Discover Oncology

Analysis

Construction and analysis of a lysosome-dependent cell death score-based prediction model for non-small cell lung cancer

Jiangping Fu[1] · Yaohua Chen[2] · Jie Li[3] · Ming Tan[4] · Rui Lin[5] · Jiang Wang[5] · Guirong Wu[5] · Yao Rao[2] · Fudao Wu[5] · Youshu Gao[6] · Maoshu Bai[7] · Pingfei Wang[2] · Fang Wu[1]

Received: 15 July 2024 / Revised: 6 August 2024 / Accepted: 13 August 2024
Published online: 30 August 2024

Abstract

Background Non-small cell lung cancer (NSCLC) is the most common type of tumor globally and the leading cause of cancer-related deaths. Although treatment strategies such as immune checkpoint inhibitors and chemotherapy have advanced, the heterogeneity among NSCLC patients results in significant variability in treatment outcomes. Studies have shown that certain patients respond poorly to immune checkpoint inhibitors, indicating that treatment response is closely related to multiple factors. Therefore, it is necessary to develop predictive models to stratify patients based on gene expression and clinical characteristics, aiming for precision therapy.
Objective This study aims to construct a stratified prognostic model for NSCLC patients based on lysosome-dependent cell death (LDCD) scoring by integrating single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing data. By analyzing the immune-related characteristics of high-risk and low-risk groups, we further explored the impact of cell death patterns on lung cancer and identified potential therapeutic targets.
Methods This study obtained single-cell RNA sequencing data and gene expression data of NSCLC patients and normal lung tissues from the GEO and TCGA databases. We used R packages such as Seurat and CellChat for data preprocessing and analysis, and performed dimensionality reduction and visualization through Principal Component Analysis (PCA) and UMAP algorithms. LASSO regression analysis was used to construct the predictive model, followed by cross-validation and ROC curve analysis. The model's effectiveness was validated through survival analysis and immune microenvironment analysis.
Results The study showed a significant increase in the proportion of monocytes in NSCLC tissues, suggesting their important role in cancer progression. Cell communication analysis indicated that macrophages, smooth muscle cells, and myeloid cells exhibit strong intercellular communication during cancer progression. Using the constructed prognostic

Jiangping Fu, Yaohua Chen, Jie Li and Ming Tan have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-024-01233-4.

✉ Pingfei Wang, 956565257@qq.com; ✉ Fang Wu, wufang@gxmu.edu.cn; Jiangping Fu, fujiangping2006@126.com; Yaohua Chen, chyh007@qq.com; Jie Li, 13618272395@163.com; Ming Tan, 349404274@qq.com; Rui Lin, 19938608569@163.com; Jiang Wang, 2365880092@163.com; Guirong Wu, 1004754004@qq.com; Yao Rao, 252733924@qq.com; Fudao Wu, wfde@163.com; Youshu Gao, 416198337@qq.com; Maoshu Bai, bmaoshu@aliyun.com | [1]Department of Radiation Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China. [2]Department of General Respiratory, Dazhou Central Hospital, Dazhou, Sichuan, China. [3]Department of Clinical Research Center, Dazhou Central Hospital, Dazhou, Sichuan, China. [4]Department of Otolaryngology-Head and Neck Surgery, The Central Hospital of Jingmen, Jingmen, China. [5]Department of Oncology, Dazhou Central Hospital, Dazhou, Sichuan, China. [6]Department of Ultrasound Imaging, Dazhou Central Hospital, Dazhou, Sichuan, China. [7]Department of Oncology, Dazhou Second People's Hospital, Dazhou Integrated Traditional Chinese Medicine and Western Medicine Hospital, Dazhou, Sichuan, China.

Discover

model based on 12 LDCD-related genes, we found significant differences in overall survival and immune microenvironment between the high-risk and low-risk groups.

Keywords Non-small cell lung cancer · Lysosome-dependent cell death · Single-cell

1 Introduction

Lung cancer is the most common tumor globally and the leading cause of cancer-related deaths. Non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancers. According to WHO guidelines, lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes [1, 2]. Many factors contribute to the progression of lung cancer, including age, gender, living environment, and smoking status. With ongoing research into NSCLC, treatment strategies have evolved, encompassing immune checkpoint-based immunotherapy and chemotherapy [3]. However, the heterogeneity of NSCLC patients significantly impacts treatment outcomes, with studies showing that some patients exhibit minimal response to immune checkpoint inhibitors [4]. This suggests that treatment response is closely linked to various factors, and not all NSCLC patients benefit from current treatment strategies. Therefore, it is necessary to develop predictive models for patient stratification considering gene expression and clinical characteristics. Through patient stratification, we can identify responses to different treatment strategies and implement appropriate treatments for different patient groups, aligning with the principles of precision therapy and rational drug use.

Lysosome-dependent death is a unique mode of cell death that has great significance for cellular life activities. Lysosomes, as cellular recycling centers, are filled with many hydrolytic enzymes that can degrade most cellular macromolecules. Lysosomal membrane permeabilization and the consequent leakage of lysosomal contents into the cytoplasmic lysate leads to so-called "lysosome-dependent death". This form of cell death is mainly carried out by lysosomal organizing proteases and can have necrotic, apoptotic, or apoptosis-like features depending on the extent of leakage and the cellular context [5]. Many studies have demonstrated that lysosomal-dependent death has an important role in the therapeutic process of tumors [6], and that tumorigenesis can be inhibited by inducing lysosomal-dependent death in tumors.

Single-cell RNA sequencing (scRNA-seq) reveals the highly complex cellular composition of the tumor microenvironment (TME) with high resolution [7]. This technique can uncover developmental changes and cell interaction information within tumor cells with extreme precision, providing new insights into tumor bioinformatics. It is also a powerful tool for future exploration of common characteristics and key differences among various immune cell subsets in the TME [8]. Meanwhile, machine learning, an important branch of artificial intelligence (AI), focuses on enabling computer systems to learn from data and make predictions or decisions. By developing and applying algorithms, machine learning allows computers to recognize patterns and regularities in data, thus improving and enhancing performance without explicit programming. In the biomedical field, researchers use machine learning to analyze clinical data and develop diagnostic and prognostic models for diseases [9]. By combining the critical tumor microenvironment revealed by single-cell RNA sequencing technology with machine learning, We can build stable prognostic models based on the clinical characteristics of non-small cell lung cancer (NSCLC) patients to explore the role of lysosome-dependent cell death in lung carcinogenesis.In summary, the aim of this study was to construct a prognostic model for stratifying NSCLC patients based on lysosomal-dependent cell death scores by integrating clinical features such as single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing (bulk RNA-seq) data. Specifically, we analyzed scRNA-seq and bulk RNA-seq data separately and performed detailed comparisons and analyses of high-risk and low-risk groups, such as those related to immune responses. This approach enabled us to gain deeper insights into the impact of different cell death modes on lung cancer and to identify potential therapeutic targets.

Through this multi-level data integration and analysis, we were able not only to predict the prognosis of NSCLC patients more accurately but also to reveal the specific mechanisms of lysosome-dependent cell death in lung cancer progression. This provides an important basis for the development of personalized treatment plans and helps to discover new and effective therapeutic targets, thereby improving the treatment outcomes and quality of life for NSCLC patients. In conclusion, this study offers a new perspective on the prognostic evaluation of NSCLC and provides important theoretical and practical foundations for exploring the impact of cell death modes on cancer development.

2 Methods and materials

2.1 Data collection and preprocessing

From GEO GSE198099 data set in database (https://www.ncbi.nlm.nih.gov/) for patients with non-small cell lung cancer (GSM5938737, GSM5938738) and normal lung tissue (GSM5938739, GSM5938740) single-celled RNA sequencing data. In addition, from the TCGA database (https://portal.gdc.cancer.gov/) and GEO GSE30219 data set in the database, respectively for 585 cases and 272 cases of patients with non-small cell lung cancer gene expression profile, Their clinical characteristics such as survival status, survival time, and TMN stage are shown in Supplementary Tables 1 and 2. The combined data were batch corrected using the "ComBat" function from "limma" (PMC4402510) and "sva" R package. TCGA was used as the training set and GSE30219 as the test set.

2.2 Processing of scRNA-seq data

Single-cell RNA sequencing data were read from 10X files, and a "Seurat" object was created. Cells with low quality were filtered out based on criteria of minimum 200 genes, maximum 4000 genes, and mitochondrial gene proportion of 20%. Differential expression genes were selected using the "FindVariableFeatures()" function, and a plot of these genes was generated using the "VariableFeaturePlot()" function. The data were standardized using the "ScaleData()" function to remove batch effects in gene expression levels. The top 10 differentially expressed genes were labeled on the plot for further analysis. Principal component analysis (PCA) was performed to reduce dimensionality, and highly variable genes were selected as features. Dimensionality reduction visualization was carried out using "tSNE" and "UMAP" algorithms. The "createCellChat" function creates a "CellChat" object for cell communication analysis, identifies overexpressed genes and ligand-receptor pairs, and maps ligands and receptors onto the protein-protein interaction network.

2.3 The identification of mononuclear cells and communication analysis

Single cell RNA sequencing (scRNA-seq) data using the Seurat packages were analyzed, and the first to use t—are initially dimension reduction, SNE in visualization of cells depending on the type of organization. Monocytes were isolated and reclustered by scale analysis and principal component analysis (PCA). Key myeloid marker genes were identified using dot plots and cell types were annotated accordingly to obtain monocyte subsets. Non-small cell lung cancer (non-small cell lung cancer, NSCLC) organization of mononuclear cells are integrated into monocytes data set for further analysis. Using CellChat package analysis intercellular communication, mainly analyzes the interaction of secretion signal. We identified the excessive expression of genes and interaction, and calculate the probability of communication, and use the network diagram visualization. We passed the heat map analysis and visualization centricity index and the signal function, highlight the outgoing and incoming signal model.

2.4 Analysis of monocyte subpopulations

Software packages such as "reshape2", "ggplot2" and "dplyr" were used to organize and visualize the data, obtain statistical analysis of cell types and generate bar charts. Subsequently, we used the Seurat software package for dimensionality reduction, clustering, and identification of monocyte subsets. Correction mass effect, the use of "harmonious" algorithm using UMAP algorithm dimensionality of data visualization. For cell communication analysis, we constructed cell communication networks using the "CellChat" package and identified and analyzed ligand-receptor pairs. In addition, we also performed visual analysis of the topology and signaling pathways of the cellular communication network.

2.5 Modularization and network analysis of monocyte scRNA-seq data using "hdWGCNA" method

We preprocessed and cleaned the raw data using the "hdWGCNA" and "Seurat" packages in R. Subsequently, we filtered genes expressed in at least 5% of cells and constructed "metacells", followed by normalization of the "metacell" expression matrix. Next, we determined the appropriate soft power based on testing soft threshold values and constructed a co-expression network. Based on this network, we generated a dendrogram of the co-expression network and obtained the TOM matrix for subsequent advanced analysis. We also calculated the module eigengenes and performed inter-modular

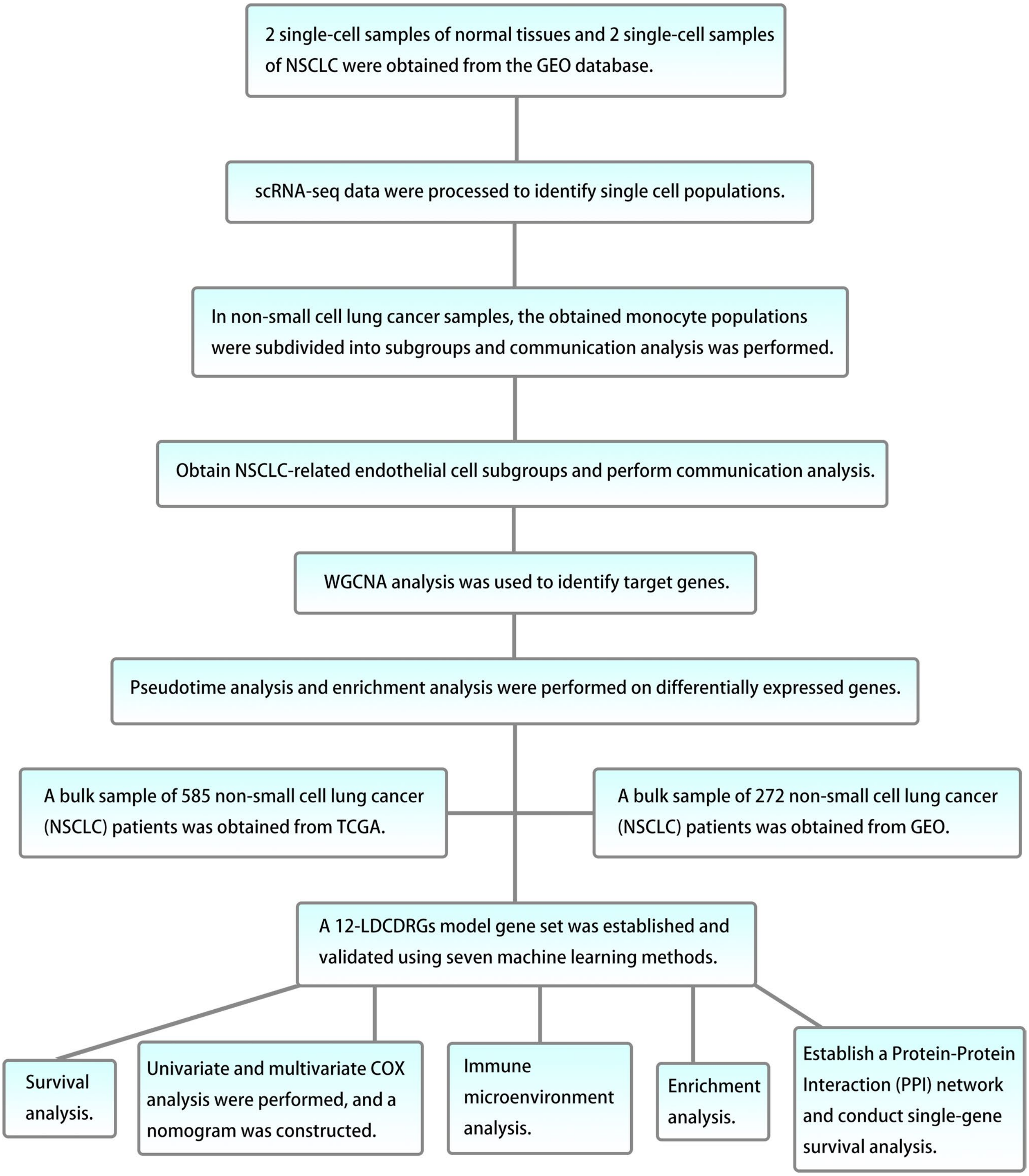

2 single-cell samples of normal tissues and 2 single-cell samples of NSCLC were obtained from the GEO database.
scRNA-seq data were processed to identify single cell populations.
In non-small cell lung cancer samples, the obtained monocyte populations were subdivided into subgroups and communication analysis was performed.
Obtain NSCLC-related endothelial cell subgroups and perform communication analysis.
WGCNA analysis was used to identify target genes.
Pseudotime analysis and enrichment analysis were performed on differentially expressed genes.
A bulk sample of 585 non-small cell lung cancer (NSCLC) patients was obtained from TCGA.
A bulk sample of 272 non-small cell lung cancer (NSCLC) patients was obtained from GEO.
A 12-LDCDRGs model gene set was established and validated using seven machine learning methods.
Survival analysis.
Univariate and multivariate COX analysis were performed, and a nomogram was constructed.
Immune microenvironment analysis.
Enrichment analysis.
Establish a Protein-Protein Interaction (PPI) network and conduct single-gene survival analysis.

Fig. 1 **A** flow chart was used to illustrate the main ideas and design steps of the study

connectivity analysis. Additionally, we generated module feature plots ranked by gene kME and identified hub genes. Also, we saved key analysis processes and performed various visualizations, including correlation plots between modules and "dotplot" of module features. In the end, we all chose green, blue, and the most significant difference in turquoise module of the 30 genes, a total of 90 genes as a core.

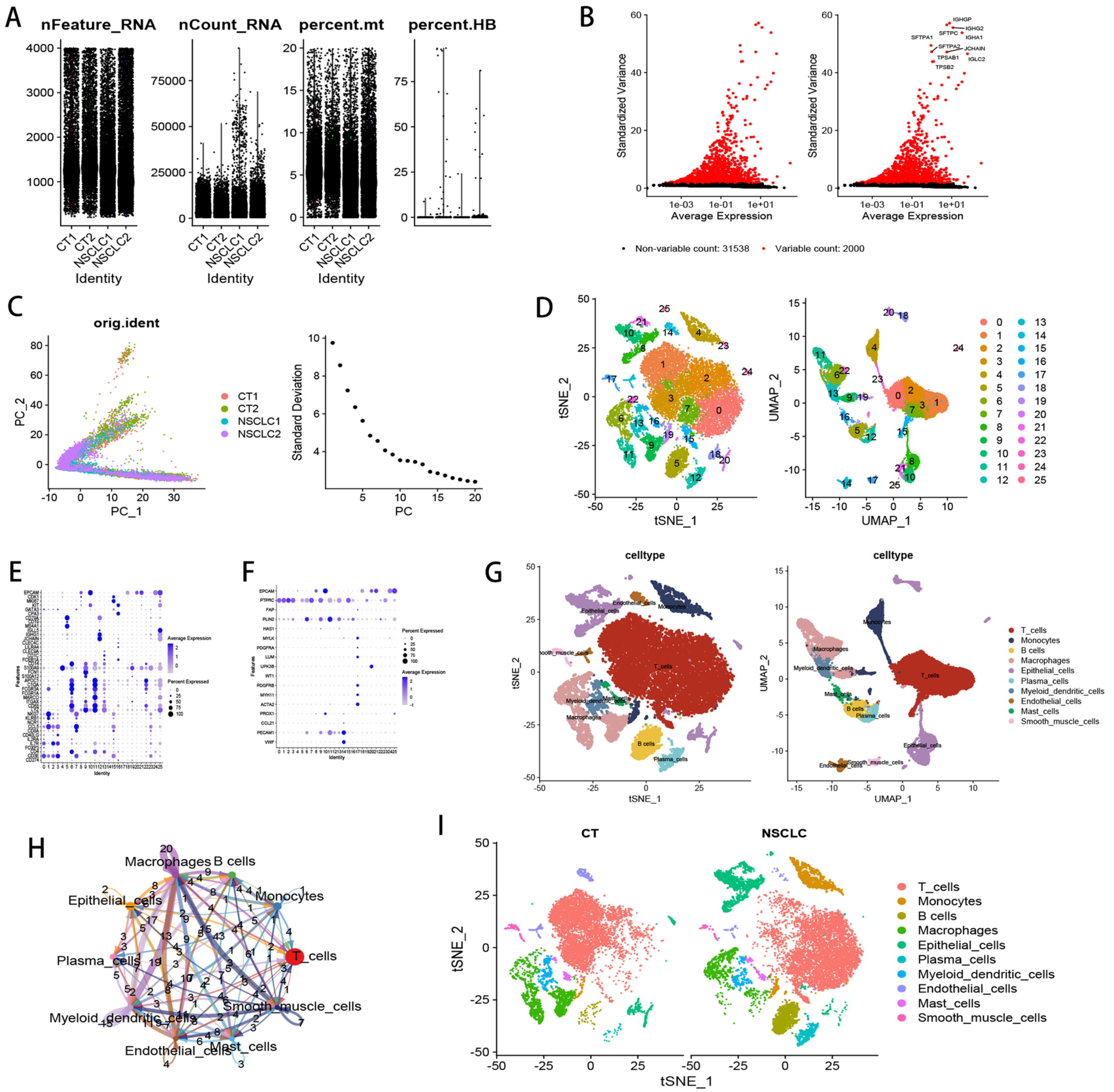


Fig. 2 **We** utilized single-cell RNA sequencing data from non-small cell lung cancer (NSCLC) and normal individuals to study the cellular landscape and potential intercellular communication within the tissue. The obtained single-cell data underwent quality control and normalization (**A**). The top 2000 variable features were selected and a scatter plot was created, with the top 10 variable features highlighted (**B**). PCA analysis was performed on the normalized data (**C**). The left panel shows the distribution of cell samples after PCA dimensionality reduction, with different colors representing different original identities (orig.ident). The right panel is an elbow plot, displaying the standard deviation of the first 20 principal components. By observing the elbow plot, we can determine the number of principal components that best represent the data structure. Dimensionality reduction visualizations of cell clusters using t-SNE and UMAP are shown (**D**). **E** shows the expression of immune-related genes, while **F** shows the expression of other genes. The cell clusters were annotated by marking, including endothelial cells, T cells, macrophages, plasma cells, myeloid dendritic cells, monocytes, B cells, mast cells, smooth muscle cells, and epithelial cells, with t-SNE and UMAP visualizations of dimensionality reduction (**G**). We also analyzed intercellular communication, resulting in a chord diagram of intercellular communication (**H**). Additionally, we plotted tSNE comparisons between normal and NSCLC samples (**I**)

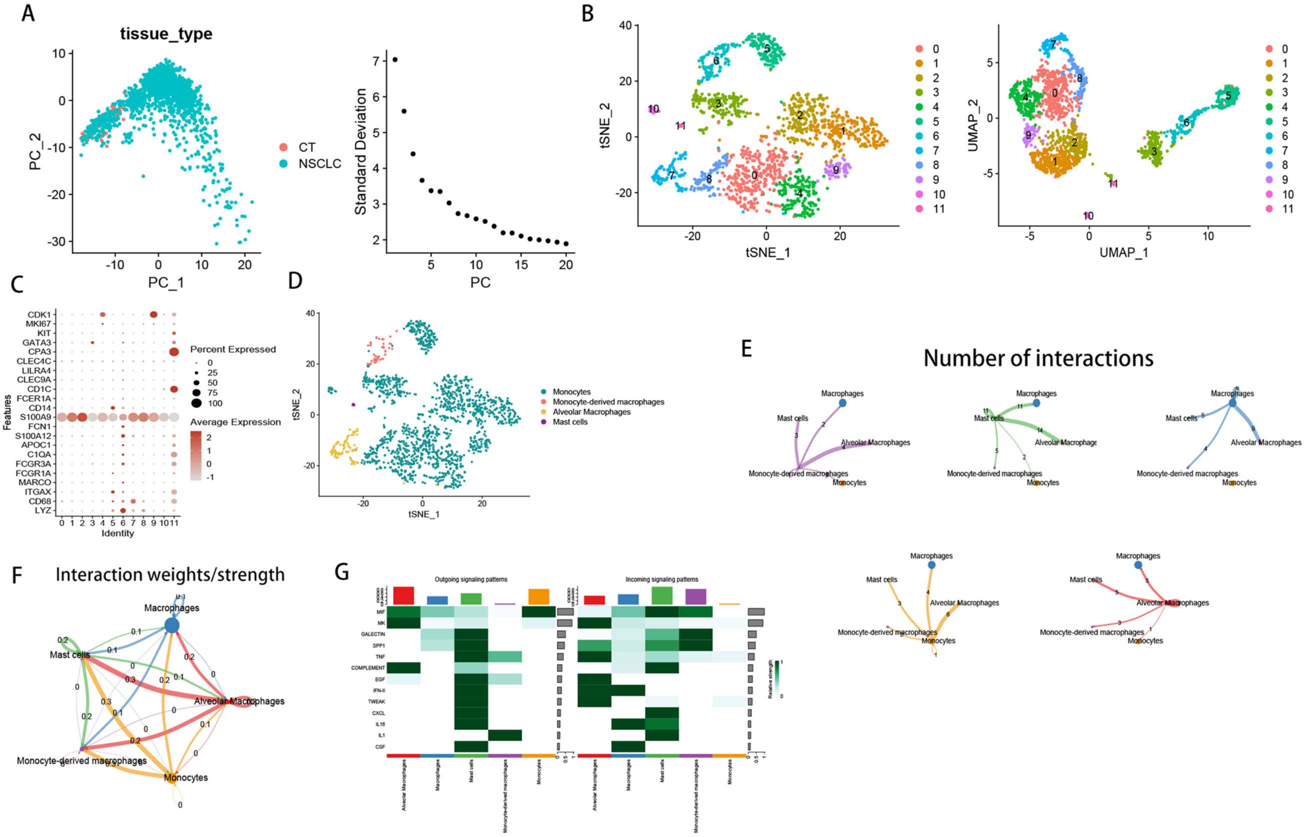


Fig. 3 We conducted pseudotime and communication analysis on monocyte subpopulations. A PCA analysis was performed on the monocytes, resulting in an elbow plot (**A**). The tSNE and UMAP dimensionality reduction results of the single-cell subpopulations (scRNA_mono) are shown (**B**). A DotPlot displays the expression distribution of marker genes in different cell clusters, with different colors representing the intensity of gene expression (**C**). The tSNE dimensionality reduction results of monocytes, monocyte-derived macrophages, alveolar macrophages, and mast cells in the single-cell subpopulations (scRNA_mono) are shown (**D**). We analyzed and visualized the trajectory of cell states in the subpopulations (**E**) and analyzed the communication between the generated subpopulations and the original monocyte groups (**F**). We used a heatmap to compare the input and output of various signaling molecules in monocytes and selected subpopulations (**G**)

2.6 Pseudo-temporal analysis of monocytes and enrichment analysis of differentially expressed genes

In addition to modularization and network analysis, we conducted pseudo-temporal analysis to better understand the gene expression changes of monocytes across different states. Initially, leveraging single-cell transcriptomic data, we performed pseudo-temporal analysis using the "monocle" package. Through this analysis, we elucidated the distribution of monocytes along states and pseudo-time trajectories. Furthermore, we integrated the results of pseudo-temporal analysis with the previously constructed co-expression network to further explore key genes associated with pseudo-temporal dynamics. Ultimately, we generated a pseudo-temporal heatmap illustrating the expression patterns of key genes associated with pseudo-temporal dynamics in monocytes. Moreover, we conducted enrichment analysis of 90 differentially expressed genes using the gprofiler website (https://biit.cs.ut.ee/gprofiler/gost).

2.7 Using LASSO analysis to obtain model 12-LDCDRGs and validation

From the results of the "hdWGCNA" analysis, 150 differentially expressed genes were obtained. The "GetHubGenes" function was utilized to retrieve important genes. Genes were then selected based on module specification, followed by single-factor logistic regression to identify significant genes. Single-factor Cox proportional hazards regression analysis was conducted on candidate genes in the training set to select feature genes associated with prognosis. Variables with p-values < 0.05 were included in the least absolute shrinkage and selection operator (LASSO) regression analysis. This analysis was performed using the "glmnet" package in R (PMC2929880) to reduce the number of genes in the final risk model. The prognosis model was constructed based on the following formula: Risk score = Gene exp1 × β1 + Gene exp2 × β2 + … + Gene expression n × βn (where gene expression represents the expression value of the gene, and β represents the corresponding LASSO regression coefficient). Standardization was applied to both the training and testing sets. Classification task objects were created using the "mlr3" package in R, initializing various learning algorithms such as logistic regression, linear discriminant analysis, support vector machines, etc. Cross-validation was performed to evaluate the performance of each learning algorithm, visualizing the performance of different algorithms including AUC curves and box plots. The naïve_bayes model with the best performance was selected for final testing, and ROC curves were plotted with AUC values calculated.

2.8 Prognostic analysis and nomogram construction for the 12-LDCDRGs model

Survival analysis was conducted using the "survival" and "survminer" packages in R. Survival curves between high-risk and low-risk groups in both the training and testing sets were compared, and relevant p-values were calculated. The "bioForest" function was employed to create a risk forest plot, and the "indep" function was used to perform univariate and multivariate Cox regression analyses, generating corresponding risk forest plots. Subsequently, the "coxph" function was utilized to fit the Cox proportional hazards model, and risk curve plots were generated. The model was then calibrated, and calibration curve plots for 1 year, 3 years, and 5 years were constructed. Finally, a nomogram was constructed based on the Cox proportional hazards model to visualize the predictive ability of the 12-LDCDRGs model for 1 year, 3 years, and 5 years survival probabilities.

2.9 Analysis of immune microenvironment

First of all, we to TCGA dataset of gene expression data preprocessing, excluding the expressed genes, using the 'voom' function in limma package remaining data standardization. Then, cell type deconvolution was performed using the normalized data using the CIBERSORT algorithm with 1000 permutations and quantile normalization. The resulting cell type proportions were screened to include only samples with a p-value less than 0.05. These proportions were further analyzed using boxplots to visualize low-risk and high-risk category differences. In addition, using GSVA packages for a single sample enrichment analysis (ssGSEA) gene set. ssGSEA scores were standardized and correlated with risk scores. In addition, we to the gene expression analysis, gene PLEKHM1 its expression level components associated with the immune cells and use the ggplot2 results visualization.

2.10 Analysis of enrichment

We extracted symbolic gene sets from the MSigDB database and performed a detailed gene set Variation Analysis (GSVA). To identify statistically significant subgroup differences, adjusted $p < 0.05$. We used the Metascape website (https://metascape.org/) for enrichment analysis of 12-LDCDRGs.

2.11 PPI was used to screen core genes and single gene survival analysis

We collected protein-protein interaction data using the string website (https://string-db.org/) and visualized the topology of the PPI network using Cytoscape software. The top 10 core genes were screened by degree. In addition, we performed survival analysis of these 10 genes using the "survival" and "survminer" R packages.

2.12 Statistical analysis

Statistical analyses were performed with the use of R software, versions 4.3.1 and R 4.1.3. We obtained data from TCGA and GEO databases for quality control and batch effect correction, used a variety of techniques for data reduction and visualization, and applied different algorithms for cell subpopulation analysis, modularity and network analysis, feature selection and model building, and immune infiltration analysis. Finally, machine learning methods were used to build prediction models, including kknn, lda, log_reg, naive_bayes, ranger, rpart and svm. Kaplan-Meier (KM) survival curve and log-rank test were used to compare the overall survival (OS) between high-risk group and low-risk group. We used

Fig. 4 We extracted differential subpopulations of monocytes from the NSCLC group and the control group, analyzing the communication between NSCLC-associated endothelial cell groups and other cell groups. We processed the data of the endothelial cell groups. The differences between these subpopulations are visually displayed through bar plots, highlighting potential biomarkers or functional characteristics specific to NSCLC-associated monocytes (**A**). Bar plots were used to compare the number of cells in the cell groups of the NSCLC group and the normal group (**B** and **C**). We processed the data of the endothelial cell groups to obtain the PCA dimensionality reduction visualization and elbow plot (**D**), and generated UMAP plots for the CT and NSCLC groups (**E**). Additionally, we used bar plots to show the expression differences of cell subpopulations between these two groups (**F**). We extracted subpopulations 0, 1, 4, 5, 6, 7, and 8 as characteristic subpopulations of NSCLC and performed quantitative and weighted communication analysis with other cell groups (**G**), displaying the specific pathways of intercellular communication through a heatmap (**H**). The communication within different subpopulations in the MIF signaling pathway is shown (**I**). The communication within different subpopulations in the TNF signaling pathway is shown (**J**). The input and output of various signals in different subpopulations are represented using a heatmap (**K**). Furthermore, we used a bubble plot to show the input and output of different cells (**L**)

the Wilcoxon test to explore differences in the performance of tumor-infiltrating immune cells and immune function between the two cohorts. Analysis of variance was used for statistical analysis, p value and false discovery rate (FDR) q value less than 0.05 were considered statistically significant. These methods provide comprehensive data processing and analysis support for research and contribute to a deeper understanding of the pathogenesis and potential therapeutic targets of non-small cell lung cancer.

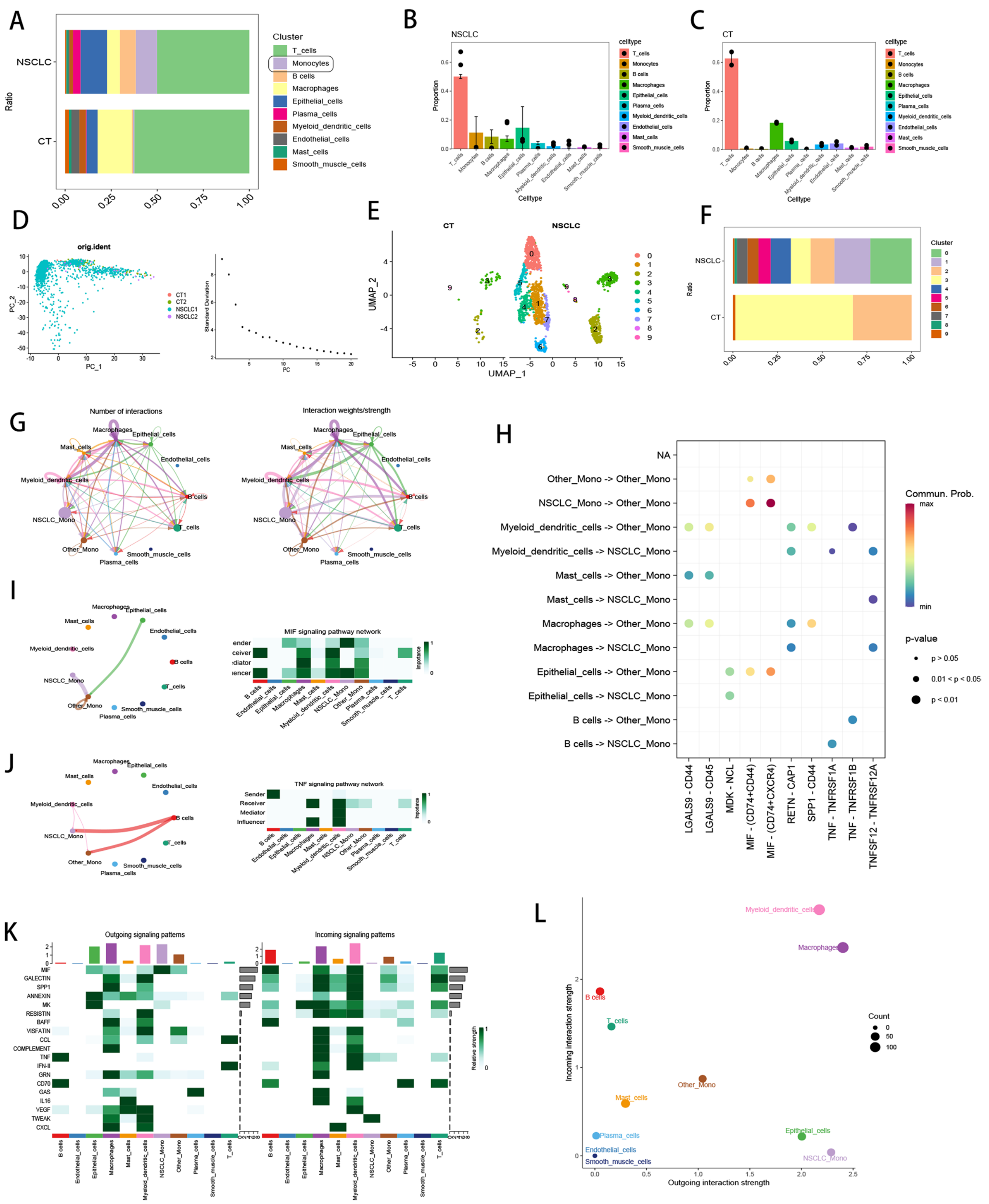
3 Result

3.1 Single cell data quality control

We have detailed the entire process of our study in Fig. 1. Four samples from the normal and NSCLC tumor tissues of patients were collected from public databases. The Seurat package was used for quality control of the single-cell data, retaining cells with a total RNA count greater than 200, less than 4000, and a mitochondrial RNA proportion less than 15% (Fig. 2A, B), to exclude low-quality or damaged cells and those potentially affected by excessive mitochondrial RNA or cellular stress. Next, based on principal component clustering analysis (Fig. 2C), the cells were divided into 25 clusters (Fig. 2D). Throughout the annotation process of these 25 cell clusters, we identified ten different cell types: T cells, monocytes, B cells, macrophages, endothelial cells, plasma cells, myeloid cells, epithelial cells, smooth muscle cells, and mast cells (Fig. 2E–G). Additionally, we conducted extensive research on intercellular communication within each cell type (Fig. 2H). Our findings indicated that macrophages, smooth muscle cells, and myeloid cells exhibited strong intercellular communication with other cells, highlighting their activity in cancer progression. We then mapped the distribution of different cell types in normal and cancerous tissues, allowing us to visualize each cell type (Fig. 2I). It was observed that the number of immune cells, such as monocytes, increased in NSCLC tissues, indicating significant immune infiltration in the cancerous tissues.

3.2 Exploring monocyte communication patterns

To explore the immune role in non-small cell lung cancer (NSCLC), we conducted an in-depth analysis of monocytes. We extracted the monocyte population for PCA dimensionality reduction clustering (Fig. 3A). The results identified 11 clusters of monocytes (Fig. 3B). By identifying the expression levels of marker genes in each cluster (Fig. 3C), we categorized these 11 clusters into four types of cells: monocytes, monocyte-derived macrophages, alveolar macrophages, and mast cells (Fig. 3D). To investigate the differences between these four cell types and the previously identified macrophage communities, we performed cell communication analysis by identifying the number of communication receptors and ligands for each cell type (Fig. 3E). The results showed strong communication between mast cells and both macrophages and alveolar macrophages. We then analyzed the communication intensity between them (Fig. 3F), which also indicated more pronounced communication between mast cells and different macrophages compared to other cells. To further explore the communication patterns of mast cells, we used a heatmap to display the incoming and outgoing communication patterns of cells under different signaling pathways (Fig. 3G). It can be seen that mast cells exhibit a large amount of signal output under different signaling pathways, and they predominantly receive signals through MIF, EGF, CXCL, and IL1. Finally, we analyzed the proportion of each cell type in the NSCLC and control groups using bar charts (Fig. 4A, B). The bar charts showed a significant increase in the proportion of monocytes in the NSCLC group compared to the


A
NSCLC
CT
Ratio
0.00
0.25
0.50
0.75
1.00
Cluster
T_cells
Monocytes
B cells
Macrophages
Epithelial_cells
Plasma_cells
Myeloid_dendritic_cells
Endothelial_cells
Mast_cells
Smooth_muscle_cells
B
NSCLC
Proportion
Celltype
celltype
T_cells
Monocytes
B cells
Macrophages
Epithelial_cells
Plasma_cells
Myeloid_dendritic_cells
Endothelial_cells
Mast_cells
Smooth_muscle_cells
C
CT
Proportion
Celltype
D
orig.ident
PC_1
PC_2
CT1
CT2
NSCLC1
NSCLC2
Standard Deviation
PC
E
CT
NSCLC
UMAP_1
UMAP_2
0
1
2
3
4
5
6
7
8
9
F
NSCLC
CT
Ratio
Cluster
0.00
0.25
0.50
0.75
1.00
G
Number of interactions
Interaction weights/strength
Macrophages
Epithelial_cells
Mast_cells
Endothelial_cells
Myeloid_dendritic_cells
B cells
NSCLC_Mono
T_cells
Other_Mono
Smooth_muscle_cells
Plasma_cells
H
NA
Other_Mono -> Other_Mono
NSCLC_Mono -> Other_Mono
Myeloid_dendritic_cells -> Other_Mono
Myeloid_dendritic_cells -> NSCLC_Mono
Mast_cells -> Other_Mono
Mast_cells -> NSCLC_Mono
Macrophages -> Other_Mono
Macrophages -> NSCLC_Mono
Epithelial_cells -> Other_Mono
Epithelial_cells -> NSCLC_Mono
B cells -> Other_Mono
B cells -> NSCLC_Mono
LGALS9 - CD44
LGALS9 - CD45
MDK - NCL
MIF - (CD74+CD44)
MIF - (CD74+CXCR4)
RETN - CAP1
SPP1 - CD44
TNF - TNFRSF1A
TNF - TNFRSF1B
TNFSF12 - TNFRSF12A
Commun. Prob.
max
min
p-value
p > 0.05
0.01 < p < 0.05
p < 0.01
I
MIF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
J
TNF signaling pathway network
Sender
Receiver
Mediator
Influencer
Importance
K
Outgoing signaling patterns
Incoming signaling patterns
MIF
GALECTIN
SPP1
ANNEXIN
MK
RESISTIN
BAFF
VISFATIN
CCL
COMPLEMENT
TNF
IFN-II
GRN
CD70
GAS
IL16
VEGF
TWEAK
CXCL
Relative strength
L
Outgoing interaction strength
Incoming interaction strength
Myeloid_dendritic_cells
Macrophages
B cells
T_cells
Other_Mono
Mast_cells
Plasma_cells
Epithelial_cells
Endothelial_cells
Smooth_muscle_cells
NSCLC_Mono
Count
0
50
100

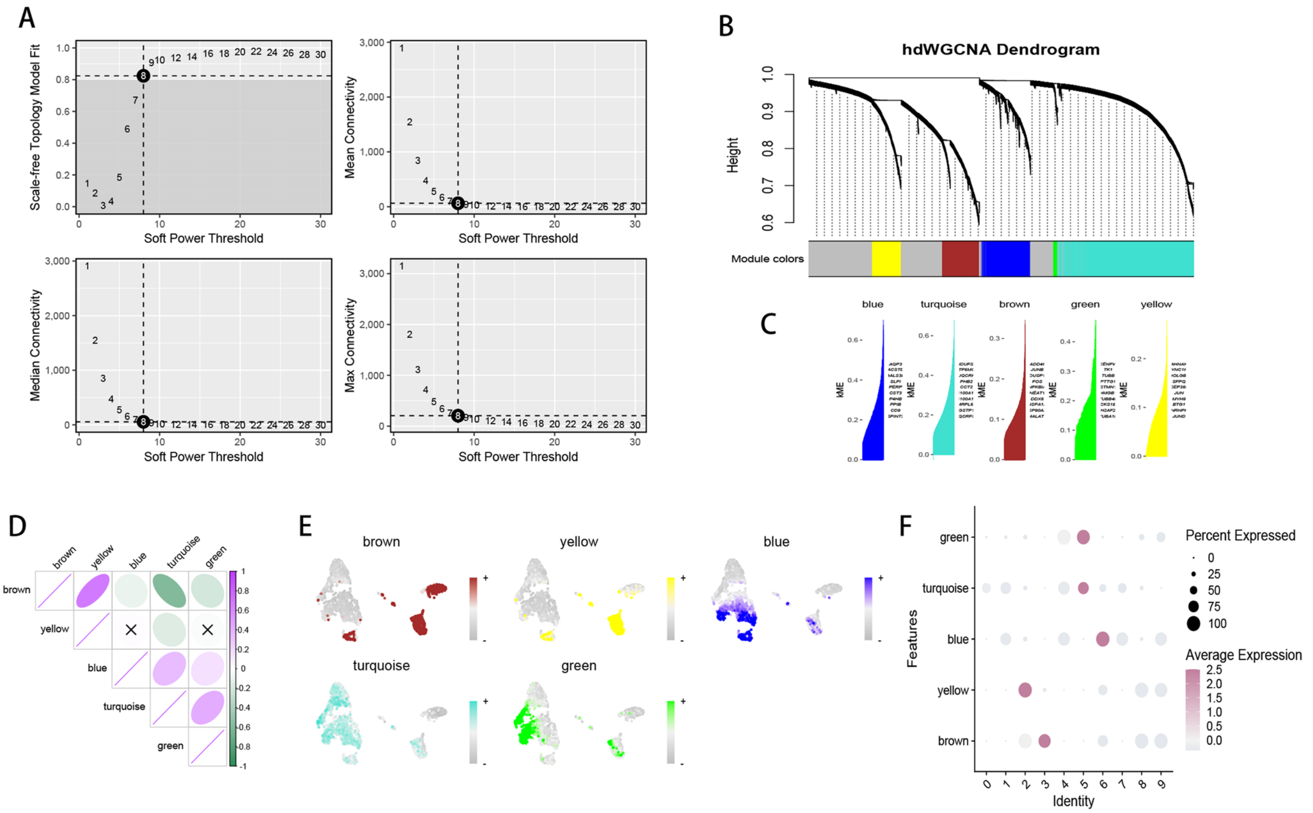
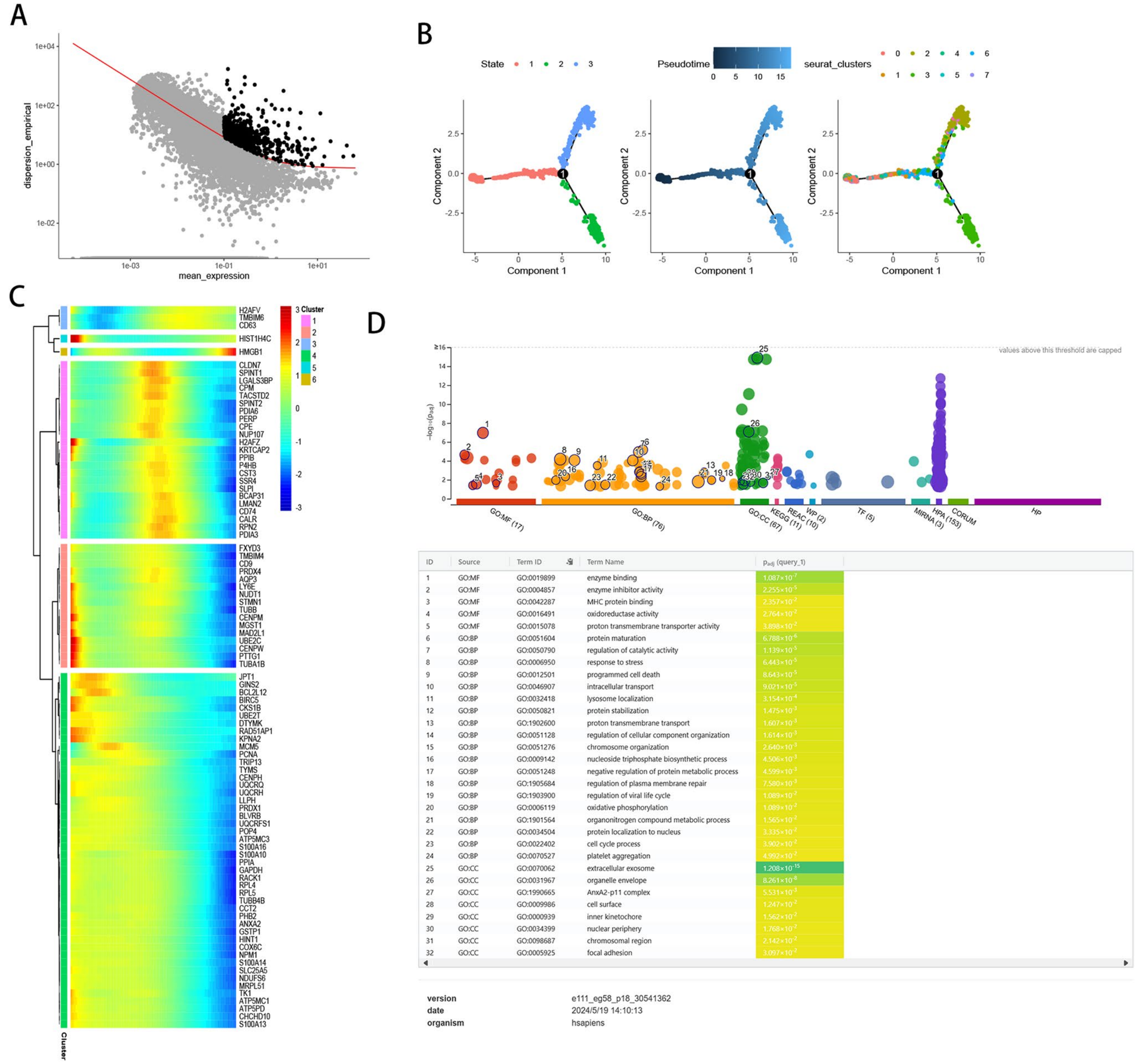


Fig. 5 Using weighted gene co-expression network analysis (WGCNA) on single-cell RNA sequencing data, we selected an appropriate soft threshold to further determine the target color modules. We determined a suitable soft threshold of 8 for WGCNA (**A**) and visualized the dendrogram of the constructed co-expression network (**B**). We obtained 100 hub genes and then plotted the module membership (KME) distribution of these hub genes using the PlotKMEs function (**C**). Additionally, we plotted the correlation heatmap between modules (**D**). By observing the expression of different colors within subgroups, we identified modules that were significantly expressed in the target subgroups (**E**) and plotted a heatmap of subgroup color expression (**F**)

control group, suggesting the important role of monocytes in cancer progression. This finding supports the potential for future research targeting monocytes to develop new therapeutic targets for NSCLC.

3.3 Revealing the total cellular communication landscape

Similarly, given the significant increase in epithelial cells in NSCLC samples (Fig. 4A), we extracted epithelial cells for PCA dimensionality reduction clustering (Fig. 4B–D). Using UMAP, we displayed the clustering results of epithelial cells from different sources (Fig. 4E) and visualized the proportion differences among different clusters (Fig. 4F). To further reveal the landscape of cell communication, we conducted a cell communication analysis of all cell types. Figure 4G shows the number of receptors and ligands as well as the intensity of cell communication between different cell types, highlighting the strong interactions between epithelial cells and other cells. To explore the details of cell communication further, we used a heatmap to display each cell type under various communication pathways (Fig. 4H). We found that the MIF pathway mediated strong cell communication, while the TNF pathway showed a decrease in communication intensity, indicating the critical roles of the MIF and TNF pathways in cell communication during NSCLC development.



ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0019899	enzyme binding	1.087×10^{-7}
2	GO:MF	GO:0004857	enzyme inhibitor activity	2.255×10^{-5}
3	GO:MF	GO:0042287	MHC protein binding	2.357×10^{-2}
4	GO:MF	GO:0016491	oxidoreductase activity	2.764×10^{-2}
5	GO:MF	GO:0015078	proton transmembrane transporter activity	3.898×10^{-2}
6	GO:BP	GO:0051604	protein maturation	6.788×10^{-6}
7	GO:BP	GO:0050790	regulation of catalytic activity	1.139×10^{-5}
8	GO:BP	GO:0006950	response to stress	6.443×10^{-5}
9	GO:BP	GO:0012501	programmed cell death	8.643×10^{-5}
10	GO:BP	GO:0046907	intracellular transport	9.021×10^{-5}
11	GO:BP	GO:0032418	lysosome localization	3.154×10^{-4}
12	GO:BP	GO:0050821	protein stabilization	1.475×10^{-3}
13	GO:BP	GO:1902600	proton transmembrane transport	1.607×10^{-3}
14	GO:BP	GO:0051128	regulation of cellular component organization	1.614×10^{-3}
15	GO:BP	GO:0051276	chromosome organization	2.640×10^{-3}
16	GO:BP	GO:0009142	nucleoside triphosphate biosynthetic process	4.506×10^{-3}
17	GO:BP	GO:0051248	negative regulation of protein metabolic process	4.599×10^{-3}
18	GO:BP	GO:1905684	regulation of plasma membrane repair	7.580×10^{-3}
19	GO:BP	GO:1903900	regulation of viral life cycle	1.089×10^{-2}
20	GO:BP	GO:0006119	oxidative phosphorylation	1.089×10^{-2}
21	GO:BP	GO:1901564	organonitrogen compound metabolic process	1.565×10^{-2}
22	GO:BP	GO:0034504	protein localization to nucleus	3.335×10^{-2}
23	GO:BP	GO:0022402	cell cycle process	3.902×10^{-2}
24	GO:BP	GO:0070527	platelet aggregation	4.992×10^{-2}
25	GO:CC	GO:0070062	extracellular exosome	1.208×10^{-15}
26	GO:CC	GO:0031967	organelle envelope	8.261×10^{-8}
27	GO:CC	GO:1990665	AnxA2-p11 complex	5.531×10^{-3}
28	GO:CC	GO:0009986	cell surface	1.247×10^{-2}
29	GO:CC	GO:0000939	inner kinetochore	1.562×10^{-2}
30	GO:CC	GO:0034399	nuclear periphery	1.768×10^{-2}
31	GO:CC	GO:0098687	chromosomal region	2.142×10^{-2}
32	GO:CC	GO:0005925	focal adhesion	3.097×10^{-2}

version	e111_eg58_p18_30541362
date	2024/5/19 14:10:13
organism	hsapiens

Fig. 6 We further analyzed the developmental trajectory and conducted enrichment analysis of differentially expressed genes related to Lysosome-dependent cell death. We obtained highly variable genes (**A**) and plotted the developmental trajectory of different endothelial cell subpopulations (**B**), validated by a heatmap of core genes (**C**). We used the website gprofiler (https://biit.cs.ut.ee/gprofiler/) to construct a PPI network (Fig. 7C). The top 10 GO enrichment histograms of core genes were plotted (Fig. 7D)

In Fig. 4I and J, we detailed the cell interactions under the MIF and TNF pathways. In the MIF pathway, monocytes associated with NSCLC were the primary signal transmitters. In the TNF pathway, B cells mediated the signal transmission. We used a heatmap to illustrate the incoming and outgoing communication patterns of all cell types under different signaling pathways (Fig. 4K). In signal transmission, epithelial cells, macrophages, myeloid cells, and NSCLC-associated monocytes played major roles. In signal reception, B cells, macrophages, and myeloid cells were the main cell types receiving signals. Figure 4L provides a more intuitive quantification of the cells involved in signal reception, showing that macrophages and myeloid cells predominantly contributed to signal communication.

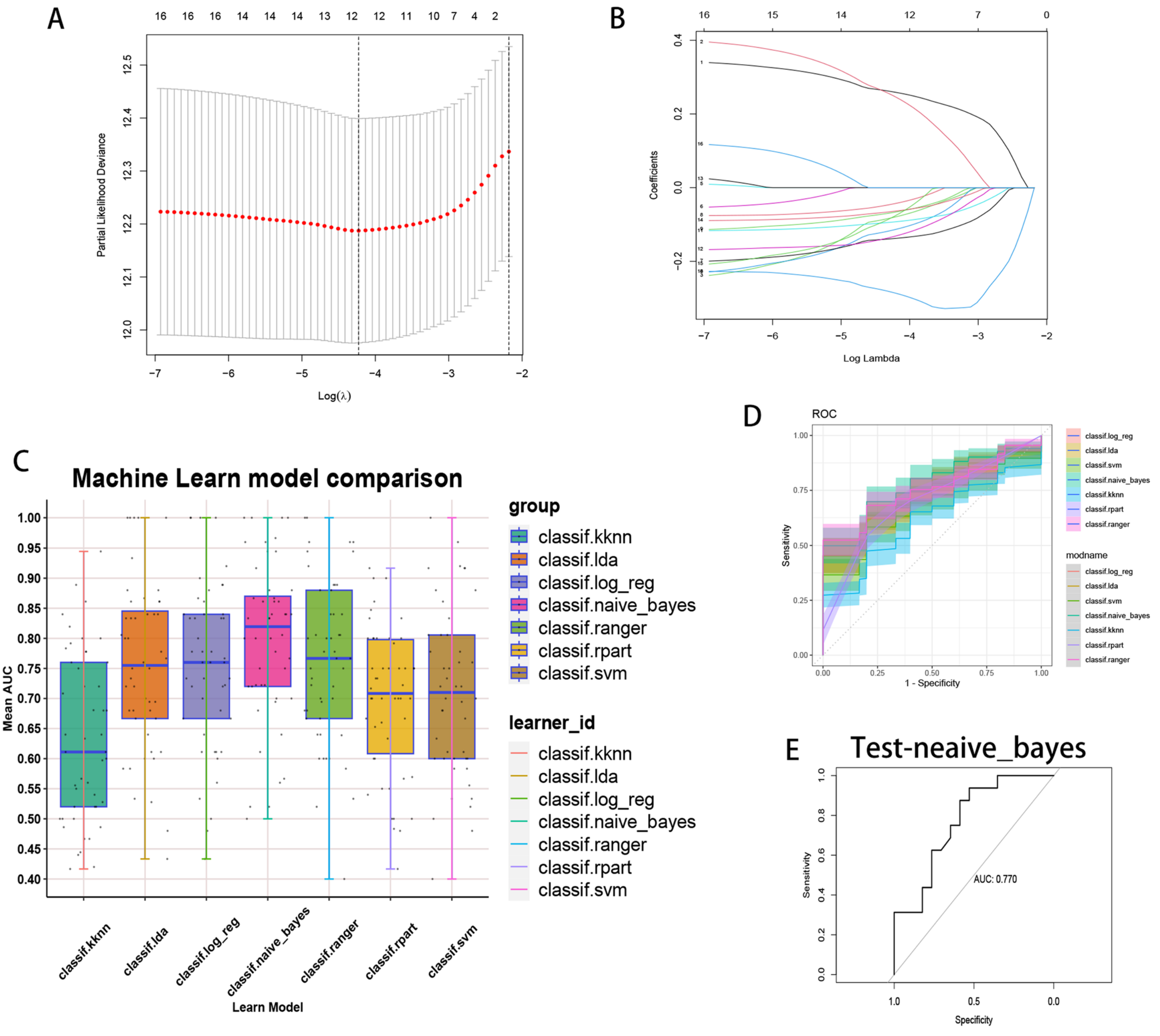


Fig. 7 We constructed a model of clock-related genes using lasso analysis and used 7 machine learning methods, including kknn, lda, log_reg, naive_bayes, ranger, rpart, and svm, for cross-validation on the training set and naive_bayes for external validation to test the model's performance. We plotted the lasso coefficient path through lasso analysis (**A**) and the cross-validation curve through lasso regression analysis (**B**). Cross-validation was performed using 7 machine learning algorithms (**C**), and ROC curves were plotted to compare AUC (**D**). We selected the naive_bayes method for testing on the external validation set and obtained the ROC curve (**E**)

3.4 Analysis of hdWGCNA in epithelial cells

To explore the key role of epithelial cells, we conducted hdWGCNA analysis to identify potential markers of epithelial cells. After setting the soft threshold to 8, we identified 5 modules (Fig. 5A). As shown in Fig. 5B and C, a total of 6 gene modules were obtained, and the 10 most influential genes were listed according to hdWGCNA. Additionally, the yellow module exhibited a strong positive correlation with the brown module (Fig. 5D). Moreover, the UMAP plot displayed the distribution of the turquoise and green modules in epithelial cells, which overlapped significantly with the epithelial cells in subcluster 5 (Fig. 5E). Interestingly, we found that the turquoise and blue modules were highly expressed in the epithelial cells of subcluster 5 (Fig. 5F). Therefore, we propose that the turquoise and green modules may represent

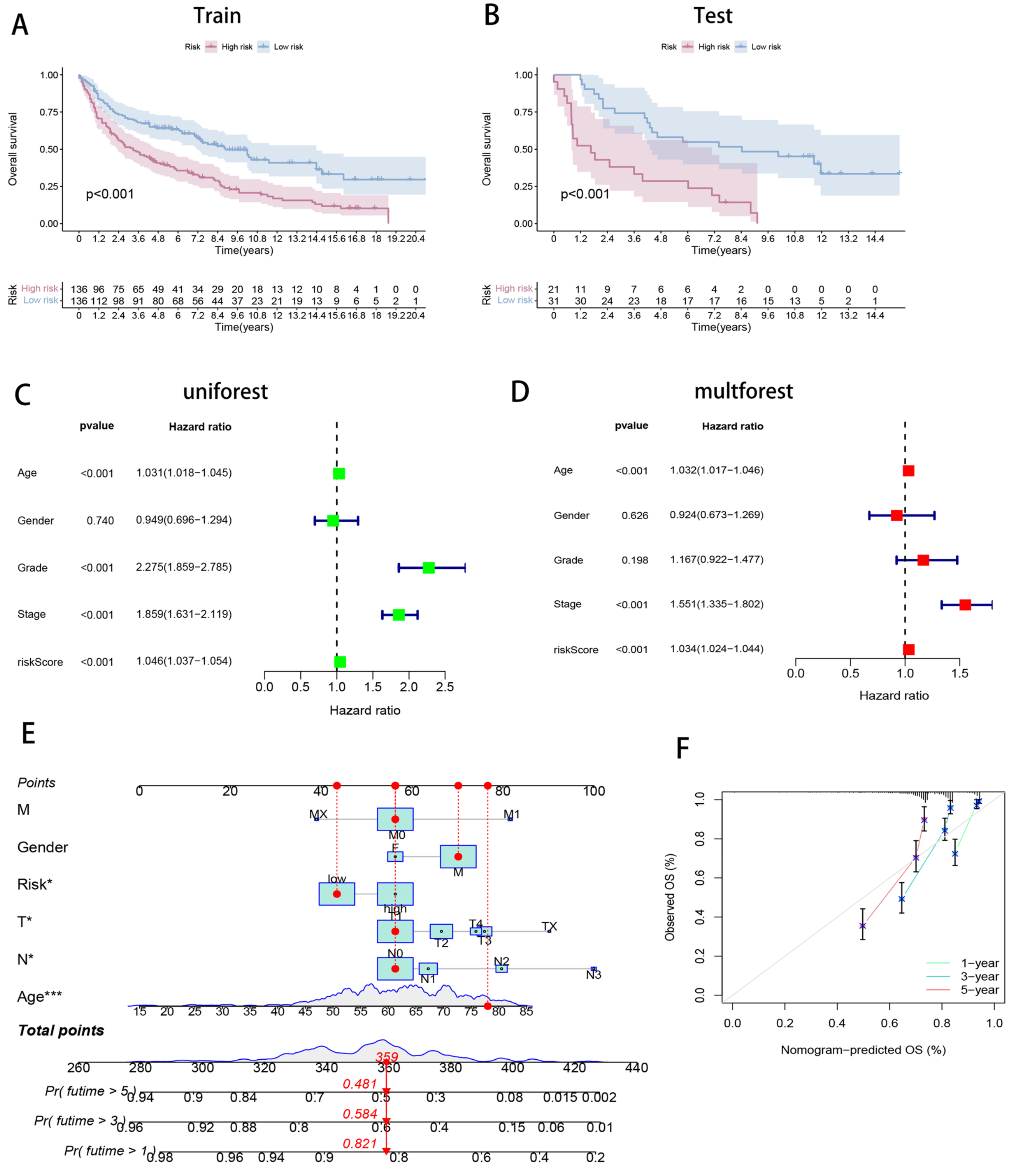


Fig. 8 We established a nomogram of clinical features through survival analysis and single- and multi-factor Cox analysis. We plotted the survival curves for the Train group (**A**) and Test group (**B**). Additionally, we plotted the forest plots for single-factor (**C**) and multi-factor (**D**) analyses. Furthermore, we developed a nomogram for various clinical features (**E**) and calibration curves for 1-year, 3-year, and 5-year survival periods (**F**)

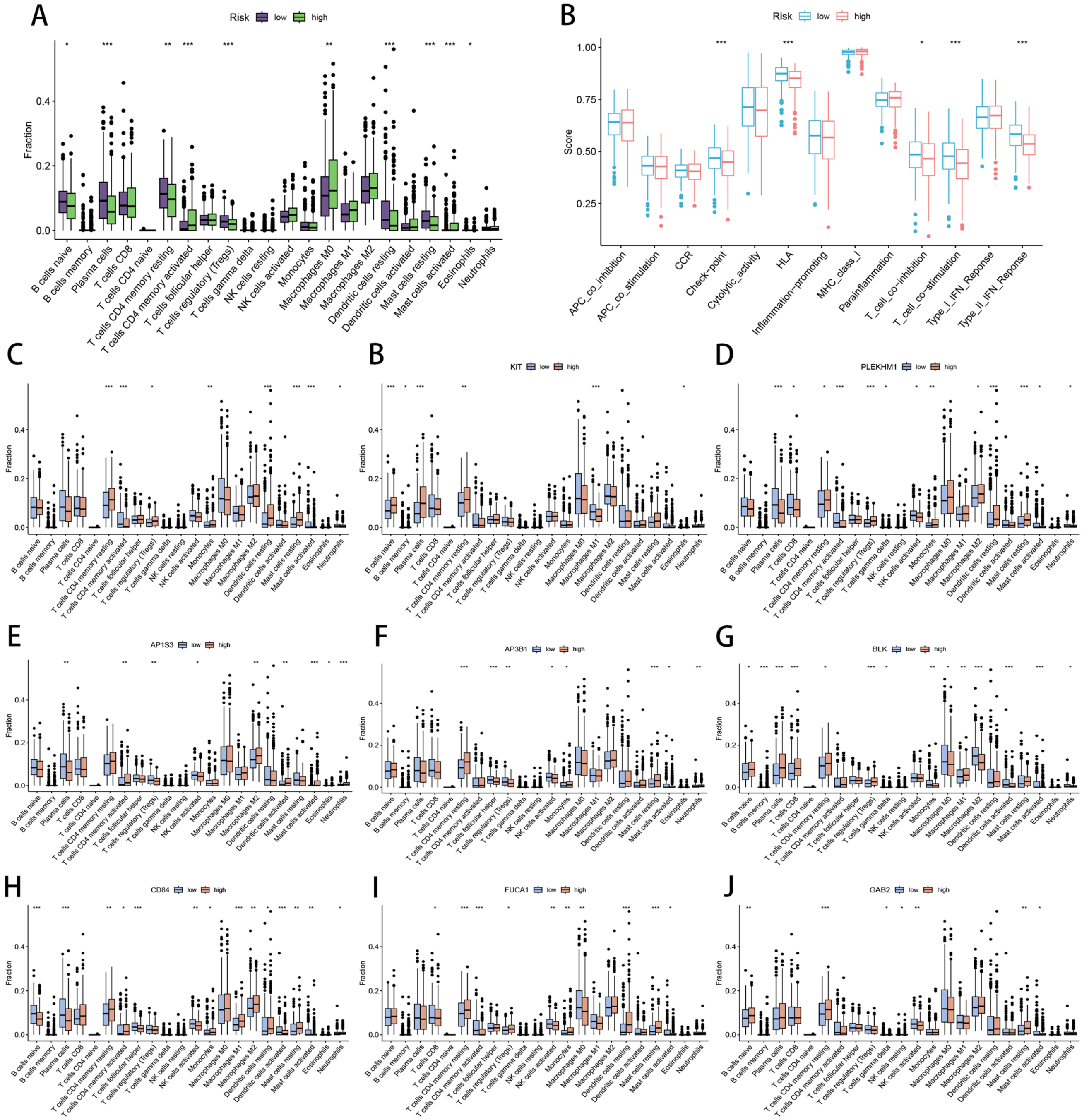


Fig. 9 We conducted immune infiltration analysis on the model genes. We plotted boxplots of immune cells (**A**) and immune functions (**B**) for the high- and low-risk groups. Additionally, we plotted boxplots of individual genes, including GGA2 (**C**), KIT (**D**), PLEKHM1 (**E**), AP1S3 (**F**), BLK (**G**), CD84 (**H**), FUCA1 (**I**), and GAB2 (**J**), showing their immune infiltration in the high- and low-risk groups

characteristics of NSCLC cells. Ninety genes from the turquoise and green modules are considered potential characteristic biomarkers of NSCLC.

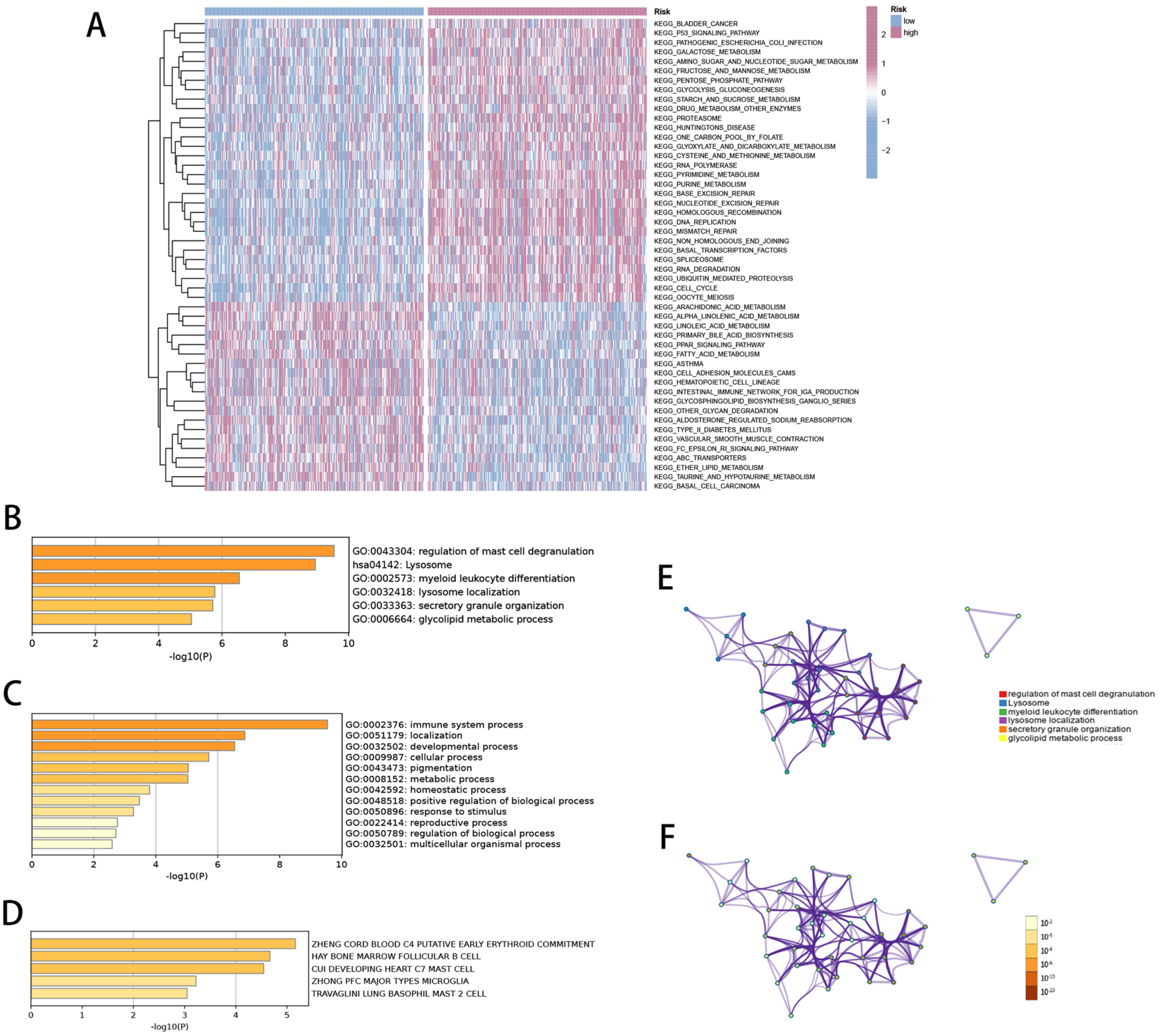


Fig. 10 We conducted enrichment analysis on the model genes. We plotted a heatmap of GSEA enrichment (**A**). We used the Metscape website (https://metascape.org/) for enrichment analysis of the 12-LDCDRGs. We obtained bar plots of GO enrichment, color-coded by p-value (**B**, **C**). Additionally, we summarized the enrichment analysis of cell type labels (**D**). We also created a network of enriched terms: colored by cluster ID, where nodes that share the same cluster ID are typically close to each other (**E**); colored by p-value, where terms containing more genes tend to have a more significant p-value (**F**)

3.5 Proposed temporal analysis of epithelial cells

To identify the characteristics of epithelial cell marker genes at different developmental stages, we conducted pseudotime analysis. Cells with similar states were grouped together, and branch points divided the cells into different states. Notably, epithelial cells in clusters 2, 3, and 4 were primarily located at the end of the pseudotime trajectory (Fig. 6A, B). Additionally, the changes in potential marker genes during differentiation were detected based on the gene expression levels of different epithelial cell subclusters (Fig. 6C). To further explore the functional roles behind the regulation of NSCLC by epithelial cells, we used enrichment analysis to investigate the module genes identified by hdWGCNA (Fig. 6D).

Fig. 11 We applied PPI to filter core genes and conducted single-gene survival analysis. We collected protein-protein interaction data from the STRING website (https://string-db.org/) and visualized the topological structure of the PPI network using Cytoscape software (**A**). We selected the top 10 core genes using the Degree method, with redder color indicating richer interactions with other proteins (**B**). Additionally, we plotted Kaplan-Meier survival curves for the core genes KIT (**C**), BLK (**D**), AP3B1 (**E**), GAB2 (**F**), CD84 (**G**), GGA2 (**H**), GATA2 (**I**), FUCA1 (**J**), AP1S3 (**K**), and PLEKHM1 (**L**) ▶

We found that epithelial cells play major roles in pathways such as enzyme binding, enzyme inhibitor activity, protein binding, and oxidoreductase activity, suggesting their potential impact on protein functions in NSCLC, particularly affecting enzyme-mediated energy metabolism processes.

3.6 Machine learning reveals prognostic value of epithelial cells

To further explore the clinical value of epithelial cell marker genes, we obtained 585 NSCLC samples from the TCGA database and 272 NSCLC patient samples from GEO. First, we obtained lysosomal autophagy-related genes by LASSO regression analysis of 90 genes from the hdWGCNA module by taking the intersection set with the lysosome-dependent death gene set, and 12 marker genes related to lysosome-dependent death of epithelial cells were obtained by LASSO regression screening (Fig. 7A, B). Next, we used seven machine learning methods to model the 12 genes and compared their performance. The results showed that the naive_bayes model had the largest AUC area compared to other machine learning methods, proving that analyzing the 12 selected genes using the naive_bayes model provides the most prognostic value (Fig. 7C). We further demonstrated the prognostic value of the naive_bayes model by plotting ROC curves (Fig. 7D and E).

In this step, we integrated transcriptome samples and used machine learning to establish a robust 12-gene NSCLC prognostic model to explore the clinical value of single-cell analysis. To evaluate the accuracy of the prognostic model, we plotted survival curves and performed calibration curve analysis at different time points. The results consistently indicated that the survival rate of the high-risk group was significantly lower than that of the low-risk group at various time points, with $p < 0.05$ indicating statistical significance. The constructed model effectively distinguished between high-risk and low-risk groups (Fig. 8A, B, and F). Additionally, calibration curves at different time points showed high predictive accuracy of the model.

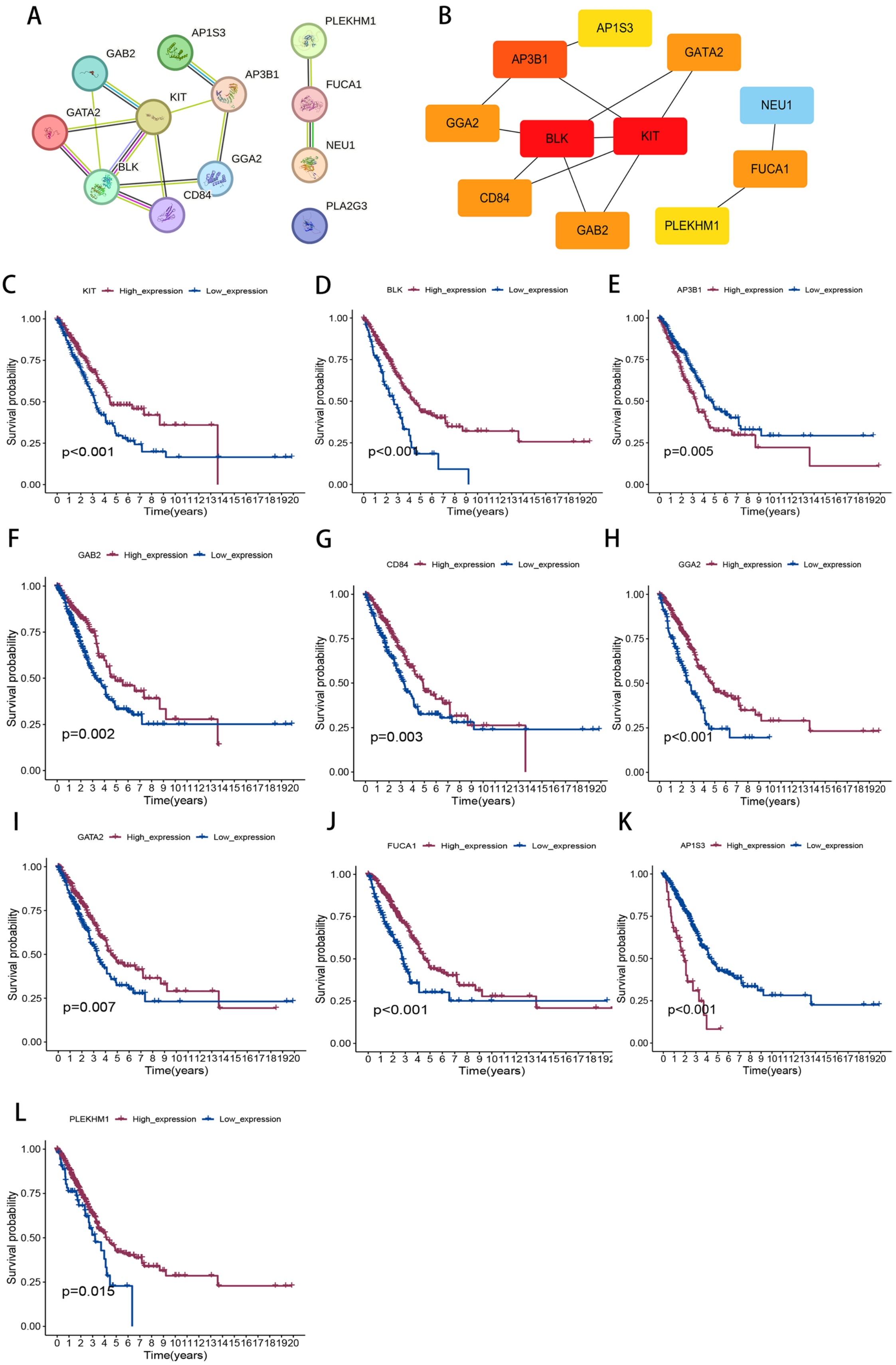
We conducted univariate and multivariate analyses to assess the impact of different clinical indicators on outcomes (Fig. 8C and D), and the results indicated that tumor stage was a significant risk factor. We also created a nomogram to visualize the results of our model (Fig. 8E).

3.7 Immune cell infiltration and analysis of the tumour microenvironment

To explore the relationship between the model, immune cell infiltration, and the tumor microenvironment, we performed deconvolution analysis using the CIBERSORT algorithm. This analysis revealed significant differences between the high-risk and low-risk groups in terms of plasma cells, macrophages M2, resting CD4 memory T cells, macrophages M0, CD8 T cells, naive B cells, Tregs, activated CD4 memory T cells, monocytes, and resting mast cells (Fig. 9A). Additionally, there were differences in immune functions, such as type II IFN response, between the two risk groups (Fig. 9B).

We also analyzed the 12 feature genes included in the model. Each gene was divided into high and low expression groups, and the differences in immune cells between these groups were observed to reveal the immune functions of these genes (Fig. 9C–J). Furthermore, we identified pathway differences between the high-risk and low-risk groups through KEGG enrichment analysis, which showed significant differences between the two groups. GO enrichment analysis indicated functional differences in regulating mast cell degranulation, immune processes, and bone marrow follicular B cells, highlighting the importance of immune function differences between the two groups (Fig. 10A–F).

Finally, we depicted the association of the 12 feature genes through PPI network analysis, with the results showing that BLK and KIT genes held key positions among all the genes (Fig. 11A, B). Additionally, survival analysis of individual genes revealed the crucial roles of these 12 key genes in tumor prognosis (Fig. 11C–K).


A
PLEKHM1
AP1S3
GAB2
AP3B1
FUCA1
KIT
GATA2
NEU1
BLK
GGA2
CD84
PLA2G3
B
AP1S3
AP3B1
GATA2
NEU1
GGA2
BLK
KIT
FUCA1
CD84
GAB2
PLEKHM1
C
KIT High_expression Low_expression
p<0.001
D
BLK High_expression Low_expression
p<0.001
E
AP3B1 High_expression Low_expression
p=0.005
F
GAB2 High_expression Low_expression
p=0.002
G
CD84 High_expression Low_expression
p=0.003
H
GGA2 High_expression Low_expression
p<0.001
I
GATA2 High_expression Low_expression
p=0.007
J
FUCA1 High_expression Low_expression
p<0.001
K
AP1S3 High_expression Low_expression
p<0.001
L
PLEKHM1 High_expression Low_expression
p=0.015
Survival probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Time(years)

4 Discussion

Non-small cell lung cancer (NSCLC) is a complex disease involving the interplay of genetic background and environmental factors, and is associated with various abnormalities such as metabolic dysregulation and apoptosis. Epithelial cells play a crucial role in the progression of NSCLC. They act as surface barriers and perform secretory functions, with some epithelial cells undergoing morphological changes in response to stimuli as a reaction to external disturbances. This process, known as epithelial-mesenchymal transition (EMT), is significant for cancer processes such as tumor invasion and metastasis. Through EMT, epithelial cells acquire invasive potential, transforming into migratory mesenchymal cells associated with tumor cell invasion [10]. EMT also leads to the upregulation of anti-apoptotic signals, making the cells more tumorigenic and less responsive to treatment. Concurrently, extracellular vesicles released by NSCLC cells drive the invasion and permeability of non-tumorigenic lung epithelial cells [11].

In this study, we described the communication of epithelial cells through single-cell analysis. We found that epithelial cells predominantly output signals rather than receive them in cell communication, with a focus on the MK and ANNEXIN signaling pathways. This suggests that epithelial cells have a unique communication pattern in the progression of NSCLC. In the hdWGCNA analysis, we identified modules most associated with NSCLC epithelial cells, thereby obtaining functional gene groups. Enrichment analysis revealed that these genes are concentrated in functions such as enzyme-linked reactions, protein maturation, and extracellular vesicles, highlighting the critical role of epithelial cells in protein synthesis and enzymatic reactions.

Using machine learning, we constructed a prognostic model comprising 12 lysosomal-dependent death genes. Studies have reported that the BLK gene is ectopically expressed in various malignancies, including breast cancer, kidney cancer, and lung cancer, and may serve as a potential therapeutic target [12]. The activation of CD84 leads to the expression of PD-L1, inhibiting T cell function and acting as a regulator of the immunosuppressive microenvironment [13]. Knockdown of FUCA1 significantly alleviates p53-dependent, chemotherapy-induced apoptotic death [14]. Silencing GAB2 suppresses the proliferation and invasion of NSCLC cells [15], and regulating KIT expression reduces NSCLC cell resistance to cisplatin [16]. All of these lysosomal-dependent death genes are closely associated with cancer development, regulation of the tumor microenvironment, while other marker genes were newly associated with NSCLC in our study. Enrichment analysis of all marker genes indicated that these genes are mainly involved in various immune processes, including mast cell degranulation.

Collectively, our results suggest that epithelial cells influence NSCLC progression by affecting multiple immune processes and protein changes. Characterizing the tumor microenvironment (TME) at single-cell resolution can provide insights into potential novel therapeutic targets. By integrating transcriptome multi-modal analysis and combining single-cell and tissue transcriptomics, we aim to identify key mechanisms within the complex network of NSCLC. Understanding the extent to which tumor cells shape their microenvironment and how the microenvironment influences tumor cells is crucial, as these mechanisms determine the clinical response to targeted or immunotherapy. In the future, single-cell methods incorporating spatial information and surface protein expression will help complete this picture [17].

In summary, we first identified a critical role for epithelial cells in non-small cell lung cancer through lysosome-dependent death. On one hand, we did not thoroughly explore the expression patterns of these genes in epithelial cells. Additionally, we lack a large clinical cohort to investigate the diagnostic value of these characteristic genes. Nevertheless, our findings may provide new insights into the prognosis and tumor microenvironment of NSCLC. Further research on the specific mechanisms and regulatory pathways of epithelial cell-mediated immunity in NSCLC development will enhance our understanding of the pathogenesis of NSCLC [18–20].

5 Conclusion

This study constructed a prognostic model for NSCLC based on LDCD scoring by integrating single-cell RNA sequencing and machine learning techniques. This model can effectively predict the prognosis of NSCLC patients, providing an important basis for precision therapy and rational medication. Additionally, the results indicate the critical role of monocytes in NSCLC progression, providing a theoretical foundation for the future development of novel therapeutic targets.

Author contributions JF, YC, JL, MT and FW conceived the study. RL, JW, GW, YR, FW, YG, MB, PWand FW drafted the manuscript. JF, YC, JL and FW performed the literature search and collected the data. JF, YC, JL and MT analyzed and visualized the data. YR, FW, YG, MB and FW helped with the final revision of this manuscript. All authors reviewed and approved the final manuscript.

Funding The Scientific Research Fund of Technology Bureau in Dazhou (22ZDYF0020), the Scientific Research Fund of Technology Bureau in Dazhou (22YYJC0014).

Data availability Single-cell RNA sequencing (scRNA-seq) data from patients were retrieved from the Gene Expression Omnibus (GEO) database. Batch RNA-seq data were obtained from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) and GEO database.

Declarations

Competing interests The authors declare that this research was conducted in the absence of any commercial or fifinancial relationships that could be construed as a potential conflflict of interest.

References

1. Bloom GS. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. JAMA Neurol. 2014;71:505–8.
2. Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clin Proc. 2019;94:1623–40.
3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71:7–33.
4. Huang MY, Jiang XM, Wang BL, Sun Y, Lu JJ. Combination therapy with PD-1/PD-L1 blockade in non-small cell lung cancer: strategies and mechanisms. Pharmacol Ther. 2021;219:107694.
5. Aits S, Jäättelä M. Lysosomal cell death at a glance. J Cell Sci. 2013;126:1905–12.
6. Kundu ST, Grzeskowiak CL, Fradette JJ, Gibson LA, Rodriguez LB, Creighton CJ, Scott KL, Gibbons DL. TMEM106B drives lung cancer metastasis by inducing TFEB-dependent lysosome synthesis and secretion of cathepsins. Nat Commun. 2018;9:2731.
7. Huang WZ, Luo MZ. A cellulose acetate membrane counter immunoelectrophoresis test for identification of the host source of mosquito blood meals. Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 1986;4:186–8.
8. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwé H, Pircher A, Van den Eynde K, Weynand B, Verbeken E, De Leyn P, Liston A, Vansteenkiste J, Carmeliet P, Aerts S, Thienpont B. Phenotype molding of stromal cells in the lung tumor microenvironment. Nat Med. 2018;24:1277–89.
9. Xu J, Gou S, Huang X, Zhang J, Zhou X, Gong X, Xiong J, Chi H, Yang G. Uncovering the impact of aggrephagy in the development of Alzheimer's disease: insights into diagnostic and therapeutic approaches from machine learning analysis. Curr Alzheimer Res. 2023;20:618–35.
10. Thiery JP. Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer. 2002;2:442–54.
11. Hasan H, Sohal IS, Soto-Vargas Z, Byappanahalli AM, Humphrey SE, Kubo H, Kitdumrongthum S, Copeland S, Tian F, Chairoungdua A, Kasinski AL. Extracellular vesicles released by non-small cell lung cancer cells drive invasion and permeability in non-tumorigenic lung epithelial cells. Sci Rep. 2022;12:972.
12. Petersen DL, Berthelsen J, Willerslev-Olsen A, Fredholm S, Dabelsteen S, Bonefeld CM, Geisler C, Woetmann A. A novel BLK-induced tumor model. Tumour Biol. 2017;39:1010428317714196.
13. Lewinsky H, Gunes EG, David K, Radomir L, Kramer MP, Pellegrino B, Perpinial M, Chen J, He TF, Mansour AG, Teng KY, Bhattacharya S, Caserta E, Troadec E, Lee P, Feng M, Keats J, Krishnan A, Rosenzweig M, Yu J, Caligiuri MA, Cohen Y, Shevetz O, Becker-Herman S, Pichiorri F, Rosen S, Shachar I. CD84 is a regulator of the immunosuppressive microenvironment in multiple myeloma. JCI Insight. 2021;6:e141683.
14. Baudot AD, Crighton D, O'Prey J, Somers J, Gonzalez PS, Ryan KM. p53 directly regulates the glycosidase FUCA1 to promote chemotherapy-induced cell death. Cell Cycle. 2016;15:2299–308.
15. Yu S, Geng S, Hu Y. Mir-486-5p inhibits cell proliferation and invasion through repressing GAB2 in non-small cell lung cancer. Oncol Lett. 2018;16:3525–30.
16. Li P, Ma L, Zhang Y, Ji F, Jin F. MicroRNA-137 down-regulates KIT and inhibits small cell lung cancer cell proliferation. Biomed Pharmacother. 2014;68:7–12.
17. Fan T, Jiang L, Zhou X, Chi H, Zeng X. Deciphering the dual roles of PHD finger proteins from oncogenic drivers to tumor suppressors. Front Cell Dev Biol. 2024;12:1403396.
18. Chen Y, et al. Global insights into rural health workers' job satisfaction: a scientometric perspective. Front Public Health. 2022;10:895659. https://doi.org/10.3389/fpubh.2022.895659.

19. Chen Y, et al. Systematic and meta-based evaluation on job satisfaction of village doctors: an urgent need for solution issue. Front Med. 2022;9:856379. https://doi.org/10.3389/fmed.2022.856379.
20. You Y, et al. Mediation role of recreational physical activity in the relationship between the dietary intake of live microbes and the systemic immune-inflammation index: a real-world cross-sectional study. Nutrients. 2024;16:6777. https://doi.org/10.3390/nu16060777.