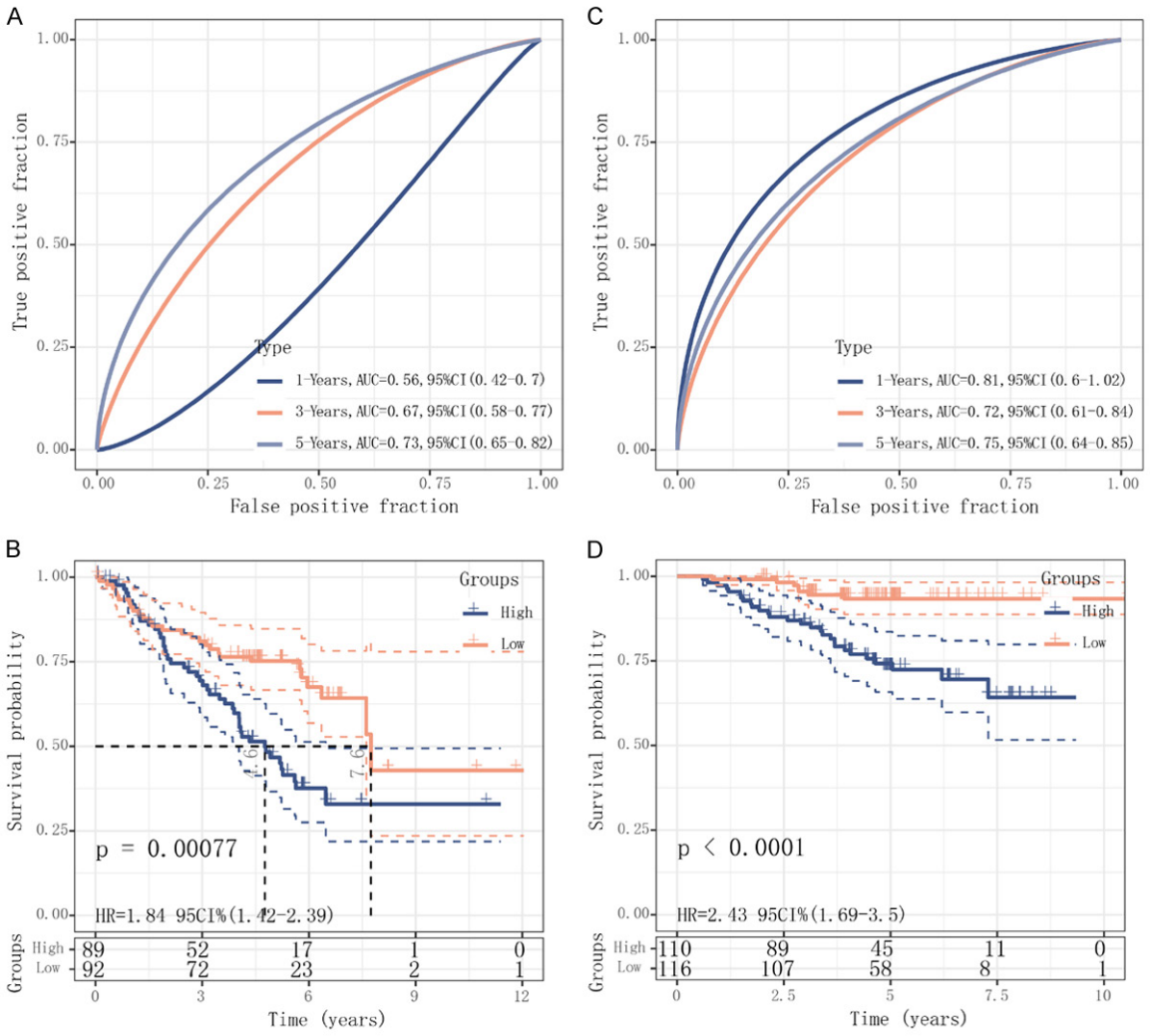
gene	amplification	deletion	diploid	amplification %	deletion %	diploid %
CHMP3	67	14	474	12.07	2.52	85.41
IL1A	56	9	490	10.09	1.62	88.29
IL1B	56	9	490	10.09	1.62	88.29
CHMP2B	29	109	420	5.2	19.53	75.27
TP63	89	77	400	15.72	13.6	70.67
IRF2	19	118	420	3.41	21.18	75.4
CASP3	19	118	417	3.43	21.3	75.27
IRF1	38	140	377	6.85	25.23	67.93
BAK1	106	34	421	18.89	6.06	75.04
GSDME	216	17	323	38.85	3.06	58.09
CYCS	216	17	321	38.99	3.07	57.94
CHMP7	29	228	297	5.23	41.16	53.61
CHMP4C	202	20	332	36.46	3.61	59.93
GSDMD	227	17	310	40.97	3.07	55.96
CASP4	69	63	422	12.45	11.37	76.17
CASP5	69	63	424	12.41	11.33	76.26
CASP1	69	63	423	12.43	11.35	76.22
IL18	62	74	419	11.17	13.33	75.5
HMGB1	13	190	358	2.32	33.87	63.81
CHMP4A	91	47	416	16.43	8.48	75.09
GZMB	91	46	417	16.43	8.3	75.27
TP53	12	204	338	2.17	36.82	61.01
CHMP6	150	12	392	27.08	2.17	70.76
ELANE	15	187	350	2.72	33.88	63.41
BAX	22	119	413	3.97	21.48	74.55
CHMP4B	105	30	421	18.88	5.4	75.72



Supplementary Figure S1. The distribution of amplifications, deletions and diploid in 27 pyroptosis-related genes. TPM, transcript per million.



Supplementary Figure S2. Gene mutations of 27 pyroptosis-related genes ranking by mutation frequency.



Supplementary Figure S3. Validation of the prognostic model in GEO datasets. (A, B) ROC curve and Kaplan-Meier survival curve in GSE50081 dataset. (C, D) ROC curve and Kaplan-Meier survival curve in GSE31210 dataset. AUC, area under ROC curve. HR, hazard ratio. CI, confidence interval.