Systematic analysis of apoptosis-related genes in the prognosis of Lung squamous cell carcinoma: a combined single-cell RNA sequencing study

Peiquan Zhu
Southwest Medical University
Wenxing Yang
Southwest Medical University
Biao Wang
Affiliated Hospital of Southwest Medical University
Zhi Hu
Affiliated Hospital of Southwest Medical University
Dengguo Zhang
Affiliated Hospital of Southwest Medical University
Ze Yang
Southwest Medical University
Kaiqiang Wang
Southwest Medical University
Jiangtao Pu (pujiangtao1972@sina.com)
Affiliated Hospital of Southwest Medical University

Research Article

Keywords: Lung squamous cell carcinoma (LUSC), prognostic model, biomarkers, Apoptosis, The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO)

Posted Date: June 16th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3054179/v1

Abstract

Purpose

Lung squamous cell carcinoma (LUSC) has a poor prognosis and lacks appropriate diagnostic and treatment strategies.Apoptosis dysregulation is associated with tumor occurrence and drug resistance, but the prognostic value of apoptosis-related genes (ARGs) in LUSC remains unclear.

Methods

We constructed an ARGs model that can predict LUSC through univariate Cox regression, least absolute shrinkage and selection operator (LASSO) regression, and multivariate Cox regression analysis based on differentially expressed ARGs. We conducted correlation analysis of prognostic ARGs by combining the dataset of normal lung tissue from the Genotype-Tissue Expression (GTEx) database. Then, we constructed a risk model and the predictive ability of the model was evaluated by using ROC (Receiver Operating Characteristic Curve) analysis. NSCLC single-cell RNA sequencing (scRNA-seq) data were downloaded from the Gene Expression Omnibus (GEO) database. Cell subgroups were determined and annotated by dimensionality reduction clustering, and the cell subgroups in disease development were clarified by establishing pseudotime analysis using Monocle.

Results

We identified four apoptosis prognostic genes and constructed a stable prognostic risk model. Kaplan-Meier curve analysis showed that the high-risk group had a poorer prognosis ($P < 0.05$). Furthermore, the ROC curve confirmed that the model had good predictive value for LUSC patients. Through analysis of single-cell sequencing data, apoptosis prognostic genes were found to be enriched in epithelial cells, smooth muscle cells, and T cells. Pseudotime analysis was used to infer the differentiation process and time sequence of cells.

Conclusions

This study identified apoptosis-related genes that are associated with prognosis in LUSC, and constructed a risk model based on these prognostic genes that accurately predicts the prognosis of LUSC. Single-cell sequencing analysis provided new insights into the cellular-level development of tumors. These findings provide more guidance for the diagnosis and treatment of LUSC patients.

1 INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths, accounting for 19.4% of cancer-related deaths annually (1). Non-small cell lung cancer (NSCLC) accounts for over 80% of all lung cancer cases and

includes two main types: non-squamous cell carcinoma, including adenocarcinoma, large cell carcinoma, and other cell types, and squamous cell carcinoma, which accounts for approximately 25–30% of NSCLC cases (2). Squamous cell carcinoma of the lung often grows in the central part of the lung and has the ability to grow to a large size (3). Compared to adenocarcinoma, patients with advanced non-small cell lung cancer have a worse prognosis, due to some clinical features that are different from other non-small cell lung cancer histologies, such as smoking history, comorbidities, age, and molecular characteristics. Overall, these factors make LUSC a particularly challenging disease (4). Comparative results from relevant clinical trials show that the median survival of patients with advanced squamous cell carcinoma who receive first-line treatment is about 30% shorter than that of patients with other non-small cell lung cancer subtypes(5). The TNM staging system is widely used for LUSC, but it has certain limitations because patients with the same TNM stage can have different survival outcomes (6). Currently, there have been significant improvements in molecular targeted therapy and immunotherapy for non-small cell lung cancer. However, the vast majority of advanced non-small cell lung cancer cases develop resistance to current treatment methods (7). Therefore, it is necessary to identify new biomarkers that can provide diagnostic and prognostic information for LUSC.

Cell apoptosis is an evolutionarily conserved, genetically regulated form of cell suicide that plays an important role in the development and maintenance of tissue homeostasis in multicellular organisms (8). Cell apoptosis plays a crucial role in eliminating cells that have undergone mutations or transformation in the body. Therefore, cancer cells often evolve multiple efficient and diverse mechanisms to evade cell apoptosis (9). Cell apoptosis can be induced through both extrinsic and intrinsic pathways. The extrinsic pathway is stimulated by death receptors, such as Fas, tumor necrosis factor receptors, and TRAIL, while the intrinsic pathway is initiated by DNA damage, energy depletion, and hypoxia, which can lead to dephosphorylation and cleavage of pro-apoptotic proteins (10). Genetic analysis of NSCLC has revealed inherited and somatic mutations in the EGFR and P53 genes, as well as somatic mutations in the KRAS, BRAF, ERBB2, MET, STK11, PIK3CA, and PARK2 genes. These gene mutations have led to the development of new strategies targeting these targets (11). Exploring the targeted apoptosis mechanisms in NSCLC represents another approach aimed at selectively killing cancer cells while preserving normal cells. However, related studies have shown that there is no correlation between the response to apoptosis-targeted drugs and the histological subtypes of NSCLC. The presence of apoptosis targets, such as TRAIL receptors and DISC compounds, anti-apoptotic BCL-2 and IAP family members, is important, but cannot predict the response to corresponding targeted therapies (12). Moreover, LUSC patients often have a poor prognosis. Therefore, identifying key molecules, establishing stable and effective predictive models, and implementing precision treatment are crucial for improving the prognosis of LUSC patients.

In recent years, single-cell RNA sequencing (scRNA-seq) has developed rapidly. Compared to conventional RNA sequencing (RNA-seq) methods, although RNA-seq studies the entire transcriptome in detail and has led to many important discoveries, RNA-seq is usually performed in "bulk", and the data represents the average gene expression patterns of thousands to millions of cells, which may mask biological differences between cells. In contrast, scRNA-seq captures the transcriptomes of individual cells and generates sequencing libraries, where the transcripts are mapped to individual cells. scRNA-seq allows for

unprecedented resolution to evaluate the fundamental biological properties of cell populations and biological systems (13). By exploring the heterogeneity of cells and the tumor microenvironment at the single-cell level, we can greatly improve our understanding of the transcriptional, genetic, metabolic, and other characteristics of thousands of individual cells, thereby deepening our analysis of the cells involved in tumor progression (14). Integrating clinical pathological information and single-cell sequencing data can facilitate the identification of novel diagnostic and prognostic biomarkers, as well as potential treatment-related cell types or states (15). However, scRNA-seq is relatively expensive, and therefore the available sample datasets are relatively limited. Nevertheless, the information obtained from scRNA-seq is highly valuable for exploring the characteristics of each cell subpopulation in the sample and the interactions of each cell in the tumor immune microenvironment (TIME) (16). In this study, we aimed to construct an apoptosis prognostic model associated with LUSC survival. Through bioinformatics analysis, we identified four ARGs that were associated with LUSC prognosis. We constructed a risk score based on these ARGs and developed a prognostic model that was associated with the risk score. The model was validated in a test set, and we also investigated the potential role of the risk score in guiding immunotherapy for LUSC patients. Finally, we further elucidated the role of the prognostic-related ARGs at the cellular level in the occurrence and development of LUSC, improving the treatment outcomes and prognosis of LUSC patients at the single-cell level.

2 Methods

2.1 Data source and pre-processing

RNA sequencing data and clinical information of LUSC patients were obtained from The Cancer Genome Atlas (TCGA) database. overall survival (OS) and other clinical features, including age, gender, T stage, N stage, M stage, and clinical stage, were extracted from the downloaded cases and missing clinical information was removed. scRNA-seq data were downloaded from the GEO database (GSE200972), and four tumor tissue samples were selected for analysis. The raw data contained 23,008 genes and 8,452 cells. The PercentageFeatureSet function was used to calculate the percentage of mitochondria and RNA, and cells with gene expression greater than 300 and less than 7000 and mitochondrial content less than 30% were selected as the basis for subsequent analysis.

2.2 Differential expression analysis of apoptosis-related genes

A series of bioinformatics analyses were performed using R programming language (version 4.2.1). The “limma” R package was used to screen for differentially expressed ARGs between tumor and non-tumor tissues in the TCGA cohort, with a false discovery rate (FDR) < 0.05 and |log2 (fold change)|> 1.

2.3 Development and validation of the prognostic model

In this study, an ARGs-related LUSC prognostic model was constructed to observe the OS of LUSC patients. The TCGA-LUSC cohort was randomly divided into a training set and a testing set, and the

prognostic features were constructed based on the training set, and then the predictive performance was validated in the validation set. Univariate Cox regression analysis was used to determine the prognostic ARGs in LUSC, and the results were visualized using a forest plot generated by the “forestplot” R package. The Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox proportional hazards regression (with R packages “glmnet”) was used to reduce the genes of the model and limit the complexity of solving the problem of overfitting. Finally, four prognostic ARGs were determined by multivariate Cox regression analysis, and a 4-gene signature was constructed. In the training and testing sets, the risk score of each patient was calculated based on the regression coefficient of the ARGs. The risk score was calculated using the formula: risk score = expression level of Gene a × coefficient a + expression level of Gene b × coefficient b + expression level of Gene c × coefficient c + …… + expression level of Gene n × coefficient n. LUSC patients were divided into high-risk and low-risk groups based on the median risk value, and Kaplan-Meier survival curves were used to compare the survival between the low-risk and high-risk groups, with $P < 0.05$ considered statistically significant. Finally, time-dependent receiver operating characteristic curves were established to reflect the predictive performance of the prognostic model.

2.4 Nomogram development and validation for prognostic risk prediction

To provide a quantitative method for predicting the 3-year, 5-year, and 7-year survival probabilities of LUAD patients for clinical doctors, we developed a nomogram integrating various clinical risk factors and apoptosis prognostic models. Then, we evaluated the calibration curve of the nomogram by plotting the predicted probability of the observed rate using the nomogram. Subsequently, 1000 repeated samplings were used for cross-validation to construct a C-index index plot to show the predictive ability of the model. The prognostic accuracy of the model was evaluated by sensitivity and specificity using the receiver operating characteristic (ROC) curve. Finally, a heatmap was constructed to evaluate the correlation between prognostic features and clinical factors.

2.5 Functional Enrichment Analysis

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the DAVID (https://david.ncifcrf.gov/). Firstly, we obtained pathway enrichment information by inputting ARGs into DAVID, and selected the top ten pathways with the smallest p-values for enrichment analysis in both GO and KEGG. The enriched pathways were displayed using the “ggplot” R package.

2.6 Gene Mutation Analysis

Tumor mutational burden (TMB) reflects the number of mutations in cancer mutation. We downloaded mutation data of LUSC patients from TCGA and performed gene mutation analysis using the “maftools” R package. We calculated the tumor mutation burden (TMB) of each patient and compared the TMB between the high-risk and low-risk groups. Survival analysis was performed based on the TMB score. The somatic mutations of prognostic genes were displayed using the cBioPortal database.

2.7 Relationship of Molecular Patterns With TME in LUSC

The ESTIMATE algorithm was used to evaluate whether the risk score of LUSC patients was associated with immune and stromal components. The single-sample gene set enrichment analysis (ssGSEA) was applied to quantify the infiltration levels of 16 immune cells and 13 immune-related pathways between the two risk groups using the “GSVA” package. Immune checkpoint genes with differential expression (with $P < 0.05$ as the threshold) were selected, and the expression differences of immune checkpoint genes between the high-risk and low-risk groups were studied.

2.8 Processing of single cell sequencing data

The scRNA-seq dataset GSE200972 of NSCLC was downloaded from the GEO database, containing 19 tissue samples from four MPLC patients (one squamous carcinoma and three adenocarcinomas). We selected four tumor samples for subsequent analysis. The samples were integrated using the anchors method in the R package "Seurat". Cells that detected genes in only three or fewer cells and cells that detected fewer than 300 genes were excluded, while cells with gene numbers between 300 and 7000, mitochondrial proportions less than 10%, blood cell proportions less than 1%, and total transcript numbers less than 100,000 were retained. The FindVariableFeatures function was used to identify the top 3000 highly variable genes, and the top 10 genes were displayed. Principal component analysis (PCA) was performed on single-cell samples, and the top 13 principal components (PC) were selected for subsequent analysis. The samples were subjected to overall dimensionality reduction analysis using the UMAP algorithm. Different cell clusters were manually annotated using the R package “singleR” package, CellMarker database, and references as auxiliary annotations.

2.9 Construction of Pseudo-Temporal Analysis Trajectory

Pseudo-temporal analysis was performed using the Monocle 2 algorithm through branching trajectory analysis. Gene expression levels and dispersion were filtered and sorted to exclude noise and irrelevant genes in the single-cell data. The orderCells function was used to sort cells and determine their position on the trajectory. Trajectories were constructed for two clusters, and differentiation trajectories of cells in different states were plotted. Subsequently, the expression of prognostic genes in the two clusters was visualized and the pseudo-temporal trajectories of several prognostic genes in T cells were observed.

2.10 Statistical Analysis

All statistical analyses were performed using R software (Version 4.2.1). A p value < 0.05 was considered statistically significant. Univariate Cox analyses, LASSO regression analysis and multivariate Cox analyses were used to select key ARGs associated with survival. Survival variations between different groups were compared using Kaplan-Meier curves. The predictive ability of this model was tested using ROC analysis.

3 Results

3.1 Identification of the differentially expressed genes in LUSC

We obtained RNA sequencing data and clinical follow-up data from 502 LUSC samples and 51 normal lung tissue samples from the TCGA dataset. Differential expression analysis was performed to determine the expression levels of ARGs in tumor and normal samples, and 56 differentially expressed genes were identified (Supplementary Table S1), most of which were enriched in tumor samples(Figure 1A ,B). Univariate Cox regression was used to screen ARGs, and 8 prognostic-related ARGs were identified Figure 1C . LASSO Cox regression analysis and 10-fold cross-validation were used to determine the optimal regularization parameter λ to obtain the best model performance(Fig. 1D, E). Finally, multivariate analysis was performed to identify 4 ARGs (BMP2, GPX3, JUN, and AIFM3) that were most relevant to the prognosis of LUSC patients. Kaplan-Meier analysis was used to analyze the impact of high and low expression of these 4 genes on prognosis, and the 3 most significant genes (BMP2, GPX3, and JUN) were identified (Figure S1A, B, C). We found that patients with high expression of BMP2, GPX3, and JUN had a worse prognosis than those with low expression. Correlation analysis revealed a close relationship between these genes (Figure S1D, E, F), suggesting a possible synergistic effect on prognosis. The high expression of BMP2, GPX3, and JUN proteins in cancer cells was confirmed through immunohistochemistry (IHC) data obtained from the Human Protein Atlas (HPA) database (Figure S1G).

3.2 Construction of a riskscore based on prognostic ARGs

The TCGA cohort was randomly divided into a training set and a testing set, and based on the median risk cutoff value of patients, they were further divided into high-risk and low-risk groups (Fig. 2A ,B). The risk score was calculated as follows: risk score= (− 0.411× expression level of AIFM3)+(0.084× expression level of BMP2) + (0.0838× expression level of GPX3)+(0.258× expression level of JUN). The patients in the training and testing sets were sorted from left to right based on their increasing risk scores. The results showed that as the risk score increased, the risk of death increased and the survival time decreased. The trend of survival status and survival time in the testing set was consistent with that in the training set, and the decrease in survival time was significantly correlated with the increase in risk score (Fig. 2C, D). Similarly, Kaplan-Meier survival curves showed that patients in the high-risk group had a lower survival rate than those in the low-risk group (Fig. 2E,F). The predictive performance of the prognostic risk score model was evaluated using time-dependent ROC curves and the area under the curve (AUC) (Fig. 2G ,H). The results showed that the risk score model had good predictive performance for LUSC patients, as confirmed by mutual verification between the training and testing sets.

3.3 Screening of independent prognostic factors and construction of nomogram

To identify independent prognostic factors, univariate and multivariate Cox analyses were performed on clinical features and risk scores. We found that RiskScore and tumor stage were independent prognostic

factors for patients, and a nomogram model was constructed to improve its predictive ability (Fig. 3A). By constructing calibration curves, it was found that the calibration curves of the 3-year, 5-year, and 7-year calibration points were relatively well matched with the standard curves, indicating that the model had good predictive performance (Fig. 3B). The time-dependent ROC curves and C-index indicated that the nomogram based on the independent prognostic factors had better prognostic predictive performance for LUSC compared to single clinical features (Fig. 3C ,D). Finally, we linked the risk score with clinical factors such as tumor stage, and the results showed that patients with high risk scores had shorter survival times (Fig. 3E). In conclusion, our study results indicate that the risk score of prognostic genes is of significant in predicting patient prognosis.

3.4 Functional analysis of differentially expressed ARGs

To explore the biological functions and important pathways of differentially expressed ARGs, we performed GO enrichment analysis to enrich the functions and pathways related to BP, CC, and MF, and selected the top 10 pathways with the smallest p-values for display. We found that regulation of apoptotic process was significantly enriched in BP (Fig. 4A), cytoplasmic components such as mitochondrial outer membrane were significantly enriched in CC (Fig. 4B), and protein complexes such as cysteine-type endopeptidase were significantly enriched in MF (Fig. 4C). Similarly, KEGG enrichment analysis was performed using the same GO enrichment analysis method, and other pathways such as the apoptosis pathway and cancer pathway were enriched (Fig. 4D). The detailed results of GO and KEGG enrichment analysis can be found in (Supplementary Table S2).

3.5 Tumor mutation burden of prognostic biomarkers associated with ARGs in LUSC samples

High TMB indicates that there are more mutations in tumor cells, which may make them more susceptible to attack by the immune system, thereby increasing the effectiveness of immunotherapy. To investigate the significance of tumor mutations in LUSC, we downloaded LUSC somatic mutation data and calculated TMB scores, comparing the genomic mutation differences between the high-risk and low-risk groups. The mutation rates in TP53, TTN, CSMD3, MUC16, and RYR2 were higher than or equal to 35% in LUSC patients in both risk groups. Interestingly, the likelihood of these gene mutations was greater in the low-risk group compared to the high-risk group (Fig. 5A ,B). To investigate the impact of TMB status on the prognosis of LUSC patients, survival analysis was performed on different TMB subgroups, and it was found that the prognosis of high TMB patients was better than that of low TMB patients (Fig. 5C). Then, we combined TMB with risk scores for survival analysis, and the results showed that patients with high TMB and low risk scores had the best prognosis (Fig. 5D). Finally, we examined the mutation rates of the four genes in the prognostic model and found that a segment of DNA sequence in the AIFM3 gene was duplicated, leading to an increase in copy number, which may contribute to the occurrence and development of tumors and an increase in the expression of other genes. However, the mutation rates of these genes were all very low (Fig. 5E).

3.6 Assessment of tumor immune infiltration and immune checkpoint

By analyzing the correlation between risk scores and tumor stromal and immune scores, the results showed that as the risk score increased, the abundance of tumor stroma and immune infiltration also increased (Fig. 6A ,B). This indicates a positive correlation between risk scores and the degree of tumor stromal and immune cell infiltration. Therefore, an increase in risk score may lead to changes in the tumor microenvironment and immune response, promoting tumor stromal and immune cell infiltration. To further understand the characteristics of the tumor microenvironment and its impact on tumor treatment, samples were divided into high-risk and low-risk groups, and the differences in TME score and tumor microenvironment evaluation indicators between different risk groups were compared. We found that both the StromalScore and ImmuneScore in the high-risk group were higher than those in the low-risk group (Fig. 6C). Previous studies have shown that the immune microenvironment has a significant impact on various aspects of tumor occurrence and development, treatment response, and the identification of new therapeutic targets (17, 18). Furthermore, our study showed that as the risk score increased, the abundance of immune infiltration also increased. We used the single-sample gene set enrichment analysis (ssGSEA) algorithm to demonstrate the enrichment levels of different immune cells in the tumor microenvironment of high- and low-risk groups, in order to understand the mechanisms of tumor immune escape and potential targets for immunotherapy. We found that compared to the low-risk group, the high-risk group had higher immune cell infiltration and more immune-related functions and pathways (Fig. 6D ,E). Next, we evaluated the expression differences of immune checkpoints between the high- and low-risk groups (Fig. 6F). The results showed that there were differences in the expression of 35 immune checkpoints between the two risk subgroups, such as TNFRSF4, CTLA4, and CD200.

3.7 Quality control and filtering of scRNA-seq data

Firstly, we filtered out the unqualified cells and used the filtered cells for subsequent analysis (Fig. 7A). Then, we performed principal component analysis on the four single-cell samples and found that the four non-small cell lung cancer samples had a high degree of aggregation (Fig. 7B), which may indicate that the transcriptional differences between these four samples are relatively small. We then conducted a correlation analysis of several quality control indicators and found that the proportion of mitochondria and hemoglobin did not increase with the increase of RNA molecules, while the number of transcriptional genes increased with the increase of RNA molecules (Fig. 7C), indicating that our quality control was relatively good and removed low-quality cells. Subsequently, we used differential analysis to screen out 3000 highly variable genes (Fig. 7D).

3.8 Identification and Localization of LUSC Cell Subtypes

We used the UMAP algorithm to divide the core cells into 16 independent cell clusters (Fig. 8A). In order to explore the heterogeneity within the tumor, we calculated the proportion of each cell in the four samples

separately (Fig. 8B,C). The results showed that the proportions of different cells were different in each sample, indicating the presence of different cell subpopulations within the tumor, which is consistent with the characteristics of tumor heterogeneity. We identified marker genes using the “singleR” package, CellMarker database, and references (19). We annotated the different clusters (Fig. 8D). Subsequently, we displayed the upregulated, significantly upregulated, downregulated, and significantly downregulated genes in the marker genes of the 16 clusters (Fig. 9A). We found differences in gene expression between different clusters, which can be further explored for their roles in cellular and biological processes. Similarly, we displayed the marker genes of the annotated eight cell types (Fig. 9B) to understand the differences in their functions and characteristics. We located four apoptosis-related prognostic genes associated with LUSC to individual cell subtypes to understand the biological characteristics and functions of these cell subtypes. Furthermore, this localization can provide clues for further research, such as exploring the interactions and regulatory mechanisms of these prognostic genes in different cell subtypes and their relationships with other biological processes. This localization of cell subtypes can help us better understand the mechanisms of LUSC occurrence and development. We observed the distribution of several apoptosis-related prognostic genes in these eight cell clusters (Fig. 9C). It can be seen that BMP2 is significantly upregulated in epithelial cells, GPX3 is expressed at a higher level in smooth muscle cells, and JUN is expressed in all cell subtypes, but relatively higher in T cells. To further understand the specific expression of prognostic genes in the eight cell clusters, we displayed the gene expression by drawing a heatmap (Fig. 9D).

3.9 Identifying lung cancer subpopulations through single-cell trajectory analysis

Temporal analysis aims to model the time information of single cells, revealing the temporal trends and expression profile differences of different cells, and exploring biological issues such as cell development, differentiation, and function, helping us to gain a deeper understanding of single-cell data, discover new biological knowledge and biomarkers. (20, 21). We selected two clusters for temporal analysis, and the trajectory displayed the transcriptional states of cell development in the two clusters (Fig. 10A). We also displayed the trajectory of the state of cluster cells (Fig. 10B) and determined the starting point of cell differentiation (Fig. 10C), revealing changes related to tumor progression. Next, we studied the expression changes of four prognostic genes in the temporal analysis (Fig. 10D). JUN was significantly decreased with cancer progression, while AIFM3, BMP2, and GPX3 did not show significant changes. As we know from the previous dimensionality reduction clustering, T cells account for a large proportion in the sample, so we studied the expression changes of four prognostic genes in the temporal analysis of T cells (Fig. 10E).

4 Discussion

Lung squamous cell carcinoma (LUSC) is a histological subtype of non-small cell lung cancer (NSCLC) with poor prognosis, high treatment difficulty, and high mortality rate mainly due to the lack of effective

treatment (22). Despite the available treatment methods for lung squamous cell carcinoma, including chemotherapy, radiotherapy, and targeted therapy, the prognosis for patients with lung squamous cell carcinoma remains unsatisfactory (23). Therefore, it is urgently needed to study new molecular therapeutic targets and prognostic models for the diagnosis and treatment of LUSC. Previous studies have shown that apoptosis plays a crucial role in maintaining the balance between cell death and division, and evading apoptosis can lead to uncontrolled cell proliferation, resulting in various diseases such as cancer (24). The mechanism of apoptosis is complex and involves many pathways. Any defect in any of these pathways can lead to malignant transformation of affected cells, tumor metastasis, and resistance to anticancer drugs. Currently, many new therapeutic strategies targeting apoptosis are feasible (25, 26). Although there is a connection between apoptosis and LUSC, there has not been a systematic study that uses apoptosis-related features as prognostic indicators to predict the prognosis of LUSC patients.

Our study conducted a systematic analysis based on the TCGA-LUSC dataset and apoptosis-related genes to identify differentially expressed ARGs in LUSC and non-tumor tissues. Through multifactor and LASSO Cox regression analysis, we screened for prognostic-related genes and established a risk model to predict the prognosis of LUSC. The genes in the model, including BMP2, GPX3, and JUN, were validated to be associated with the prognosis of LUSC patients through survival analysis and the HPA database.

Bone morphogenetic proteins (BMPs) are multifunctional cytokines, belonging to members of the transforming growth factor-βsuper-family(27). MP can trigger the occurrence and progression of tumors through the signaling mediators of ligands and receptors. At the same time, BMP can promote cell differentiation, including inhibiting the process of epithelial-mesenchymal transition, which can prevent the malignant progression of cancer in the later stages (28). BMP can activate ERK, phosphoinositide 3-kinase (PI3K), protein kinase A (PKA), PKC, and PKD. Through the initiation of these pathways, BMP can induce its effects on cell survival, apoptosis, migration, and differentiation (29). SMAD is the classical pathway mediated by BMP2. Studies have shown that BMP2 is highly overexpressed in human non-small cell lung cancer and is associated with tumor grading and metastasis. Through mouse models, BMP2 has been shown to promote lung cancer metastasis. Depletion of BMP2 or its receptor BMPR2 greatly reduces cell migration and invasiveness, and BMP2/BMPR2-mediated cell migration involves the activation of the SMAD1/5/8 signaling pathway. Depletion of SMAD1/5/8 significantly reduces cell migration by inhibiting SMAD1/5/8 (30). Epigenetic silencing of glutathione peroxidase 3 (GPX3), a member of the important antioxidant selenoprotein family (31), maintains genome integrity by inactivating Reactive oxygen species (ROS) (32). Reactive oxygen species (ROS) are believed to play various roles in cancer development. For example, when the gene expression of important molecules that control cell proliferation, apoptosis, or the cell cycle is abnormal, oxidative stress can lead to persistent DNA damage and may induce cancer (33). Research has shown that GPX3 is associated with ovarian cancer metastasis and cancer progression. In this study, a stable OVCAR3 GPx3 knockdown cell line was constructed using a lentiviral shRNA, and it was found that reducing GPx3 expression inhibited colony formation and anchorage-independent cell survival (34). In a study on thyroid cancer, methylation-specific PCR (MSP), immunohistochemistry staining, Transwell experiments, and siRNA knockdown were used to

investigate the role of GPX3. It was found that GPX3 inhibits thyroid cancer metastasis by suppressing the Wnt/β-catenin signaling pathway. Silencing GPX3 expression promotes human thyroid cancer metastasis. (35). c-Jun has been found to be an oncogenic transcription factor in most cancers, and its overexpression plays an important role in various biological functions such as cell apoptosis, proliferation, invasion, and migration (36, 37). A study showed that enforced expression of c-Jun increased anchorage-independent growth of human bronchial epithelial cell lines, and constitutive expression of a significant c-Jun-negative mutant suppressed anchorage-independent but not anchorage-dependent growth of lung cancer cell lines (38). The activity of c-Jun is regulated by post-translational modifications, which are mainly controlled by components of the mitogen-activated protein kinase (MAPK) family of kinases, including c-Jun N-terminal kinase (JNK), extracellular signal-regulated kinase (ERK), and p38 kinase (39). The significant and unique function of JNK is as an activator of c-Jun. Overexpression or activation of c-Jun has been shown to have anti-apoptotic effects in various cancer cell lines, and its targeting may sensitize drug-resistant cancer cells to DNA-damaging agents (40). Therefore, these key apoptotic prognostic-related genes are closely associated with LUSC and its prognosis, demonstrating the validity of establishing a prognostic model based on apoptotic prognostic-related genes in this study.

After modeling, we conducted validation by dividing LUSC patients into different risk groups based on the median risk score. Patients in the high-risk group had significantly worse prognosis. The results of ROC curve analysis showed that the prognostic model had good predictive performance. Furthermore, through analysis of the risk score of the model and other clinical features, the risk score of the model was found to be an independent prognostic indicator, and C-index analysis showed that the risk score had better predictive value than other traditional clinical parameters. Therefore, the good predictive value of the model was confirmed once again.

In addition, we conducted functional enrichment analysis on the apoptotic differentially expressed genes that we screened, exploring the key pathways through which they play their functions. According to GO and KEGG pathway analysis, we found that in cancer, cell apoptosis is mediated by changes in the mitochondrial membrane, which is a multifactorial process involving BCL-2 family proteins, cysteine proteases, and large molecular complexes. Studies have shown that in the pre-initiation stage of mitochondria, different pro-apoptotic signal transduction or damage pathways are activated. When these signals or pathways converge on the mitochondria, the permeability of the inner and outer membranes increases, leading to the execution stage of the apoptosis process (41). The regulation of apoptotic mitochondrial events is achieved through the Bcl-2 family of proteins. The Bcl-2 family consists of more than 30 proteins and belongs to the Bcl-2 superfamily, which includes anti-apoptotic proteins, pro-apoptotic proteins, and BH3-only proteins (42). In the presence of apoptotic stimuli, the expression of BH3 proteins increases, competitively binding to the anti-apoptotic protein Bcl-2 to release Bax/Bak from inhibition. Free Bax and Bak form oligomers, causing cytochrome C to form a channel through the outer membrane of the mitochondria, releasing it from the intermembrane space of the mitochondria into the cytoplasm. The released cytochrome C activates the caspase cascade reaction to induce cell apoptosis (43, 44). Therefore, mitochondria can be considered as the main integrator of the death pathway.

Apoptosis depends on the activation of the above different signaling pathways, and these pathways are often dysregulated in cancer, providing a direction for further research on the treatment of apoptosis in lung squamous cell carcinoma.

Furthermore, in our multifaceted study of the immune microenvironment, we found significant differences in TMB and TP5 between high- and low-risk groups. Studies have shown that TMB in blood can be used to evaluate the efficacy of using camrelizumab in combination with chemotherapy in advanced lung squamous cell carcinoma patients. During treatment, blood TMB levels are positively correlated with patient efficacy, indicating that higher TMB leads to better treatment efficacy and longer overall survival (OS) and progression-free survival (PFS) for patients. These findings are consistent with the results of our study (45). In this study, we can see the differences in gene mutations between high and low-risk groups, and the most common type of gene mutation, TP53, has significant differences in mutation types between high and low-risk groups. Previous studies have shown that TP53 is one of the most common mutated genes in lung cancer, and its mutation is closely related to the occurrence and development of lung cancer. The type of TP53 mutation is related to prognosis, and patients with nonsense mutations have a poorer prognosis (46). This is consistent with our research results, where patients in the high-risk group had a poorer prognosis.

Although previous studies have investigated the relationship between apoptosis and cancer, there have been few studies on its relationship with tumor immunity. Stromal cells and immune cells are the main elements of the TME, and an important aspect of our research is to explore the correlation between the risk model of LUSC patients and tumor immunity. We used the ESTIMATE algorithm to calculate these scores and found that the high-risk group had higher immune and stromal scores. This indicates that changes in immune status can also affect the process of cell apoptosis. We found that NK cells and macrophages were highly expressed in the high-risk group, and they have an inherent connection with cell apoptosis. Studies have shown that the mitochondrial apoptosis (mtApoptosis) pathway is crucial for efficient NK killing, and NK cells can pre-activate cancer cells for mtApoptosis. Pre-activated NK cells bind BH3 mimetics to NK cells, synergistically killing cancer cells in vitro and inhibiting tumor growth in vivo (47). This study suggests that mtApoptosis can enhance NK-based immunotherapy. An interesting study found that macrophages are a heterogeneous group of cells in the innate immune system that are crucial for the initiation, progression, and resolution of inflammation. They have significant functional plasticity and can respond to abnormalities and initiate programs to overcome them and restore normalcy (48, 49). The cytokines and intracellular components released by apoptotic cells can activate macrophages, causing them to shift from the M1 to M2 phenotype, thereby promoting tumor growth and metastasis. In addition, apoptotic cells can also induce macrophages to produce growth factors and extracellular matrix components, promoting tumor cell proliferation and invasion. If the interaction between apoptotic cells and macrophages is not disrupted, surviving tumor cells may receive excessive support from the reaction induced by local macrophages in apoptotic tumor cells, thereby enhancing tumor recurrence (50). Indeed, this provides valuable ideas and strategies for researchers to further explore the anti-tumor potential of cell apoptosis and tumor immunity.

Tumor heterogeneity is a significant challenge in cancer treatment, as different subpopulations of cells may exhibit varying sensitivity and resistance to therapeutic drugs (51). Therefore, understanding tumor heterogeneity is of great significance for developing personalized treatment plans and predicting treatment outcomes. Compared to traditional 'bulk' RNA-sequencing methods, which average potential differences in cell-specific transcriptomes, single-cell RNA sequencing (scRNA-seq) analyzes the gene expression patterns of each individual cell, providing a clear insight into the entire tumor ecosystem, such as the mechanisms of intra- and inter-tumor heterogeneity (52). Through single-cell sequencing analysis, this study ultimately divided cells into eight subgroups. To further understand the expression patterns and functions of apoptosis-related prognostic genes in different cell subgroups, we compared the expression of prognostic genes in different cell subgroups to determine their functions in different cell subgroups. We found that BMP2 was highly expressed in epithelial cells, which may indicate that BMP2 has important biological functions in epithelial cells. In a study on breast cancer, elevated levels of BMP2 led to excessive activation of the BMPR1B signaling pathway, which is the receptor for BMP2. When BMP2 binds to BMPR1B, it activates the BMPR1B signaling pathway, promoting the transformation of epithelial cells. These transformed epithelial cells have increased proliferation and invasiveness, thereby promoting tumor development (53). Similarly, in a study on lung injury, BMP2 was found to activate the BMP signaling pathway, leading to increased BMP activity and ultimately resulting in the transformation of epithelial cells and the disruption of epithelial barrier function. This study successfully inhibited the increase in BMP activity and the disruption of epithelial barrier function by suppressing the expression of BMP2, thereby avoiding lung injury (54). This provides a good explanation of the role of prognostic genes in tumor development at the cellular level.

There are different subpopulations of cells within tumors, which have different gene expression profiles, leading to heterogeneity within the tumor. Sequencing multiple tumor regions can reveal the evolutionary pattern of the tumor, where tumor cells have different gene mutations and expression profiles at different time and space points, forming a branched evolutionary tree structure (55). Pseudotime analysis is a method for inferring trajectories from scRNA-seq data. It sorts cells along a trajectory based on the similarity of their expression patterns and determines the lineage structure by identifying branching events (56). In this study, we determined trajectories with different differentiation states based on scRNA-seq data of non-small cell lung cancer, and located the selected prognostic genes on the cell differentiation trajectories. We observed the changes of prognostic genes in different clusters and the same cell types. Combining pseudotime analysis and the localization of cell types in different clusters over time, we were able to intuitively witness the evolution of cells in non-small cell lung cancer. This provides new insights into the mechanisms and driving factors of tumor evolution.

Although the prognostic model established in this study has good predictive performance and is a positive prognostic indicator for LUSC patients, and provides more accurate information for the treatment and prognosis evaluation of LUSC patients by combining single-cell sequencing analysis, there are still some limitations that need to be considered. The analysis conducted in this study was based on retrospective data from the TCGA and GEO databases, and due to the scarcity of scRNA-seq data in the GEO database, we could not obtain complete single-cell sequencing results for LUSC samples. Our results

need to be further functionally validated. Despite these limitations, the research in this paper can serve as a valuable concept validation study, identifying biomarkers and targets for future research and providing meaningful references for personalized treatment of LUSC patients.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Peiquan Zhu : Conception and design. Wenxing Yang and Kaiqiang Wang: provision of study materials. Biao Wang: collection and assembly of data. Zhi Hu: data analysis and interpretation. Dengguo Zhang and Ze Yang: project administration. Jiangtao Pu: revise the draft. All authors have read and agreed to the published version of the manuscript.

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics Statement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Acknowledgments

Not applicable.

References

1. Siegel RL et al (2022) Cancer statistics, 2022. Cancer J Clin 72:7–33
2. Bray F et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 68:394–424
3. Youlden DR et al (2008) The International Epidemiology of Lung Cancer: geographical distribution and secular trends. J Thorac oncology: official publication Int Association Study Lung Cancer

3:819–831
4. Socinski MA et al (2018) Current and Emergent Therapy Options for Advanced Squamous Cell Lung Cancer. J Thorac oncology: official publication Int Association Study Lung Cancer 13:165–183
5. Paz-Ares LG et al (2013) PARAMOUNT: Final overall survival results of the phase III study of maintenance pemetrexed versus placebo immediately after induction treatment with pemetrexed plus cisplatin for advanced nonsquamous non-small-cell lung cancer. J Clin oncology: official J Am Soc Clin Oncol 31:2895–2902
6. Woodard GA et al (2016) Lung Cancer Staging and Prognosis. Cancer Treat Res 170:47–75
7. Wang M et al (2021) Toward personalized treatment approaches for non-small-cell lung cancer. Nat Med 27:1345–1356
8. Shivapurkar N et al (2003) Apoptosis and lung cancer: A review. J Cell Biochem 88:885–898
9. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100:57–70
10. Liu G et al (2017) Role of Autophagy and Apoptosis in Non-Small-Cell Lung Cancer. Int J Mol Sci, 18
11. Gandara DR et al (2009) Evolving treatment algorithms for advanced non-small-cell lung cancer: 2009 looking toward 2012. Clin Lung Cancer 10:392–394
12. Pore MM et al (2013) Targeting apoptosis pathways in lung cancer. Cancer Lett 332:359–368
13. Olsen TK, Baryawno N (2018) Introduction to Single-Cell RNA Sequencing. Curr Protoc Mol Biol, 122
14. Lei Y et al (2021) Applications of single-cell sequencing in cancer research: progress and perspectives. J Hematol Oncol 14:91
15. Rozenblatt-Rosen O et al (2020) The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. Cell 181:236–249
16. Zheng L et al (2021) Six Novel Biomarkers for Diagnosis and Prognosis of Esophageal squamous cell carcinoma: validated by scRNA-seq and qPCR. J Cancer 12:899–911
17. Shiravand Y et al (2022) Immune Checkpoint Inhibitors in Cancer Therapy. Curr Oncol (Toronto Ont) 29:3044–3060
18. Thorsson V et al (2018) The Immune Landscape of Cancer. Immunity 48:812–830e814
19. Ma KY et al (2019) Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response-related genes. JCI insight, 4
20. La Manno G et al (2018) RNA velocity of single cells. Nature 560:494–498
21. Trapnell C et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32:381–386
22. Lau SCM et al (2022) Squamous cell lung cancer: Current landscape and future therapeutic options. Cancer Cell 40:1279–1293
23. Perez-Moreno P et al (2012) Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities. Clin cancer research: official J Am Association Cancer Res 18:2443–2451

24. Chaudhry GE et al (2022) Cancer and Apoptosis. Methods in molecular biology (Clifton N J) 2543:191–210
25. Jan R, Chaudhry GE (2019) Understanding Apoptosis and Apoptotic Pathways Targeted Cancer Therapeutics. Adv Pharm Bull 9:205–218
26. Wong RS (2011) Apoptosis in cancer: from pathogenesis to treatment. J experimental Clin cancer research: CR 30:87
27. Wakefield LM, Hill CS (2013) Beyond TGFβ: roles of other TGFβ superfamily members in cancer. Nat Rev Cancer 13:328–341
28. Davis H et al (2016) Mechanisms of action of bone morphogenetic proteins in cancer. Cytokine Growth Factor Rev 27:81–92
29. Bragdon B et al (2011) Bone morphogenetic proteins: a critical review. Cell Signal 23:609–620
30. Wu CK et al (2022) BMP2 promotes lung adenocarcinoma metastasis through BMP receptor 2-mediated SMAD1/5 activation. Sci Rep 12:16310
31. Chang C et al (2020) Extracellular Glutathione Peroxidase GPx3 and Its Role in Cancer. Cancers, 12
32. Chen B et al (2011) GPx3 promoter hypermethylation is a frequent event in human cancer and is associated with tumorigenesis and chemotherapy response. Cancer Lett 309:37–45
33. Ushio-Fukai M (2009) Compartmentalization of redox signaling through NADPH oxidase-derived ROS. Antioxid Redox Signal 11:1289–1299
34. Worley BL et al (2019) GPx3 supports ovarian cancer progression by manipulating the extracellular redox environment. Redox Biol 25:101051
35. Zhao H et al (2015) Silencing GPX3 Expression Promotes Tumor Metastasis in Human Thyroid Cancer. Curr Protein Pept Sci 16:316–321
36. Katiyar S et al (2007) Somatic excision demonstrates that c-Jun induces cellular migration and invasion through induction of stem cell factor. Mol Cell Biol 27:1356–1369
37. Zhao Y et al (2016) Inhibitor of Differentiation/DNA Binding 1 (ID1) Inhibits Etoposide-induced Apoptosis in a c-Jun/c-Fos-dependent Manner. J Biol Chem 291:6831–6842
38. Maeno K et al (2006) Altered regulation of c-jun and its involvement in anchorage-independent growth of human lung cancers. Oncogene 25:271–277
39. Lopez-Bergami P et al (2010) Emerging roles of ATF2 and the dynamic AP1 network in cancer. Nat Rev Cancer 10:65–76
40. Vasilevskaya I, O'dwyer PJ (2003) Role of Jun and Jun kinase in resistance of cancer cells to therapy. Drug Resist updates: reviews commentaries Antimicrob anticancer Chemother 6:147–156
41. Jacotot E et al (1999) Mitochondrial membrane permeabilization during the apoptotic process. Ann N Y Acad Sci 887:18–30
42. Cory S, Adams JM (2002) The Bcl2 family: regulators of the cellular life-or-death switch. Nat Rev Cancer 2:647–656

43. Hongmei Z (2012) Extrinsic and Intrinsic Apoptosis Signal Pathway Review [M]. Apoptosis and Medicine.
44. Xiong S et al (2014) Mitochondria-mediated apoptosis in mammals. Protein Cell 5:737–749
45. Jiang T et al (2022) On-treatment blood TMB as predictors for camrelizumab plus chemotherapy in advanced lung squamous cell carcinoma: biomarker analysis of a phase III trial. Mol Cancer 21:4
46. Fan Z et al (2022) Genomic landscape and prognosis of patients with TP53-mutated non-small cell lung cancer. Annals of translational medicine 10:188
47. Pan R et al (2022) Augmenting NK cell-based immunotherapy by targeting mitochondrial apoptosis. Cell 185:1521–1538e1518
48. Lavin Y, Merad M (2013) Macrophages: gatekeepers of tissue integrity. Cancer Immunol Res 1:201–209
49. Wynn TA et al (2013) Macrophage biology in development, homeostasis and disease. Nature 496:445–455
50. Weigert A et al (2016) Killing Is Not Enough: How Apoptosis Hijacks Tumor-Associated Macrophages to Promote Cancer Progression. Adv Exp Med Biol 930:205–239
51. Dagogo-Jack I, Shaw AT (2018) Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol 15:81–94
52. Baslan T, Hicks J (2017) Unravelling biology and shifting paradigms in cancer with single-cell sequencing. Nat Rev Cancer 17:557–569
53. Chapellier M et al (2015) Disequilibrium of BMP2 levels in the breast stem cell niche launches epithelial transformation by overamplifying BMPR1B cell response. Stem cell reports 4:239–254
54. Helbing T et al (2013) Inhibition of BMP activity protects epithelial barrier function in lung injury. J Pathol 231:105–116
55. Gerlinger M et al (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366:883–892
56. Wu F et al (2021) Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. Nat Commun 12:2540

Figures

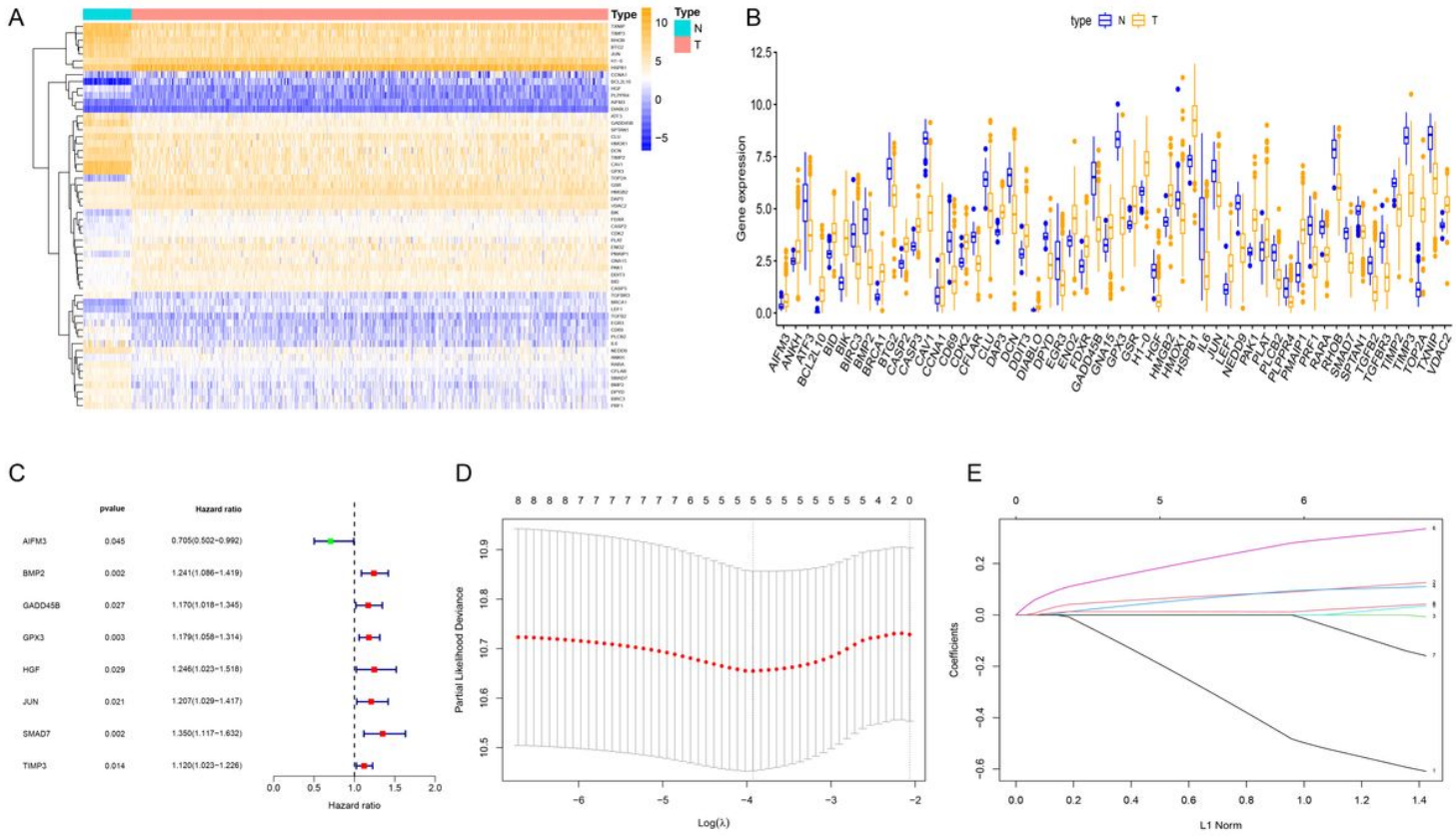

Figure 1

The differential expression of apoptosis-related genes in LUSC was analyzed, and genes that are associated with the prognosis of LUSC were screened using regression algorithms. A Heat maps of 56 differentially expressed ARGs. B Boxplot of 56 differentially expressed ARGs in LUSC and normal lung tissue. (C) Forest maps of 8 prognostic ARGs were obtained by univariate Cox analysis. (D) Ten-time cross-validation for tuning parameter selection in the least absolute shrinkage and selection operator (LASSO) model. E Distribution plot of gene coefficients generated with log (λ) in the LASSO model.

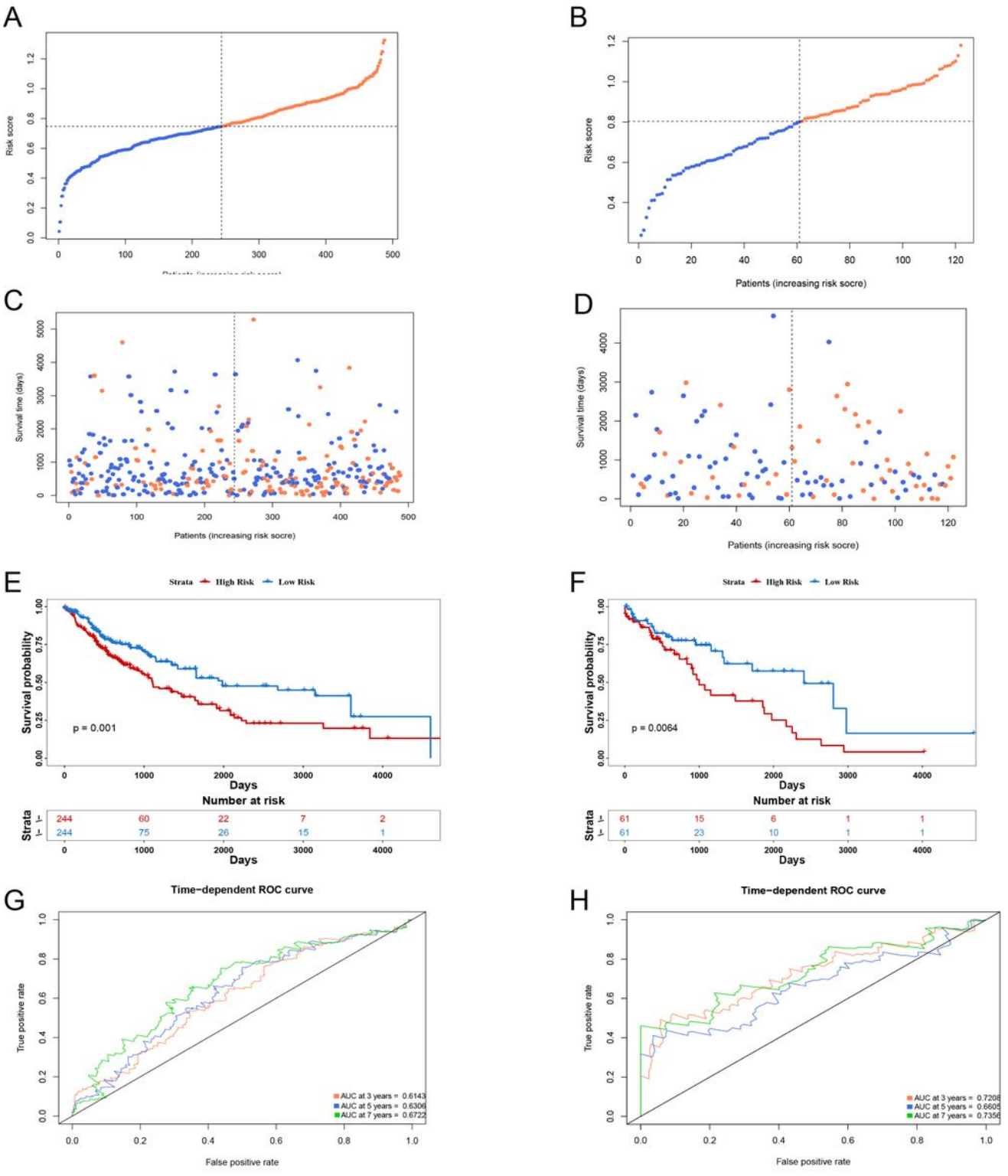

Figure 2

Construction of a risk model based on apoptosis-related genes associated with prognosis in lung squamous cell carcinoma (LUSC). (A-B) The distribution of risk scores for patients in the training and testing groups. (C-D) The survival status of patients in the training and testing groups. E-F Kaplan-Meier analysis of the training and testing groups showing different overall survival (OS) between high and low-risk groups. G-H TimeROC curve predicting overall survival rate for the training and testing groups.

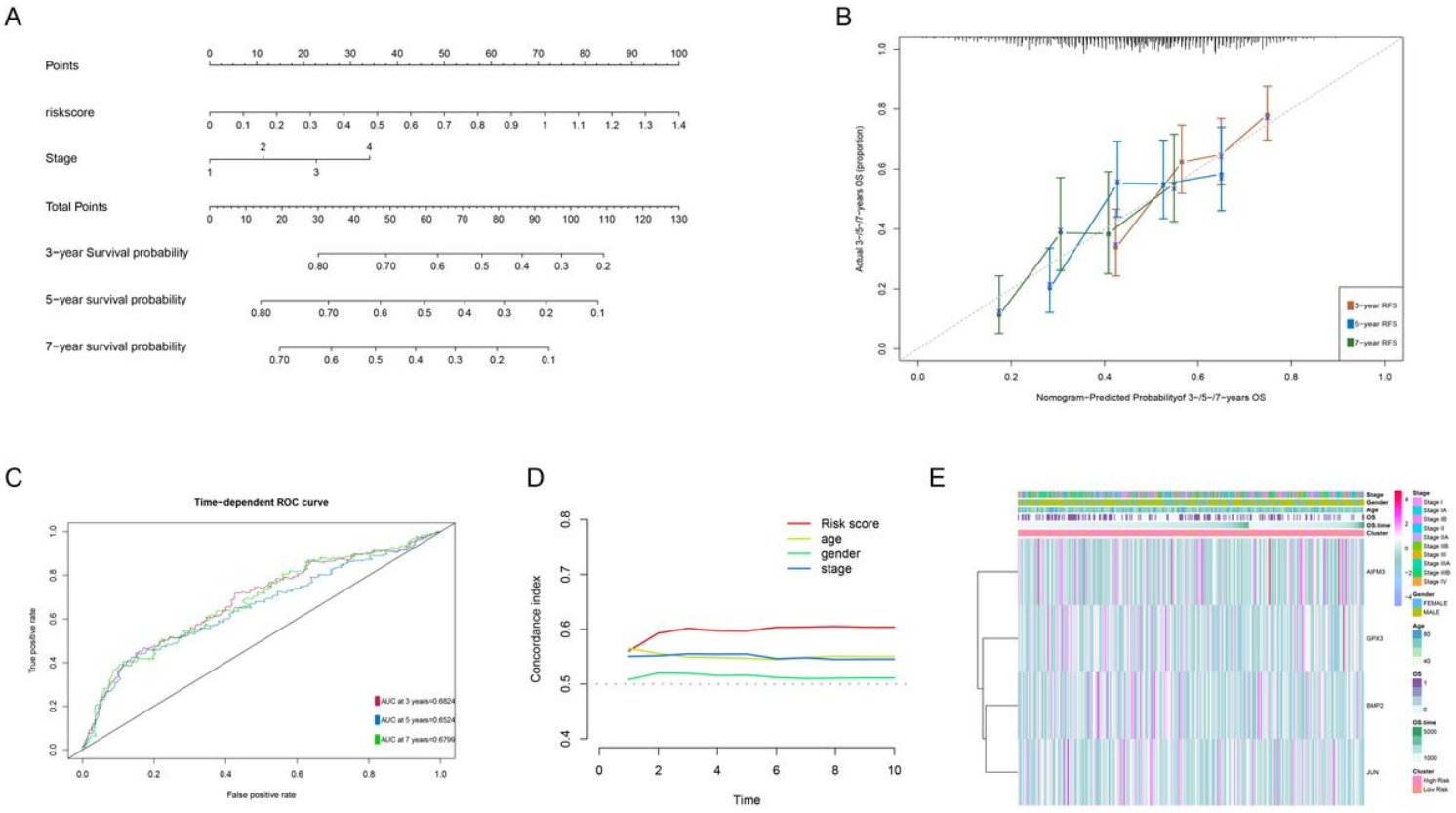

Figure 3

Establishment and validation of the nomogram prediction model.

(A) Nomogram combining risk score and other clinical factors to predict 3-year, 5-year, and 7-year overall survival (OS) of TCGA-LUSC patients. B Calibration plot of the nomogram predicting 3-year, 5-year, and 7-year survival probabilities. (C) TimeROC curve predicting 3-year, 5-year, and 7-year survival of LUSC patients. (D) The C-index indicates that the predictive accuracy of the risk model is superior to other clinical parameters. (E) Heatmap of 4 prognostic ARGs combined with different clinical features.

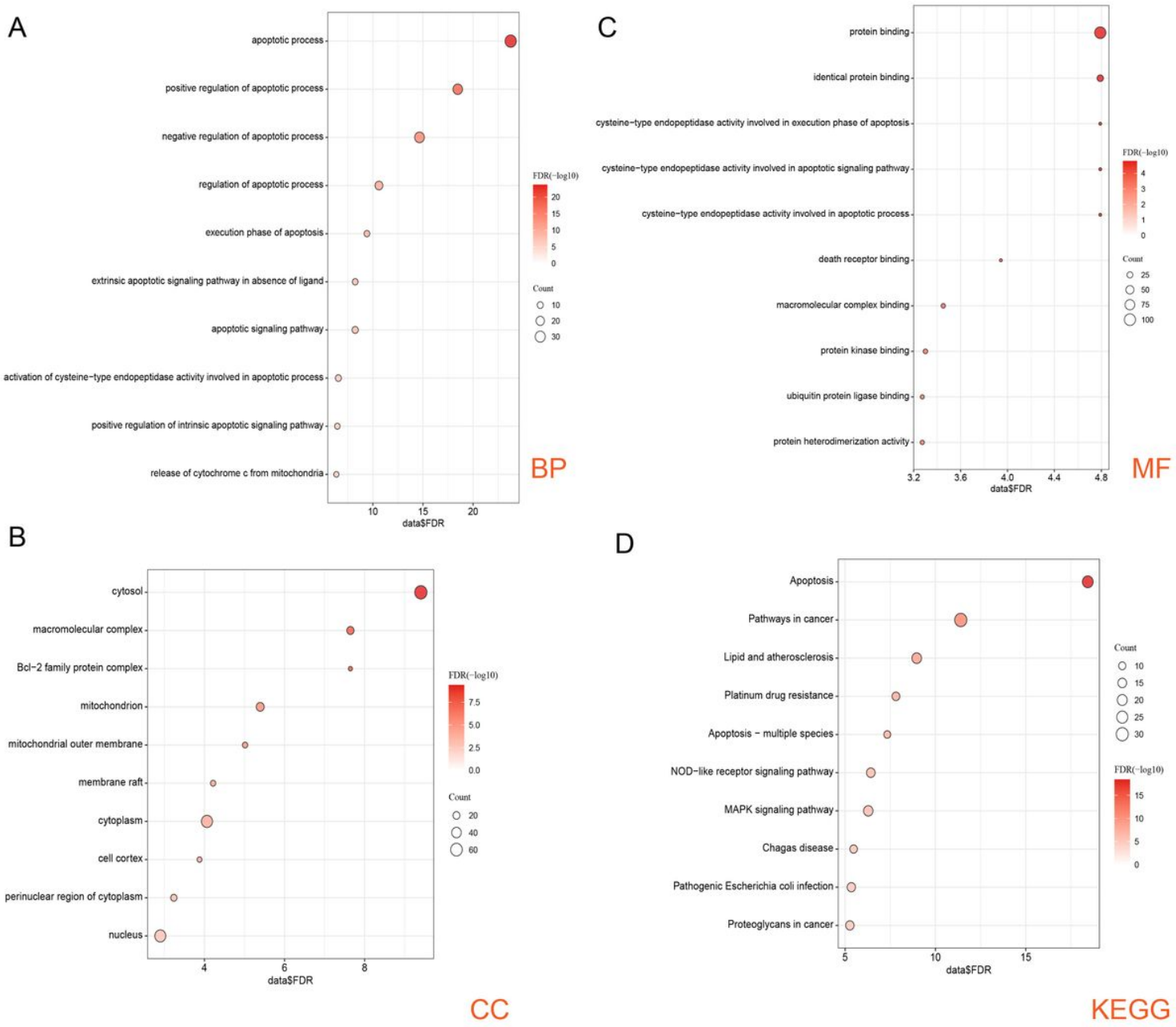

Figure 4

Function analysis of differentially expressed ARGs.

A-C GO functional analysis of differentially expressed ARGs. (BP: biological processes, CC: cellular components, MF: molecular function). (D) KEGG pathway analysis of differentially expressed ARGs.

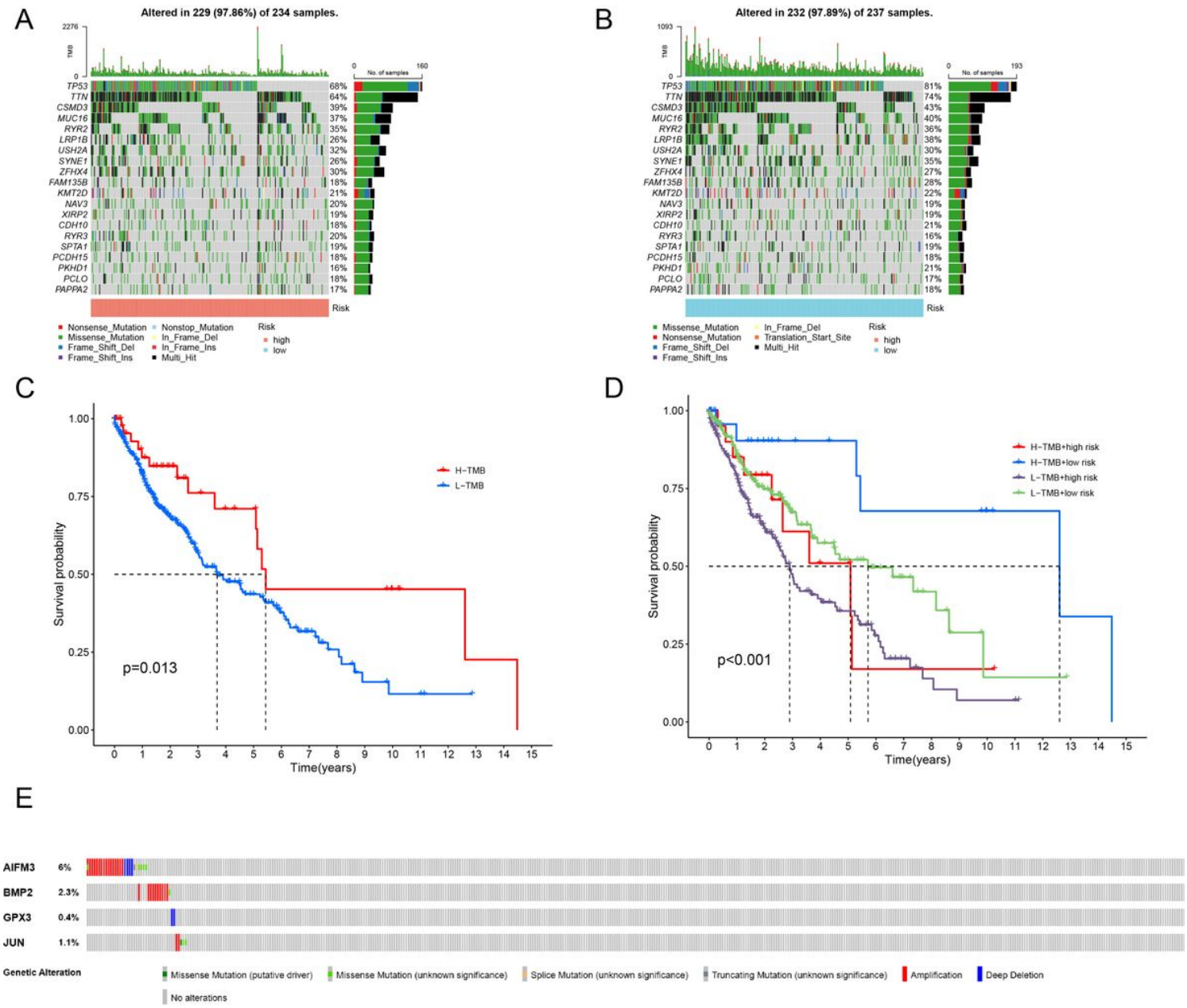

Figure 5

Waterfall plot of high-risk and low-risk groups and prognostic gene somatic mutations.

A-B Waterfall plot of somatic mutations in high-risk and low-risk groups. (C) Survival curves of high TMB group and low TMB group. (D) Survival curves based on TMB and risk score. (E) Study of mutation rates of 4 prognostic genes using the cBioPortal database.

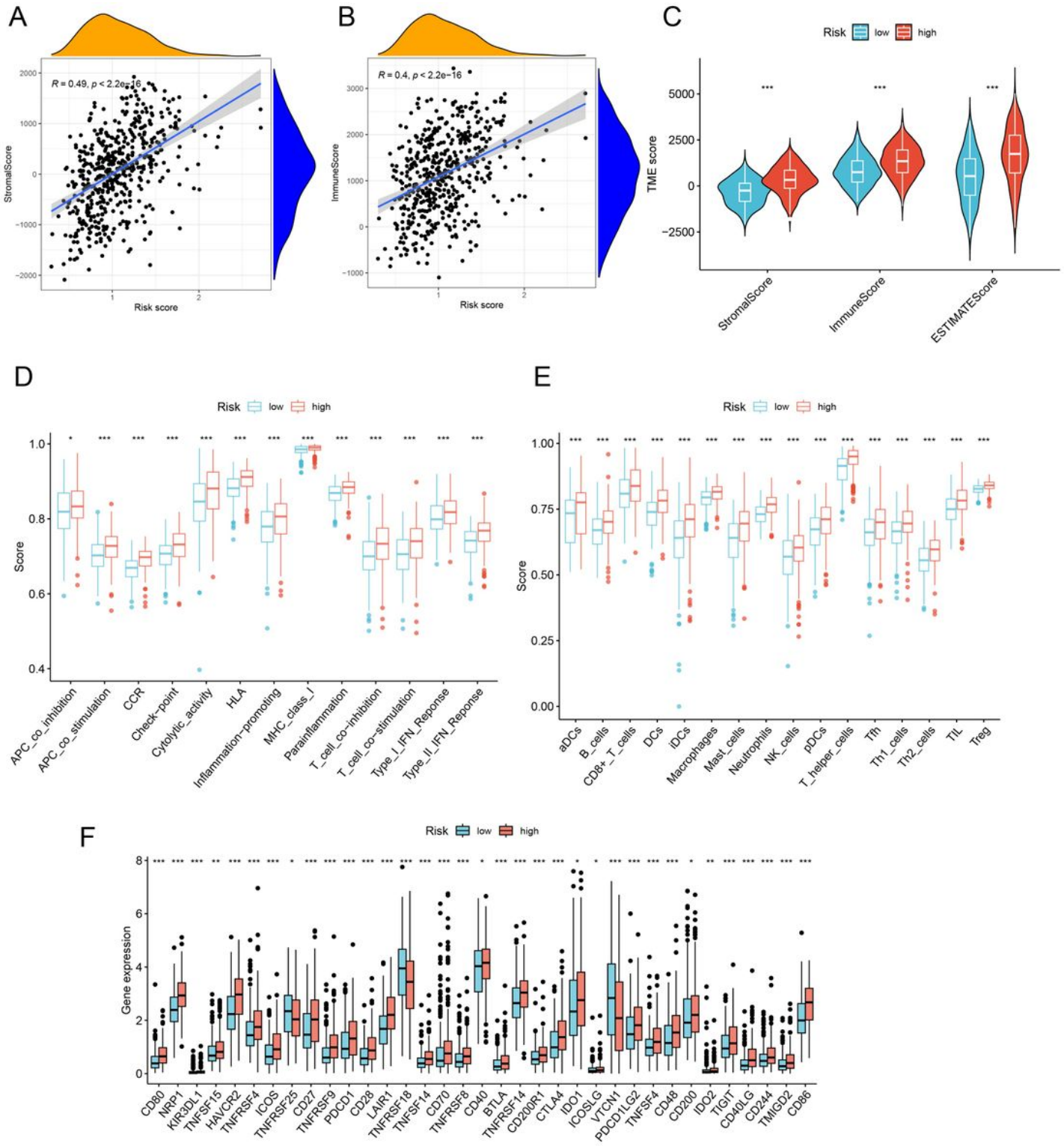

Figure 6

Analysis of TME and immune-related characteristics in high-risk and low-risk groups. A Correlation between risk score and stromal score. (B) Correlation between risk score and immune score. (C) Evaluation of tumor microenvironment using TME score. (D) Box plots showing the scores of various immune-related functions in high-risk and low-risk groups. (E) Box plots showing the scores of multiple

immune cells in high-risk and low-risk groups. (F)Expression of immune checkpoints in the high and low-risk groups. ($p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***).

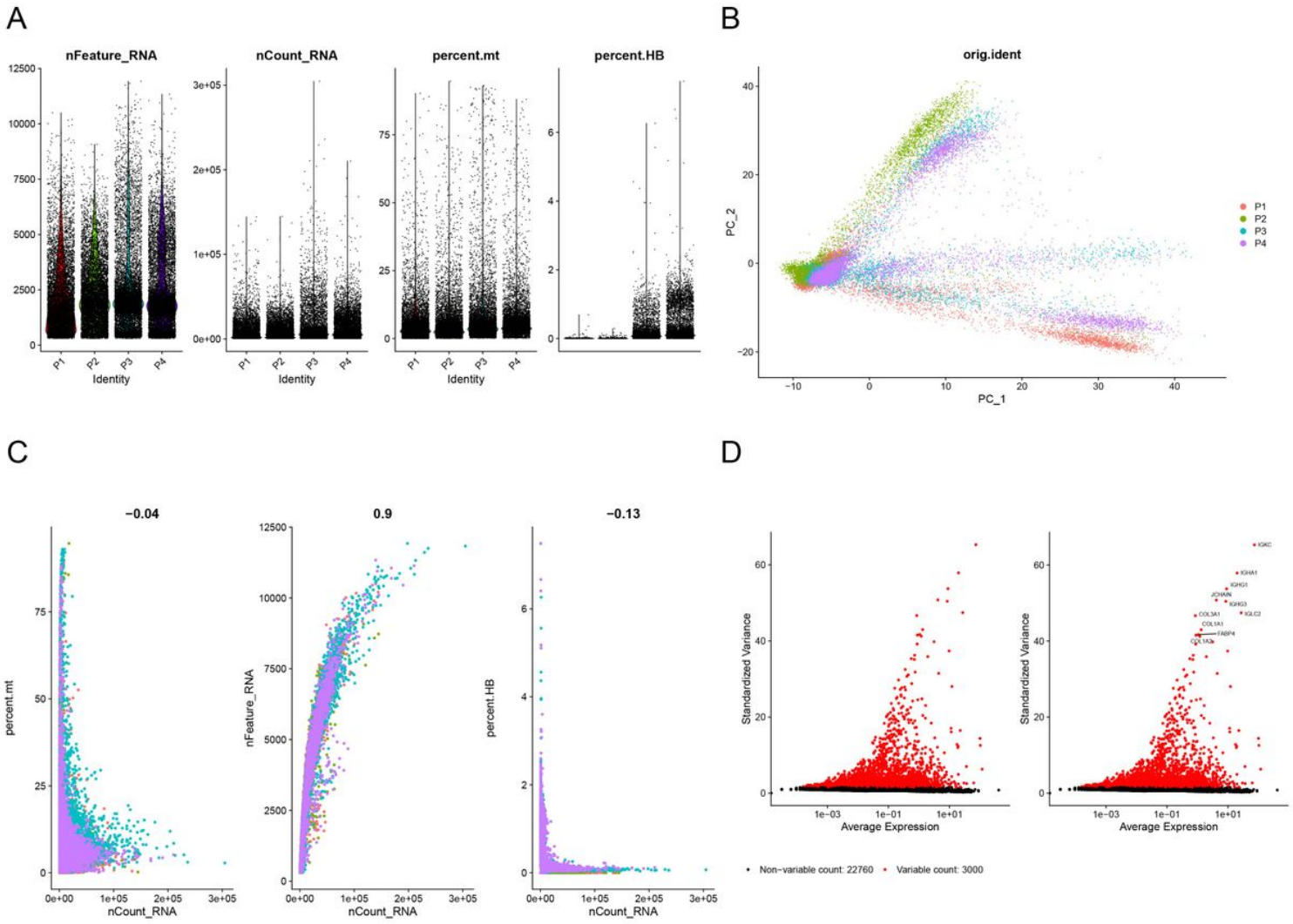

Figure 7

Quality control and filtering of scRNA-seq data. A The number of genes detected, sequencing depth, mitochondrial gene content, and HB content in four non-small cell lung cancer samples. (B)PCA-based preliminary dimensionality reduction of the data from four samples. (C) There is a negative correlation (R= -0.04) between sequencing depth and mitochondrial gene content, a positive correlation (R=0.9) between sequencing depth and the number of intracellular genes, and a negative correlation (R= -0.13) between sequencing depth and intracellular Hb content in the four samples. (D) The volcano plot shows the genes that are active in all samples, with the top 10 genes' names given in the red-marked highly variable genes.

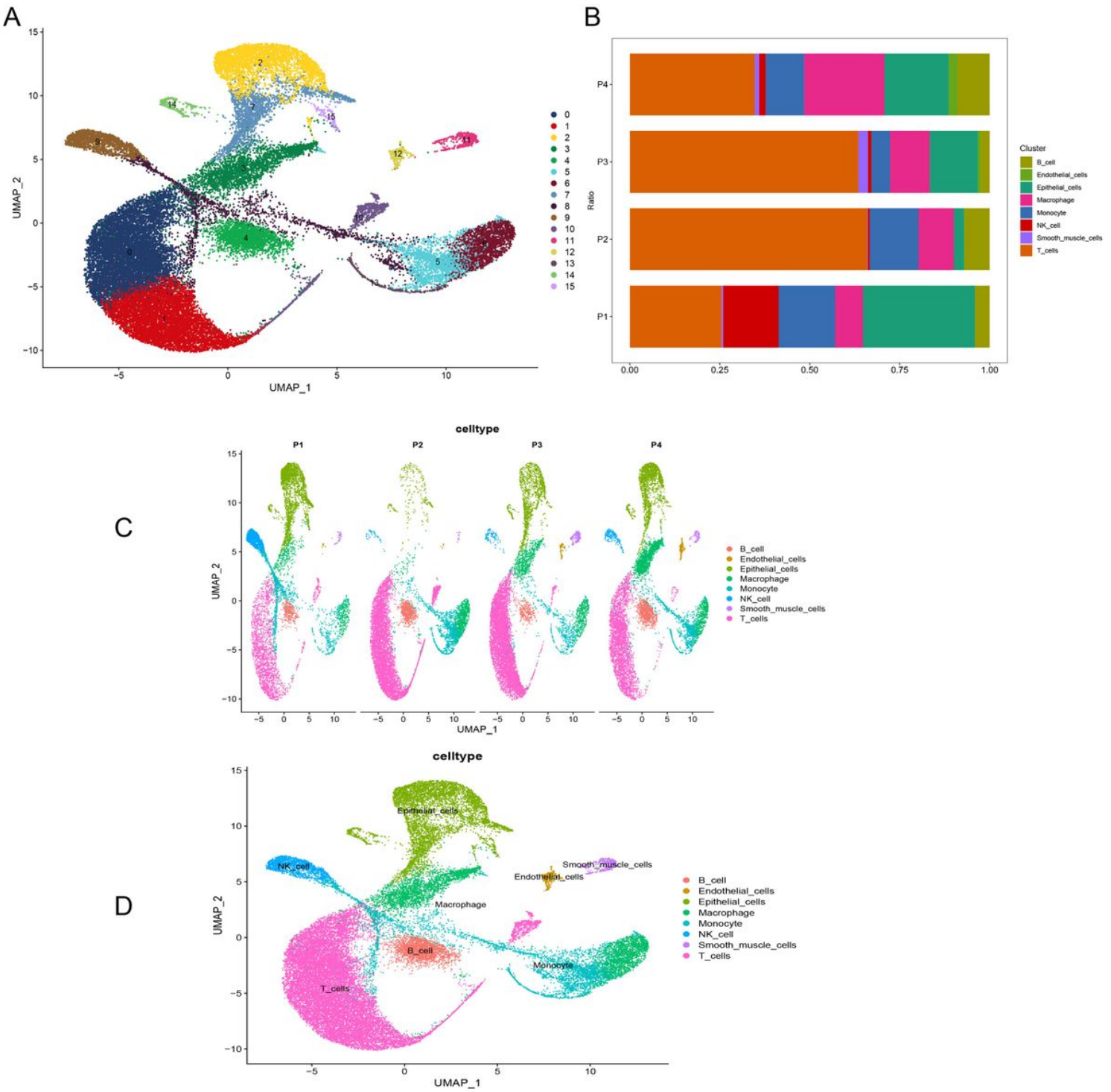

Figure 8

The proportion and annotation of cells in the samples. A Clustering of cells using uMAP. (B) The proportion of cells in the four samples. (C) The distribution of cells in each of the four samples. (D) The distribution and annotation of cell types.

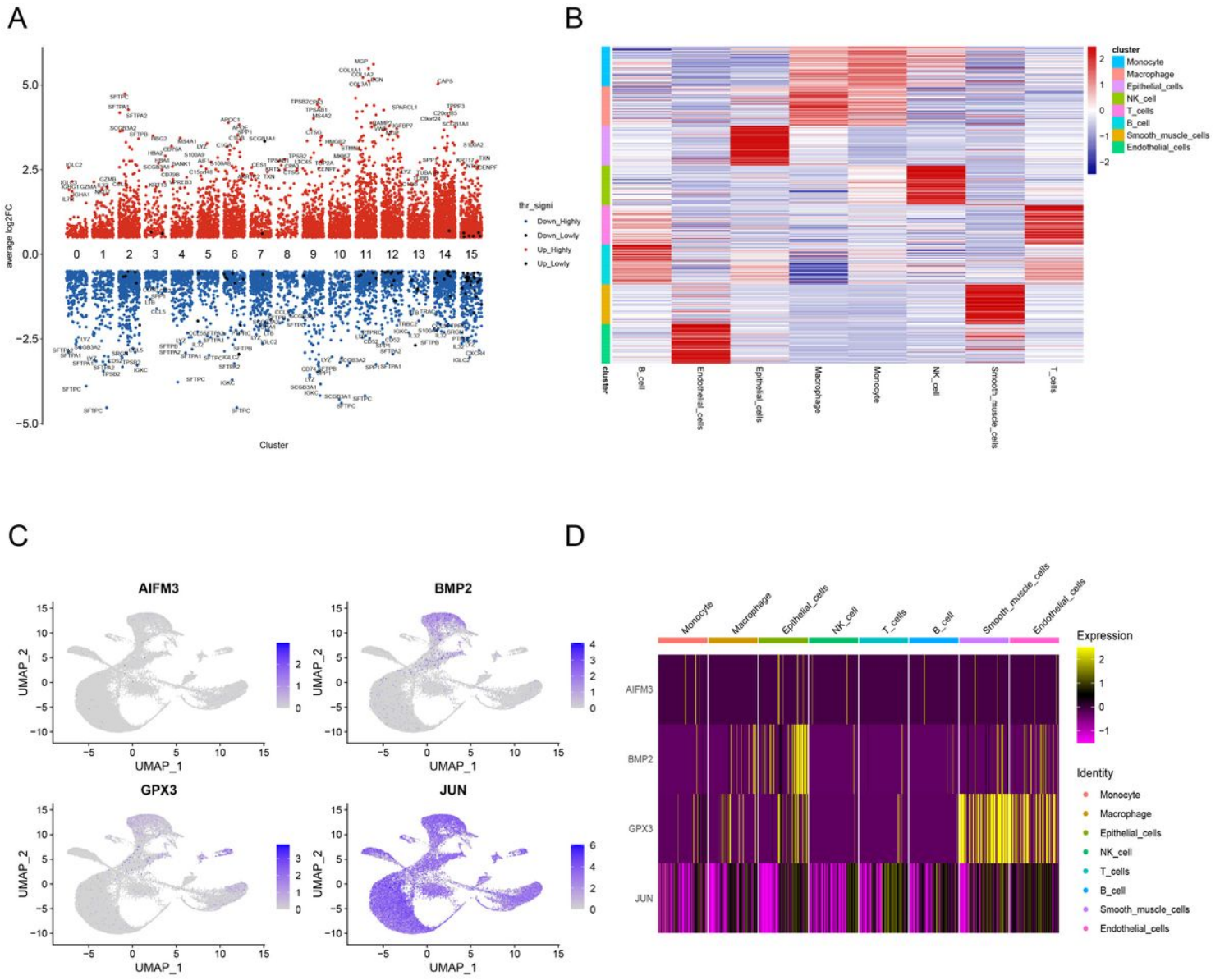

Figure 9

The classification and expression of differential genes and prognostic genes in cell subpopulations. A Volcano plot showing significantly upregulated and downregulated genes in 16 clusters. (B) Heatmap showing marker genes in 8 cell types. (C) The distribution of 4 prognostic genes (AIFM3, BMP2, GPX3, JUN) in 8 cell types. (D) Heatmap showing the specific expression of AIFM3, BMP2, GPX3, JUN in each cell type.

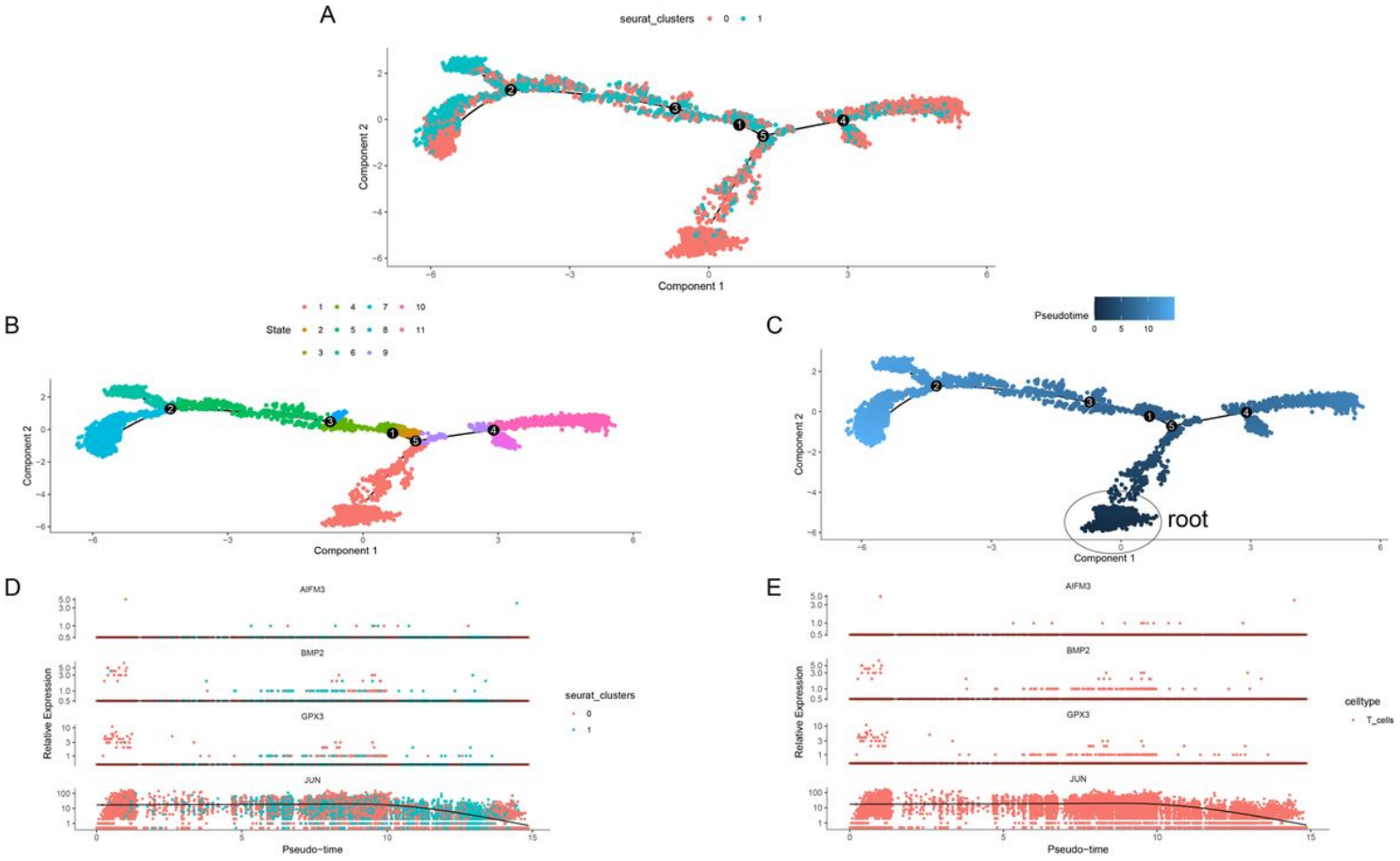

Figure 10

Trajectory analysis of cell subpopulations and prognostic genes. A-C Trajectory analysis of clusters and cell State. Different colored dots represent corresponding clusters or cell states, arranged along the pseudo-temporal branch, and their pseudo-temporal curve. Dark blue denotes an earlier time. (D) The pseudo-temporal changes of AIFM3, BMP2, GPX3, JUN in two clusters. (E) The pseudo-temporal changes of AIFM3, BMP2, GPX3, JUN in T cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xls
- TableS2.xlsx
- FigureS1.jpg