Original Article

Neutrophil estimation and prognosis analysis based on existing lung squamous cell carcinoma datasets: the development and validation of a prognosis prediction model

Youyu Wang[1#], Dongfang Li[2#], Qiang Li[3], Alina Basnet[4], Jimmy T. Efird[5,6], Nobuhiko Seki[7]

[1]Department of Thoracic Surgery, The Second People's Hospital of Shenzhen, The First Affiliated Hospital of Shenzhen University, Shenzhen, China; [2]Department of Thoracic Surgery, Shenzhen Hospital of Southern Medical University, Shenzhen, China; [3]Department of Oncology, Shenzhen Hospital of Southern Medical University, Shenzhen, China; [4]Division of Hematology-Oncology, Upstate Cancer Center, Upstate Medical University, Syracuse, NY, USA; [5]VA Cooperative Studies Program Coordinating Center, Boston, MA, USA; [6]Department of Radiation Oncology, Case Western Reserve University School of Medicine, Cleveland, OH, USA; [7]Division of Medical Oncology, Department of Internal Medicine, Teikyo University School of Medicine, Tokyo, Japan

Contributions: (I) Conception and design: Y Wang, D Li; (II) Administrative support: Q Li; (III) Provision of study materials or patients: Y Wang; (IV) Collection and assembly of data: Y Wang, D Li; (V) Data analysis and interpretation: Y Wang, D Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Qiang Li, MD. Department of Oncology, Shenzhen Hospital of Southern Medical University, No. 1333 Xinhu Road, Bao'an District, Shenzhen 518000, China. Email: 110184356@qq.com.

Background: Notwithstanding the rapid developments in precision medicine in recent years, lung cancer still has a low survival rate, especially lung squamous cell cancer (LUSC). The tumor microenvironment (TME) plays an important role in the progression of lung cancer, in which high neutrophil levels are correlated with poor prognosis, potentially due to their interactions with tumor cells via pro-inflammatory cytokines and chemokines. However, the precise mechanisms of how neutrophils influence lung cancer remain unclear. This study aims to explore these mechanisms and develop a prognosis predictive model in LUSC, addressing the knowledge gap in neutrophil-related cancer pathogenesis.

Methods: LUSC datasets from the Xena Hub and Gene Expression Omnibus (GEO) databases were used, comprising 473 tumor samples and 195 tumor samples, respectively. Neutrophil contents in these samples were estimated using CIBERSORT, xCell, and microenvironment cell populations (MCP) counter tools. Differentially expressed genes (DEGs) were identified using DEseq2, and a weighted gene co-expression network analysis (WGCNA) was performed to identify neutrophil-related genes. A least absolute shrinkage and selection operator (LASSO) Cox regression model was constructed for prognosis prediction, and the model's accuracy was validated using Kaplan-Meier survival curves and time-dependent receiver operating characteristic (ROC) curves. Additionally, genomic changes, immune correlations, drug sensitivity, and immunotherapy response were analyzed to further validate the model's predictive power.

Results: Neutrophil content was significantly higher in adjacent normal tissue compared to LUSC tissue ($P<0.001$). High neutrophil content was associated with worse overall survival (OS) ($P=0.02$), disease-free survival (DFS) ($P=0.02$), and progression-free survival (PFS) ($P=0.03$) using different software estimates. Nine gene modules were identified, with blue and yellow modules showing strong correlations with neutrophil prognosis ($P<0.001$). Eight genes were selected for the prognostic model, which accurately predicted 1-, 3-, and 5-year survival in both the training set [area under the curve (AUC) value =0.60, 0.63, 0.66, respectively] and validation set (AUC value =0.58, 0.58, 0.59, respectively), with significant prognosis differences between high- and low-risk groups ($P<0.001$). The model's independent prognostic factors included risk group, pathologic M stage, and tumor stage ($P<0.05$). A further molecular mechanism analysis revealed differences between risk groups were revealed in immune checkpoint and human leukocyte antigen (HLA) gene expression, hallmark pathways, drug sensitivity, and immunotherapy responses.

Conclusions: This study established a risk-score model that effectively predicts the prognosis of LUSC patients and sheds light on the molecular mechanisms involved. The findings enhance the understanding of neutrophil-tumor interactions, offering potential targets for personalized treatments. However, further experimental validation and clinical studies are required to confirm these findings and address study limitations, including reliance on public databases and focus on a specific lung cancer subtype.

Keywords: Neutrophils; lung cancer; lung squamous cell carcinoma (LUSC); prognosis; public databases

Submitted May 08, 2024. Accepted for publication Aug 06, 2024. Published online Aug 19, 2024.
doi: 10.21037/tlcr-24-411
View this article at: https://dx.doi.org/10.21037/tlcr-24-411

Introduction

With an incidence of 11.4%, lung cancer is the second most commonly diagnosed cancer after female breast cancer and is the leading cause of cancer-related death worldwide (1,2). Even with significant advancements in diagnosis and treatment, the overall survival (OS) rate of lung cancer patients remains poor, especially lung squamous cell cancer (LUSC). Several prognostic biomarkers and prediction models for LUSC have been identified, including genetic mutations, expression levels of specific proteins, and various gene signatures (3-5). These models have been used to predict patient outcomes and guide treatment decisions. Researchers have even begun to explore the cell-free DNA methylation profile of LUSC in order to evaluate its potential in the diagnosis of LUSC (6). However, many of these biomarkers and models have limitations, such as variability in predictive power across different populations, high cost, and the need for complex technological platforms. A recent study has shown that the tumor microenvironment (TME) plays a crucial role in the development and progression of lung cancer (7). Among the various components of the TME, neutrophils have emerged as important players in the pathogenesis of lung cancer. Neutrophils are the most abundant type of white blood cell and are believed to have a critical role in innate immune response. However, their role in cancer is complex and multifaceted (8,9). Therefore, it is necessary to explore the molecular mechanisms of neutrophils in lung cancer and develop a predictive model that can effectively predict the prognosis of individual patients.

Highlight box

Key findings

- This study investigated the prognostic implications of neutrophil content in lung squamous cell carcinoma (LUSC). We developed and validated a reliable prognostic model consisting of eight neutrophil-associated genes through a dataset. This model provides a comprehensive approach for predicting the outcomes of LUSC patients.

What is known, and what is new?

- Similar to previous studies, this research highlighted the association between neutrophil content and clinical outcomes in lung cancer. The identification of a prognostic signature involving multiple genes is consistent with the approach used in some previous studies to identify the molecular markers associated with patient prognosis.
- This study used a combination of software tools to estimate neutrophil content and undertook a comprehensive analysis of the molecular mechanisms of LUSC, including gene mutation and immune checkpoint gene expression.

What is the implication, and what should change now?

- This study contributes to the existing literature on the prognostic role of neutrophils in lung cancer by providing valuable insights into the potential prognostic signatures and molecular mechanisms of LUSC. The findings expand understandings of the complex interactions between neutrophils and tumor biology, offering potential targets for personalized treatments.

Several hypotheses have been proposed based on current knowledge of neutrophil biology and tumor immunology. For instance, it has been suggested that neutrophils may interact with tumor cells through the release of pro-inflammatory cytokines and chemokines, which can promote tumor cell proliferation, angiogenesis, and immune evasion (10). However, the precise mechanism underlying the association between neutrophils and lung cancer has yet to be fully elucidated.

In recent years, high-throughput sequencing technologies and bioinformatics tools have revolutionized the field of cancer research. These technologies enable the analysis of large-scale genomic and transcriptomic data, which can provide insights into the molecular mechanisms underlying cancer development and progression. In the context of lung cancer and neutrophils, several bioinformatics studies have been conducted to identify the differentially expressed genes (DEGs) and pathways associated with neutrophil infiltration in lung tumors (11-13), and they have reported a correlation between high levels of neutrophils and a poor prognosis in lung cancer patients (11-13). However, there is still a need for further research to fully understand the role of neutrophils in lung cancer and to identify potential therapeutic targets.

Given the potential strengths of neutrophils as prognostic markers, including their abundance and involvement in key cancer-related processes, there is a critical need to identify new biomarkers and develop prognostic models based on neutrophils. In this study, we compared the neutrophil content of tumor and normal tissue samples. We then conducted a gene expression analysis and weighted gene co-expression network analysis (WGCNA) to establish and verify a multivariable prediction model of lung cancer prognosis. These findings suggest potential targets for therapeutic interventions and provide insights into the role of neutrophils in lung cancer progression and drug resistance. We present this article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-411/rc).

Methods

Acquisition and screening of data

The LUSC data set and corresponding clinical information [including age, gender, tumor-node-metastasis (TNM) staging, tumor stage, family history, examined lymph node count, neoplasm histologic grade, primary diagnosis, site of resection or biopsy, and disease type] and survival information for each sample were downloaded from the Xena Hub database (https://xenabrowser.net/datapages/?hub=https://gdc.xenahubs.net:443) (14). A total of 550 samples, comprising 501 cancer tissue samples and 49 adjacent tissue samples, were included in the study. Of these samples, 473 cancer tissue samples with prognostic information were used as training sets for subsequent model construction analysis.

The GSE37745 data set, which included 195 tumor samples, was downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) as a validation set for the subsequent model validation analysis (15). We used the pre-processed and normalized probe expression matrix and downloaded the corresponding platform annotation file to convert probes to gene symbols, averaging the values to obtain gene expression values for the subsequent analysis.

The lung cancer gene mutation data from the National Institutes of Health (NIH) squamous carcinoma database (https://gdc.cancer.gov/about-data/publications/PanCan-Squamous-2018) were downloaded to analyze genomic changes between the different prognostic risk groups. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Neutrophil estimation and prognosis analysis

Based on the gene expression levels of the LUSC samples in The Cancer Genome Atlas (TCGA) data set, relative mode and absolute mode were each selected to calculate the proportions of 22 immune cells using CIBERSORT (https://cibersort.stanford.edu/index.php) (16). xCell (https://xcell.ucsf.edu/) was used to estimate the relative abundance of immune cells and stromal cells for each sample by inputting all the messenger RNA (mRNA) expression matrices (17). Microenvironment cell populations (MCP) counter (https://github.com/ebecht/MCPcounter) was used to estimate the relative infiltration abundance of nine immune cells in each sample based on the expression matrix of all the mRNA (18). The content of neutrophils in each tumor sample was obtained using the CIBERSORT, xCell, and MCP software, and the Kaplan-Meier (product-limit) survival curves, which included OS, disease-free survival (DFS) and progression-free survival (PFS), was plotted using the R package (version 3.5-5) (https://cran.r-project.org/web/packages/survival/) (19). The logarithmic-rank test was used to compute P values.

Differential expression gene screening

Based on TCGA gene expression data, the DEseq2 package (version 1.36.0) (https://bioconductor.org/packages/release/bioc/html/DESeq2.html), which provides linear regression and empirical Bayes methods, was used to compare the DEGs between the lung cancer and adjacent normal tissue samples (20). The Benjamini-Hochberg step-up method was used for multiple testing correction, and the adjusted (adj.) P value. The differential expression threshold was set as an adj. P value <0.05 and a |log fold change| >2, and DEGs were evaluated based on the fold change and significance level.

Screening of neutrophil-related drug resistance genes based on a co-expression network

Using the WGCNA (version 1.71) (https://cran.r-project.org/web/packages/WGCNA/index.html) (21), we performed a WGCNA of the top 5,000 genes, selected the scale-free network fit index and average connectivity, and computed and selected β=8 as the soft-threshold for this data set. We also established a model of lung cancer samples versus normal samples, analyzed the expression levels of the hub genes between the two groups, and used the Pearson correlation test to compute the correlations between the module genes and neutrophil compositional phenotypes, screen out the immune defense co-expressed module related to neutrophils, as well as the hub genes in the module, with a screening threshold of abs[datKME(, c)] >0.8 and |gene significance (GS)| >0.1. We performed a Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the module genes for the functional annotation, and set the correlation coefficient threshold to 0.85, with a parameter of β=8, to construct the immune defense gene co-expression network.

Functional enrichment analysis of the GO and KEGG pathways

The clusterProfiler (version 4.4.4) (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) (22) was used to perform the GO biological process (BP) analysis of the neutrophil-related genes from the obtained module, and the enrichR tool (https://mirrors.sjtug.sjtu.edu.cn/cran/web/packages/enrichR/index.html) (23) was used for the KEGG pathway analysis. A P value <0.05 and a minimum gene number of two were set as the criteria for filtering the significantly enriched pathways.

Identification of neutrophil genes associated with prognosis

Based on the DEGs in the neutrophil-related module obtained from the WGCNA, a single-factor Cox regression analysis using the R package (19) was performed on the clinical prognostic information of TCGA-LUSC samples to identify the neutrophil-related genes significantly associated with OS prognosis. A P value <0.05 was set as the threshold for significant correlation filtering.

Construction and validation of a neutrophil-related prognostic model

Based on the neutrophil-related genes significantly associated with the survival prognosis obtained in the previous step, we used the least absolute shrinkage and selection operator (LASSO), Cox regression (penalized maximum likelihood) model in the glmnet package (version 4.1.7) of R language (https://cran.r-project.org/web/packages/glmnet/index.html) (24) to further screen the prognosis-related neutrophil-gene combinations. We used the survival prognosis information of the training set samples and the expression values of the genes in different samples to perform a five-fold cross-validation analysis.

Based on the regression coefficients of the genes in the neutrophil-gene combinations related to prognosis and the expression levels of these genes in TCGA-LUSC samples, we constructed a risk-score model. The following formula was used to compute the risk score:

$$Risk\ score = \sum \beta gene \times Expgene \qquad [1]$$

where βgene represents the LASSO regression coefficient of the gene, and Expgene represents the expression level of the gene in the TCGA data set.

To validate the accuracy of the model, the risk-score calculation formula was used with the same regression coefficients to estimate the risk-score value of each sample in the GEO data set. Based on the median risk-score value, all the GEO samples were divided into a high-risk (a risk score ≥ median value) group and a low-risk (a risk score < median value) group. The Kaplan-Meier curve plotting method in the R survival package was used to evaluate the difference in the survival prognosis between the high- and low-risk groups. Furthermore, high-risk or low-risk score distribution curves, survival distribution graphs, and gene expression heatmap models were plotted based on the above grouping. Similarly, the Kaplan-Meier survival curve was plotted using the "survminer" package in R software to compare the OS of the two groups. The "timeROC" package was used to draw time-dependent receiver operating characteristic (ROC) curves and compute the areas under the curves (AUCs) of the sample OS at 1-, 3-, and 5-year to evaluate the ability of the model to predict patient prognosis. The higher the AUC value, the better the model's performance. Generally, if the AUC value of the

ROC curve is above 0.6, the model is considered to meet the requirements.

Independent prognostic factor screening

According to the grouping method mentioned above, TCGA samples were divided into two groups; that is, the high- and low-risk groups. The chi-square test in R language was used to statistically compare and analyze the following factor variables: age, risk group, gender, pathologic M, pathologic N, pathologic T, and tumor stage. For the continuous variable age, the samples were grouped based on age >60 years, and the intergroup *t*-test was used to compare the significant differences between the two groups of samples. Single-factor and multivariable Cox analyses were performed based on the risk group and clinical information to select the independent factors.

Nomogram construction

To determine whether the risk-score model mentioned above could be used as an independent prognostic factor, a single-factor Cox regression analysis was performed on age, risk group, gender, pathologic M, pathologic N, pathologic T, and tumor stage separately. Variables with a P value <0.05 were included in the multivariable Cox regression analysis. A further screening was conducted to select variables with a P value <0.05 and draw a nomogram. In addition, a calibration curve was plotted to assess the accuracy of the model.

Analysis of the genomic changes between the different risk groups

Based on the lung cancer gene mutation data in the NIH database, the maftools package (version 2.12.0) (https://bioconductor.org/packages/release/bioc/html/maftools.html) (25) in R language was used to identify the top 20 mutated genes by mutation frequency and compute the tumor mutation burden (TMB) of the tumor samples. The TMB difference between the different risk groups was compared. Based on the median TMB value, TCGA samples were divided into a TMB-high group (a TMB ≥ median value) group and a TMB-low group (a TMB < median value), and the Kaplan-Meier curve plotting method in the R survival package was used to evaluate the survival prognosis difference between the TMB-high and TMB-low groups.

Prognostic model gene and immune correlation analysis

The expression data of immune checkpoint genes, such as *PD1* (*PDCD1*), *PD-L1* (*CD274*), *CTLA-4* (*CTLA4*), *CD278* (*ICOS*), *TIM3* (*HAVCR2*), *LAG3*, *CD47*, *BTLA*, *TIGIT*, *MYD1* (*SIRPA*), *OX40* (*TNFRSF4*), *4-1BB* (*TNFRSF9*), and *B7-H4* (*VTCN1*), and human leukocyte antigen (HLA) family genes were obtained from TCGA expression. The non-parametric Deuchler-Wilcoxon test was used to compare the expression differences of immune checkpoint genes and HLA family genes between the high- and low-risk groups mentioned above.

Molecular mechanism analysis

A molecular mechanism analysis was conducted using the R package gene set variation analysis (GSVA) (version 1.46.0) (https://www.bioconductor.org/packages/release/bioc/html/GSVA.html) (26) for the pathway enrichment analysis. The hallmark gene set from the Molecular Signatures Database (MSigDB) database (version 1.46.0) (http://software.broadinstitute.org/gsea/msigdb/index.jsp) (27) was used as the enrichment background, and the single-sample gene set enrichment analysis method was used for the pathway enrichment analysis between the high- and low-risk groups, with a significant enrichment threshold set at a false discovery rate (FDR) <0.05. R package clusterProfiler (version 4.4.4) (clusterProfiler) was used for the pathway enrichment analysis between the high- and low-risk groups, with MSigDB database c2.cp.kegg pathway as the enrichment background, and the significant enrichment threshold set at a FDR <0.05.

Prediction of chemotherapy sensitivity and immunotherapy responses

The sensitivity of patients to chemotherapy drugs was evaluated using the Cancer Drug Sensitivity Genomics (https://www.cancerrxgene.org/) database. The half-maximal inhibitory concentration (IC_{50}) was quantified using the pRRophetic package (version 0.5) in R package (https://osf.io/5xvsg/) (28). The Deuchler-Wilcoxon test was used to compare differences in drug sensitivity between the high- and low-risk groups.

The immunotherapy response was determined by analyzing tumor immune dysfunction and Tumor Immune Dysfunction and Exclusion (TIDE) (https://github.com/jingxinfu/TIDEp) (29). TIDE (http://tide.dfci.harvard.

edu/) is an analytical technique that uses tumor immune evasion mechanisms to predict immunotherapy response. Specifically, this analysis technique combines differences and correlations of immune therapy predictive markers, such as CD8A/PD-L1 expression levels, cytolytic activity (CYT) score, microsatellite instability (MSI), immune cell proportion score (IPS), and tertiary lymphoid structure. The flowchart of the method is shown in Figure S1.

Statistical analysis

R software is used for statistical analysis of the data. The non-parametric Deuchler-Wilcoxon test was used to compare two groups of samples. P<0.05 is considered statistically significant.

Results

Neutrophil estimation and prognosis analysis

The differences in the neutrophil content between the lung cancer tissue samples and adjacent normal tissue samples were compared, and the results showed that the neutrophil content in the normal samples estimated by CIBERSORT, xCell, or MCP software was significantly higher than that in the tumor tissue samples (P<0.001) (*Figure 1A*).

Using the neutrophil content obtained from CIBERSORT, xCell, or MCP software for each tumor sample, combined with clinical information on OS, DFS, and PFS, Kaplan-Meier (product-limit) curves were plotted with a truncation value of Kaplan-Meier and compared using the Mantel logarithmic-rank test. The results showed that in patients with LUSC, compared with a low neutrophil content, a high neutrophil content estimated by MCP software was significantly associated with worse OS (P=0.02), while a high neutrophil content estimated by CIBERSORT software was significantly associated with worse DFS (P=0.02) and PFS (P=0.03) (*Figure 1B*).

Neutrophil-related drug resistance gene screening, functional enrichment, and identification of neutrophil prognostic signatures

Based on TCGA data set gene expression profiles, we obtained 9,603 DEGs by comparing the gene expression levels between the lung cancer and adjacent normal tissue samples (Figure S2). According to the WGCNA of the top 5,000 genes, we identified nine modules related to the subtyping by constructing an immune defense gene co-expression network (*Figure 2A*). We analyzed the correlation between these modules and the neutrophil content obtained from MCP, CIBERSORT, and xCell software. As *Figure 2B* shows, the blue and yellow modules were significantly associated with neutrophil prognosis, and the correlation was most significant in the MCP predicted scores (blue module: r=0.71, P<0.001; yellow module: r=0.77, P<0.001). Moreover, the blue and yellow modules showed a significant positive correlation with the neutrophil subgroup in the immune cells (*Figure 2C*). To investigate the co-expression regulatory network of the neutrophil defense in tumor development, we analyzed the blue and yellow modules with positive correlations and performed an enrichment analysis based on the genes included in each module while annotating the related pathways (GO BPs) of each module (Figure S3).

Construction and validation of neutrophil-related prognostic signature

According to the genes identified from the blue and yellow modules related to neutrophils by the WGCNA, we performed a single-factor Cox regression analysis and selected 157 genes. These genes were sorted based on their significance P value; *Figure 3A* shows the top 10 ranked genes (i.e., *CCDC68*, *TGM2*, *RETN*, *CSF2*, *FGG*, *APOH*, *FGA*, *SLC22A3*, *CHIA*, and *TERM2*).

Based on the 157 neutrophil-related genes that were significantly related to survival and prognosis obtained above, combined with their expression values in TCGA samples and the survival time and survival state of the samples, the optimal eight characteristic gene combinations and the prognostic regression coefficients were screened by a LASSO Cox regression algorithm. These eight genes were *CSF2*, *EPDR1*, *AOC1*, *CCDC68*, *FGA*, *TGM2*, *RETN*, and *FGG* (*Figure 3B*).

Using the LASSO regression coefficients of the eight optimal feature genes and their expression levels in TCGA GSE37745 data set and the GEO data set, a risk-score model was constructed. Risk-score values were obtained for TCGA data set and the GEO validation set samples, respectively (*Figure 3C*, left). Using the median value of the risk score as the threshold, the samples in TCGA training set and the GEO validation set were divided into high- and low-risk groups, and the prognostic differences between these groups were evaluated (*Figure 3C*, middle). Based on the survival time and status of the samples in TCGA

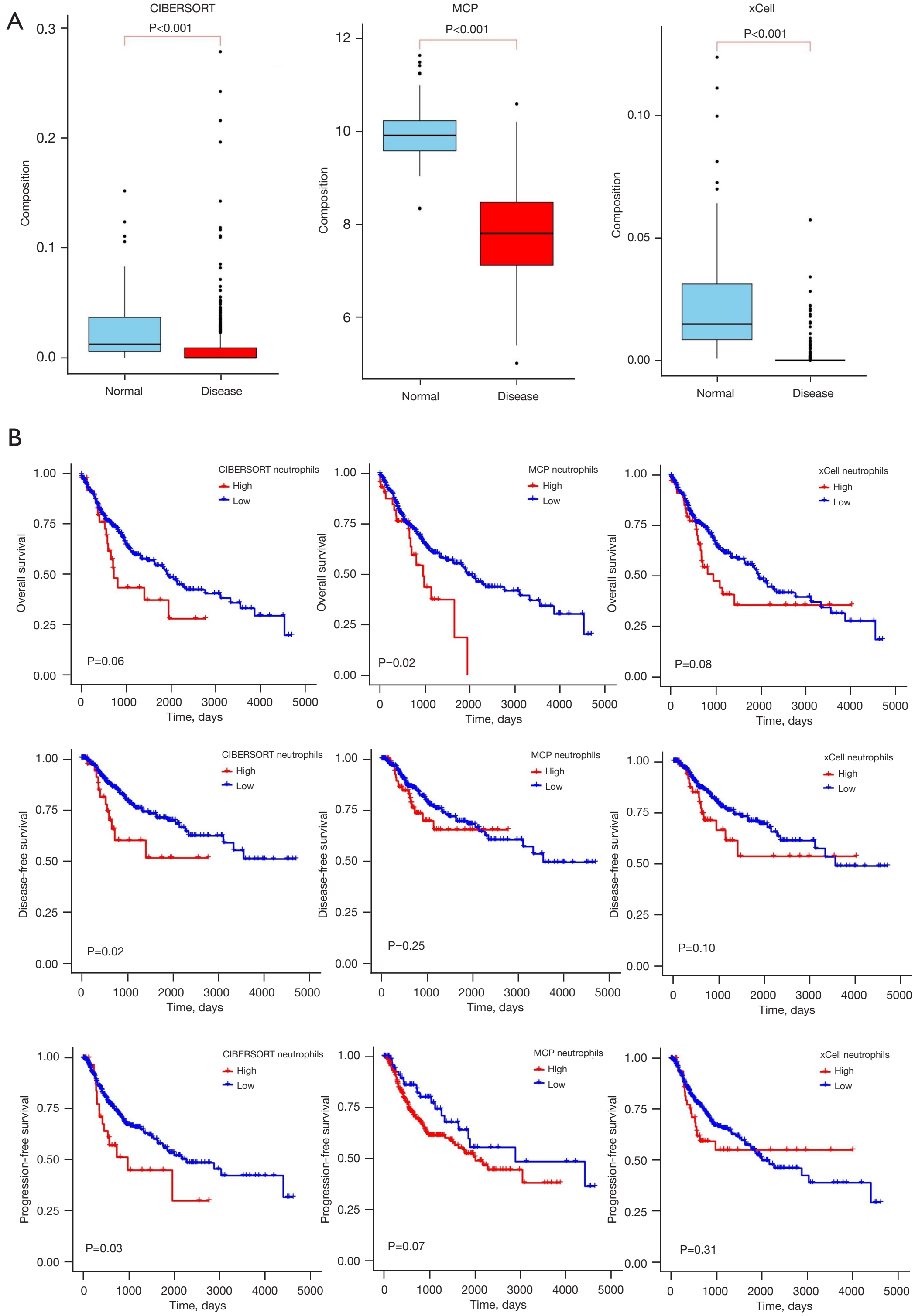
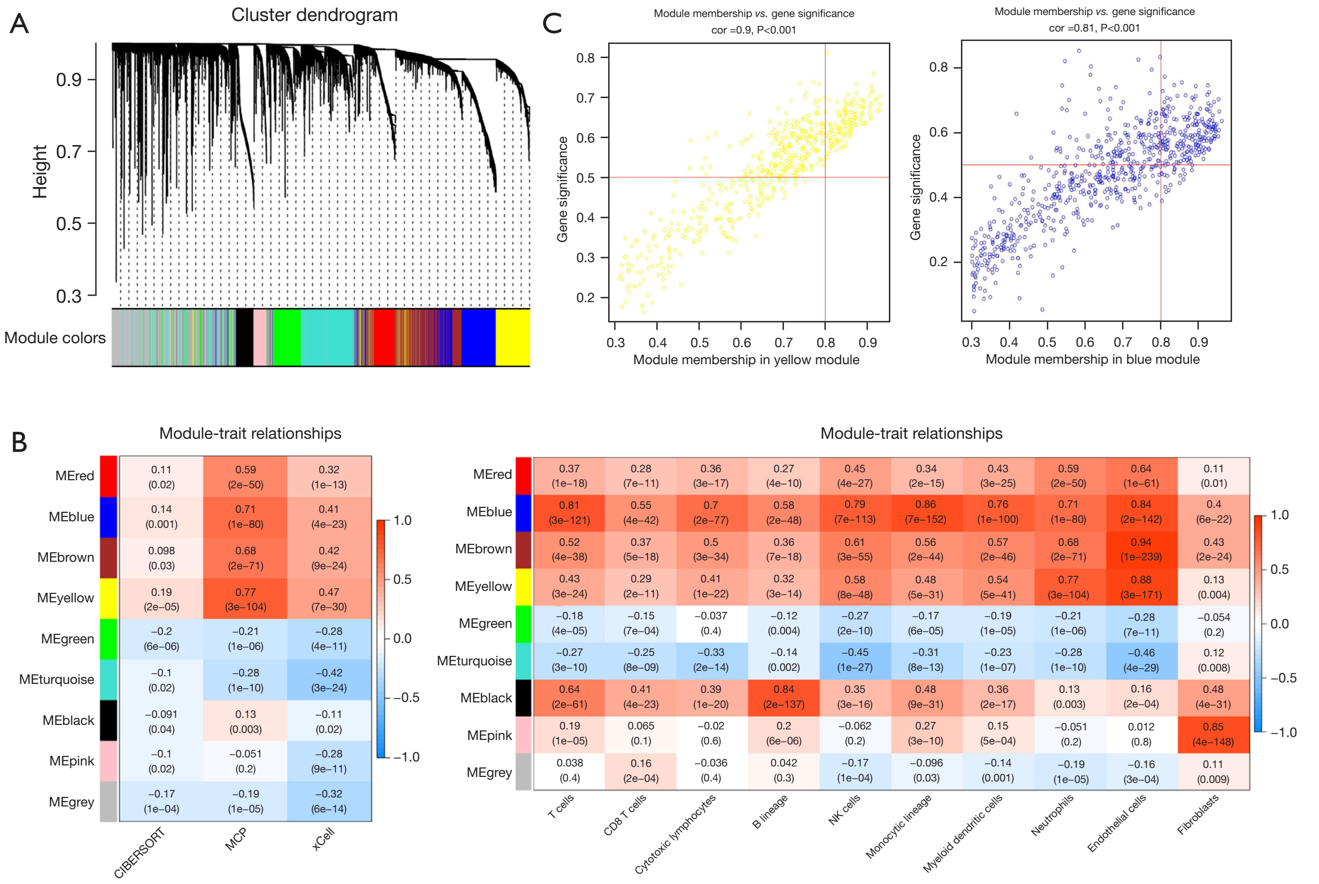


Figure 1 Correlation analysis between neutrophils and lung cancer. (A) Comparison of neutrophil content between normal and lung cancer tissue samples. (B) Analysis of association between neutrophil content and lung cancer prognosis using CIBERSORT, xCell, or MCP software. MCP, microenvironment cell populations.



Module-trait relationships

	CIBERSORT	MCP	xCell
MEred	0.11 (0.02)	0.59 (2e−50)	0.32 (1e−13)
MEblue	0.14 (0.001)	0.71 (1e−80)	0.41 (4e−23)
MEbrown	0.098 (0.03)	0.68 (2e−71)	0.42 (9e−24)
MEyellow	0.19 (2e−05)	0.77 (3e−104)	0.47 (7e−30)
MEgreen	−0.2 (6e−06)	−0.21 (1e−06)	−0.28 (4e−11)
MEturquoise	−0.1 (0.02)	−0.28 (1e−10)	−0.42 (3e−24)
MEblack	−0.091 (0.04)	0.13 (0.003)	−0.11 (0.02)
MEpink	−0.1 (0.02)	−0.051 (0.2)	−0.28 (9e−11)
MEgrey	−0.17 (1e−04)	−0.19 (1e−05)	−0.32 (6e−14)

Module-trait relationships

	T cells	CD8 T cells	Cytotoxic lymphocytes	B lineage	NK cells	Monocytic lineage	Myeloid dendritic cells	Neutrophils	Endothelial cells	Fibroblasts
MEred	0.37 (1e−18)	0.28 (7e−11)	0.36 (3e−17)	0.27 (4e−10)	0.45 (4e−27)	0.34 (2e−15)	0.43 (3e−25)	0.59 (2e−50)	0.64 (1e−61)	0.11 (0.01)
MEblue	0.81 (3e−121)	0.55 (4e−42)	0.7 (2e−77)	0.58 (2e−48)	0.79 (7e−113)	0.86 (7e−152)	0.76 (1e−100)	0.71 (1e−80)	0.84 (2e−142)	0.4 (6e−22)
MEbrown	0.52 (4e−38)	0.37 (5e−18)	0.5 (3e−34)	0.36 (7e−18)	0.61 (3e−55)	0.56 (2e−44)	0.57 (2e−46)	0.68 (2e−71)	0.94 (1e−239)	0.43 (2e−24)
MEyellow	0.43 (3e−24)	0.29 (2e−11)	0.41 (1e−22)	0.32 (3e−14)	0.58 (8e−48)	0.48 (5e−31)	0.54 (5e−41)	0.77 (3e−104)	0.88 (3e−171)	0.13 (0.004)
MEgreen	−0.18 (4e−05)	−0.15 (7e−04)	−0.037 (0.4)	−0.12 (0.004)	−0.27 (2e−10)	−0.17 (6e−05)	−0.19 (1e−05)	−0.21 (1e−06)	−0.28 (7e−11)	−0.054 (0.2)
MEturquoise	−0.27 (3e−10)	−0.25 (8e−09)	−0.33 (2e−14)	−0.14 (0.002)	−0.45 (1e−27)	−0.31 (8e−13)	−0.23 (1e−07)	−0.28 (1e−10)	−0.46 (4e−29)	0.12 (0.008)
MEblack	0.64 (2e−61)	0.41 (4e−23)	0.39 (1e−20)	0.84 (2e−137)	0.35 (3e−16)	0.48 (9e−31)	0.36 (2e−17)	0.13 (0.003)	0.16 (2e−04)	0.48 (4e−31)
MEpink	0.19 (1e−05)	0.065 (0.1)	−0.02 (0.6)	0.2 (6e−06)	−0.062 (0.2)	0.27 (3e−10)	0.15 (5e−04)	−0.051 (0.2)	0.012 (0.8)	0.85 (4e−148)
MEgrey	0.038 (0.4)	0.16 (2e−04)	−0.036 (0.4)	0.042 (0.3)	−0.17 (1e−04)	−0.096 (0.03)	−0.14 (0.001)	−0.19 (1e−05)	−0.16 (3e−04)	0.11 (0.009)

Figure 2 Construction of immune defense gene co-expression network. (A) Cluster dendrogram. (B) Correlation between neutrophil score and prognosis predicted by MCP software. The adjacent color bar of the heat map indicates the degree of correlation between modules and the infiltration level of neutrophils or other immune cells, with red shades representing positive correlation and blue shades representing negative correlation. (C) Correlation of blue and yellow modules with neutrophil subsets in immune cells. ME, module eigengenes; MCP, microenvironment cell populations.

training set and the GEO validation set, combined with the risk-score value of each sample, ROC curves were drawn to predict 1-, 3-, and 5-year survival (*Figure 3C*, right). The results revealed a significant correlation between the different risk groups obtained by the risk-score model and the actual prognosis.

Independent prognostic factor selection

A single-factor Cox analysis was performed on the risk group and clinical information (age, gender, risk group, pathologic M stage, pathologic N stage, pathologic T stage, and tumor stage), and the results showed that the risk group, pathologic M stage, and tumor stage were all significantly correlated with prognosis ($P<0.05$) (*Figure 4A*). The multivariable Cox regression analysis also confirmed that risk group, pathologic M stage, and tumor stage could serve as independent prognostic factors ($P<0.05$) (*Figure 4B*). A column chart (nomogram) and calibration curve were further drawn to demonstrate the accuracy of the model (*Figure 4C,4D*).

Molecular mechanism analysis

Based on the lung cancer gene mutation data from the NIH database, we counted the top 20 genes in terms of the mutation frequency and computed the TMB for each sample (Figure S4A). We compared the TMB between the different risk groups and found no significant difference between the two groups (Figure S4B). Using the median

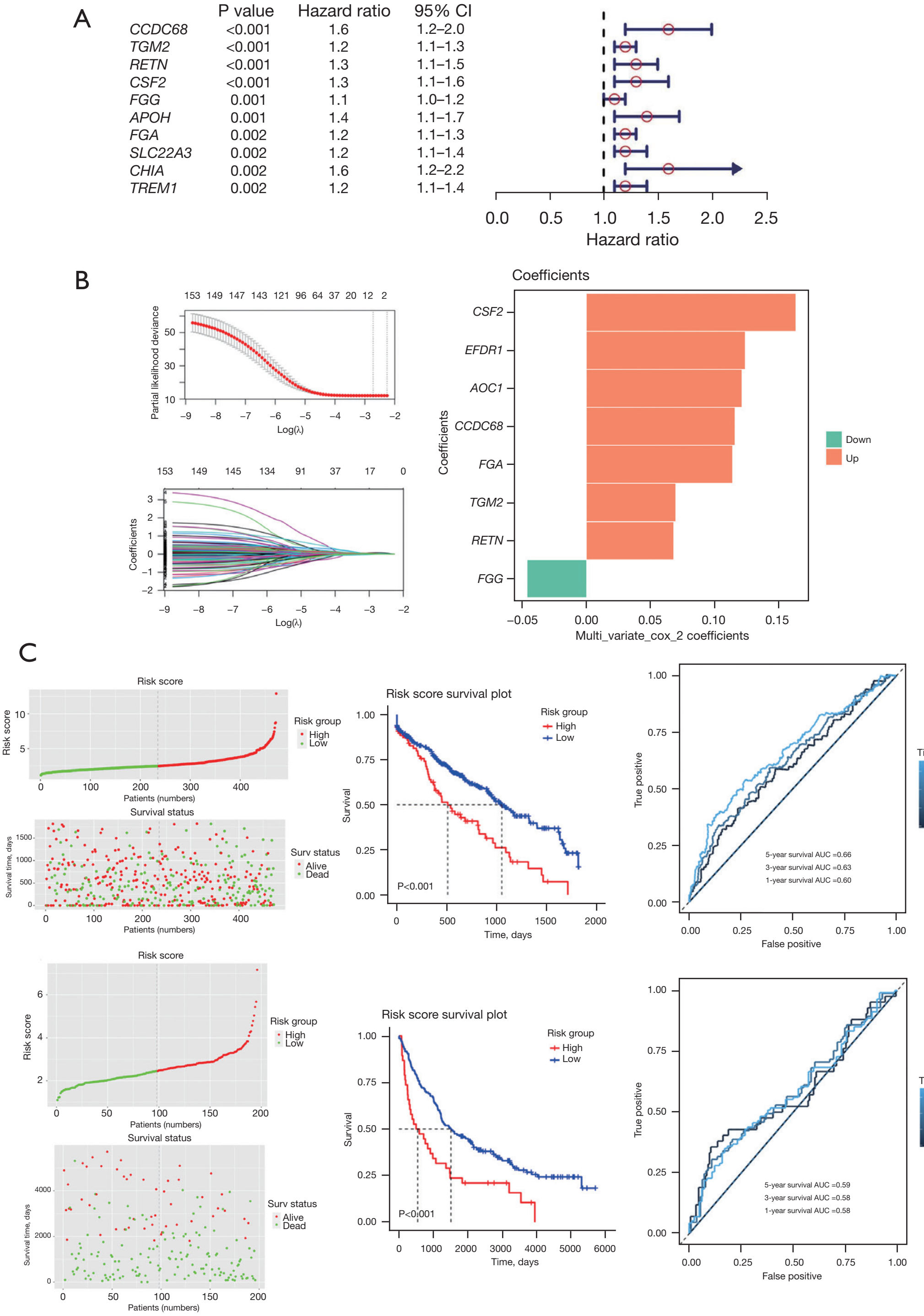


Figure 3 Construction of a risk-score model. (A) Univariate Cox regression analysis. (B) Screening of the optimal eight characteristic gene combinations and prognostic regression coefficients. (C) Construction and performance verification of the risk-score model (left: the distribution of risk score and survival status in TCGA data set and GEO validation set; middle: prognostic difference between the high- and low-risk groups; right: ROC curves for predicting 1-, 3-, and 5-year survival). CI, confidence interval; AUC, area under the curve; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; ROC, receiver operating characteristic.

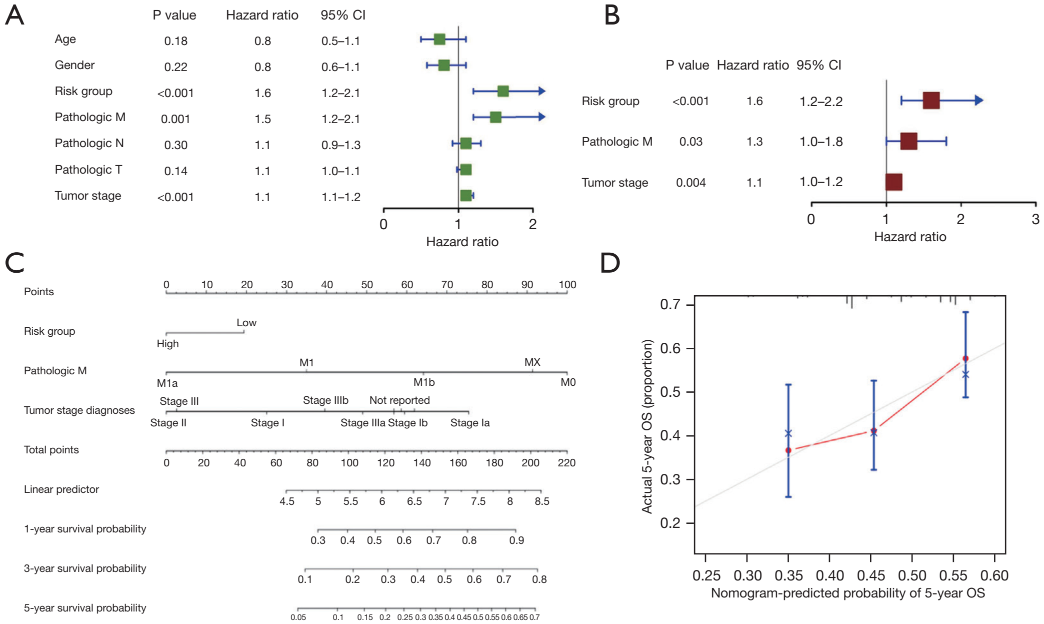


Figure 4 Construction of a prognostic model combined with clinical information and risk groups. (A) Univariate Cox analysis of risk group and clinical information. (B) Multi-factorial Cox analysis of risk group and clinical information. (C) Nomogram construction. (D) Validation of the efficacy of the model in predicting 5-year survival. CI, confidence interval; OS, overall survival.

value of the TMB as the cut-off value, all TCGA samples were divided into the TMB-high (a TMB ≥ median value) and TMB-low (a TMB < median value) groups. We further evaluated the survival prognosis difference between the TMB-high and TMB-low groups and found that patients in the TMB-high group had significantly better PFS than those in the TMB-low group (P=0.006) (Figure S4C,S4D).

From TCGA-LUSC sample expression data, we extracted the expression data of the immune checkpoint genes. The Deuchler-Wilcoxon test was used to compare the expression differences of the immune checkpoint genes and HLA family genes among the subtypes. As *Figure 5A* shows, the *B2M* and *HLA-E* genes were positively correlated with neutrophils, while the *PDCD1* and *CTLA4* genes were negatively correlated with neutrophils.

A GSVA enrichment analysis was performed on the hallmark pathway of the high- and low-risk groups, and a KEGG analysis was performed on each group separately. Based on the FDR <0.05 threshold, 33 KEGG pathways [normalized enrichment score (NES) >0] were significantly enriched in the high-risk group and 57 KEGG pathways (NES <0) were significantly enriched in the low-risk group. After sorting based on the absolute value of NES, only the top eight pathways with the largest absolute value of NES in the high- and low-risk groups were displayed (*Figure 5B*). The enrichment plot data has been normalized, so all values displayed in the figures are positive.

Drug sensitivity prediction

The sample expression levels in TCGA-LUSC were quantified for IC_{50} using the pRRophetic package in R. The Deuchler-Wilcoxon test was used to compare the differences in drug sensitivity between the high- and low-risk groups. The results showed that there were significant differences in drug sensitivity between the different risk groups for some drugs, including axitinib, sunitinib, cisplatin, vinorelbine, and vinblastine (Figure S5).

Prediction of immunotherapy responses

To determine the immune therapy responses in TCGA-LUSC samples, we conducted a TIDE analysis to analyze the escape process of lung cancer samples through the non-small cell lung cancer (NSCLC) immune escape pathway. The results showed that the TIDE in the high-risk group was significantly higher than that in the low-risk group (P<0.05), and the most significant differences were observed in the cancer-associated fibroblasts (CAF), CD8, CD274, and MSI immune escape values (Figure S6).

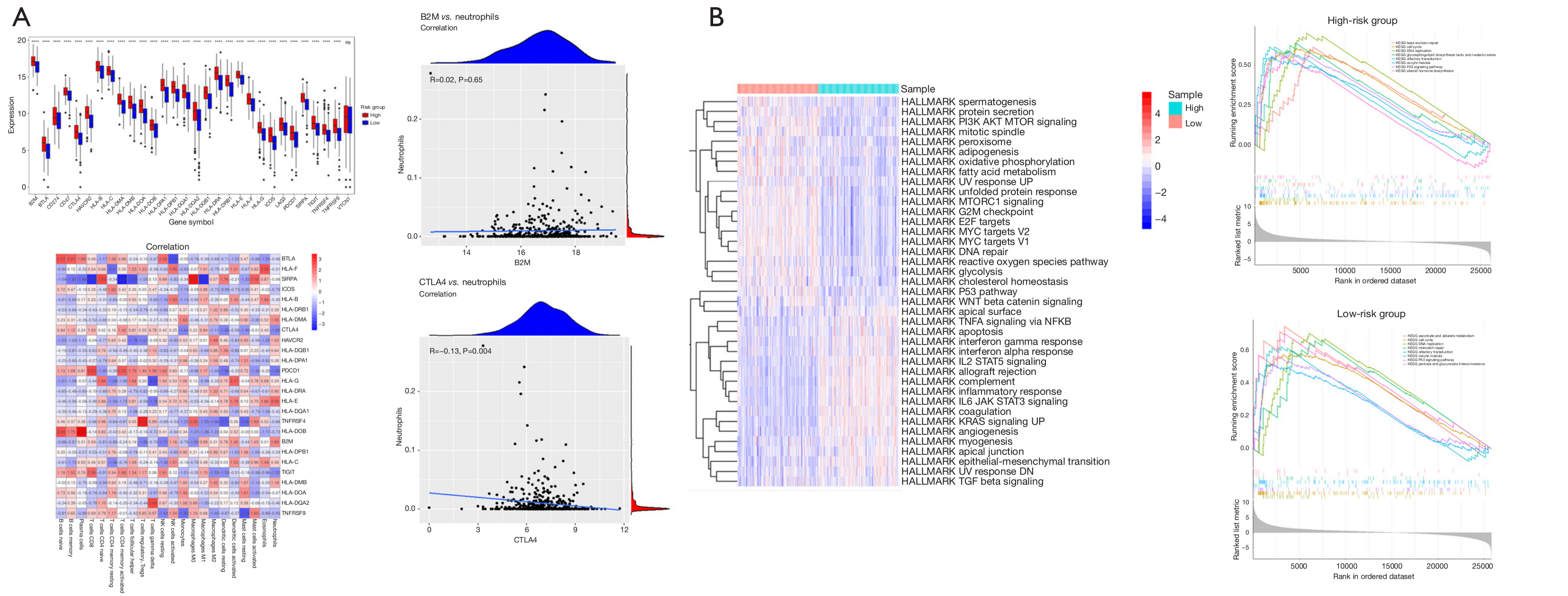


Figure 5 Molecular differences between the high- and low-risk groups. (A) Differences in the expression of immune checkpoint genes and HLA family genes. Upper left: differences in gene expression between two subtypes. Lower left: heat map of the correlation between immune genes and immune infiltrating cells. The adjacent color bar of the heat map indicates the degree of correlation between the expression of immune genes and immune cell infiltration level, with red shades representing positive correlation and blue shades representing negative correlation. Upper right: correlation between B2M expression and neutrophils in TCGA samples. Lower right: correlation between CTLA4 expression and neutrophils in TCGA samples. (B) Differences in the hallmarks pathway and KEGG pathway. Left: hallmarks enrichment score analysis among two groups. The adjacent color bar of the heat map indicates the degree of enrichment, with red shades representing higher enrichment and blue shades representing lower enrichment. Upper right: enrichment of KEGG pathway among the high-risk group in TCGA samples. Lower right: enrichment of KEGG pathway among the low-risk group in TCGA samples. ****, P<0.05; ns, no significance. HLA, human leukocyte antigen; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The Cancer Genome Atlas.

Discussion

In recent years, emerging evidence has indicated that the host immune response plays a critical role in lung cancer development and progression, and modulating the immune microenvironment may represent a promising strategy for improving therapeutic outcomes (30,31). One of the key cellular components in the TME are neutrophils, a type of white blood cell primarily responsible for immune defense against bacterial and fungal infections (9). In addition to their antimicrobial functions, neutrophils also possess the ability to interact with tumor cells and modulate the cancer immune response, leading to varying effects on cancer progression (8,9). Recent studies have suggested that the presence and abundance of neutrophils in the TME could be an important predictor of prognosis in lung cancer patients (10,11). In order to investigate the role of neutrophils in the prognosis of LUSC patients, a risk score model that can effectively identify the prognosis of LUSC patients was conducted, based on the expression level and correlation coefficient of neutrophil-related genes, and the related molecular mechanism of this model was explored. Our findings provide valuable insights into the prognostic significance of neutrophils in lung cancer and offer potential targets for personalized treatments.

Previous studies have explored the relationship between the ratio of neutrophils to lymphocytes or the genetic characteristics of circulating platelet-bound neutrophils and the prognosis of lung cancer (10,11,32-34). In this study, the analysis revealed that the neutrophil content was significantly higher in the normal samples than the tumor samples, which indicated that the reduction of neutrophils might affect the occurrence and development of LUSC, also reflected the key role of neutrophils. However, subsequent analyses using three different neutrophils level analysis software revealed that higher levels of neutrophils were associated with poorer prognosis in patients with LUSC, which appears to be a contradictory result compared to the previous one. Specifically, a high neutrophil content, as estimated by MCP software, was significantly associated with worse OS, while a high neutrophil content, as estimated by CIBERSORT software, was significantly associated with worse DFS and PFS. Obviously, inconsistent results were obtained when using different analysis software, suggesting the correlation between the level of neutrophils and prognosis is not absolute. The neutrophil level alone cannot be used as a prognostic factor of LUSC. In addition to the neutrophil level, it is necessary to explore multiple dimensions as prognostic factors of LUSC.

Therefore, the establishment of a prediction model comprising eight genes (i.e., *CSF2*, *EPDR1*, *AOC1*, *CCDC68*, *FGA*, *TGM2*, *RETN*, and *FGG*) through a WGCNA represents a notable contribution of this study. These genes have been implicated in various biological function, such as inflammatory responses, extracellular matrix remodeling, and immune regulation (35-42). The LASSO Cox regression algorithm effectively screened these genes and a risk score model was built that demonstrated good predictive performance for patient prognosis. Moreover, the validation of the model using external data sets increased its reliability and generalizability.

This study also investigated the molecular mechanisms of this risk-score model in lung cancer. By analyzing the gene mutation data, the top frequently mutated genes were identified, showcasing the genetic landscape of lung cancer and providing potential targets for future research. The TMB analysis revealed that patients with a higher TMB had better PFS. This finding is consistent with the findings of previous studies that a higher TMB is associated with an increased likelihood of a response to immune checkpoint inhibitors (43,44). The differential expression of immune checkpoint-related genes between different risk groups is an interesting observation. The negative correlation between neutrophils and genes such as *PDCD1* (*PD-1*) and *CTLA4*, which are important immune checkpoint regulators (45), suggests that neutrophils have a potential immunosuppressive role in the TME. Conversely, the positive correlation between neutrophils and the *B2M* and *HLA-E* genes, which are crucial for antigen presentation (46,47), suggests a possible interaction between neutrophils and the adaptive immune response. These findings contribute to understandings of the complex interactions between neutrophils and the immune system in the TME.

The precise mechanism underlying the association between neutrophils and lung cancer remains to be fully elucidated; however, several hypotheses have been proposed based on current knowledge of neutrophil biology and tumor immunology. For instance, it has been suggested that neutrophils may interact with tumor cells through the release of pro-inflammatory cytokines and chemokines, which promotes tumor cell proliferation, angiogenesis, and immune evasion (13,48). Alternatively, neutrophils may act as a double-edged sword in the TME, exhibiting tumor-suppressive effects under certain circumstances (13). Comparing this study with previous literature, several similarities and differences should be noted. Similar to previous studies (10,11,49), this research highlighted the association between neutrophil content and clinical outcomes in lung cancer. The establishment of a prognostic signature involving multiple genes is consistent with the approach used in some previous studies (50-53) to identify the molecular markers associated with patient prognosis. However, this study added value by using a combination of software tools to estimate the neutrophil content and by performing comprehensive molecular mechanism analyses, including gene mutation and immune checkpoint gene expression analyses. These approaches provide a more comprehensive understanding of the role of neutrophils in lung cancer.

This study provided valuable insights; however, there are several limitations to consider. First, the data used for the analysis were obtained primarily from public databases, which may have introduced biases and limitations in terms of the sample size, diversity, and data quality. Second, this study focused on a specific subtype of lung cancer; thus, the findings may not be directly applicable to other subtypes or cancers. Future studies should aim to include a more diverse set of lung cancer patients to ensure broader applicability. Additionally, the identified prognostic signature and molecular mechanisms require further experimental validation to confirm their clinical utility and functional relevance.

Conclusions

In conclusion, this study contributed to establish a risk score model that can effectively identify the prognosis of LUSC patients and explore the related molecular mechanism of this model. The findings expand understandings of the complex interactions between neutrophils and tumor biology, providing potential targets for personalized treatments. However, further experimental validation and clinical studies are necessary to confirm and implement these findings in clinical practice. The limitations of the study, such as the reliance on public databases and the focus on a specific lung cancer subtype, should be considered when interpreting the results. Future research should explore these aspects to establish a broader and more robust understanding of the prognostic role of neutrophils in LUSC.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-411/rc

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-411/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-411/coif). N.S. obtained research grants from Eli Lilly, Chugai Pharmaceutical, Taiho Pharmaceutical, Pfizer, Ono Pharmaceutical, Nippon Kayaku, Takeda Pharmaceutical, and Boehringer Ingelheim and received speaking honoraria from Eli Lilly, AstraZeneca, MSD, Chugai Pharmaceutical, Taiho Pharmaceutical, Pfizer, Ono Pharmaceutical, Nippon Kayaku, Takeda Pharmaceutical, Daiichi Sankyo, Boehringer Ingelheim, and Bristol Myers Squibb. The other authors have no conflicts of interest to declare.

Ethical 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).

References

1. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. CA Cancer J Clin 2023;73:17-48.
2. Li C, Lei S, Ding L, et al. Global burden and trends of lung cancer incidence and mortality. Chin Med J (Engl) 2023;136:1583-90.
3. Xu Y, Li H, Huang Z, et al. Predictive values of genomic

variation, tumor mutational burden, and PD-L1 expression in advanced lung squamous cell carcinoma treated with immunotherapy. Transl Lung Cancer Res 2020;9:2367-79.

4. Zhao X, Yuan C, He X, et al. Identification and in vitro validation of diagnostic and prognostic biomarkers for lung squamous cell carcinoma. J Thorac Dis 2022;14:1243-55.
5. Cao L, Zhong J, Liu Z, et al. Increased LOXL2 expression is related to poor prognosis in lung squamous cell carcinoma. J Thorac Dis 2024;16:581-92.
6. Du C, Cai J, Tang J, et al. Cell-free DNA methylation profile potential in the diagnosis of lung squamous cell carcinoma. J Thorac Dis 2024;16:553-63.
7. de Visser KE, Joyce JA. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer Cell 2023;41:374-403.
8. Gibellini L, Borella R, Santacroce E, et al. Circulating and Tumor-Associated Neutrophils in the Era of Immune Checkpoint Inhibitors: Dynamics, Phenotypes, Metabolism, and Functions. Cancers (Basel) 2023;15:3327.
9. Zhong J, Zong S, Wang J, et al. Role of neutrophils on cancer cells and other immune cells in the tumor microenvironment. Biochim Biophys Acta Mol Cell Res 2023;1870:119493.
10. Yu X, Li C, Wang Z, et al. Neutrophils in cancer: dual roles through intercellular interactions. Oncogene 2024;43:1163-77.
11. Cao W, Yu H, Zhu S, et al. Clinical significance of preoperative neutrophil-lymphocyte ratio and platelet-lymphocyte ratio in the prognosis of resected early-stage patients with non-small cell lung cancer: A meta-analysis. Cancer Med 2023;12:7065-76.
12. Huai Q, Luo C, Song P, et al. Peripheral blood inflammatory biomarkers dynamics reflect treatment response and predict prognosis in non-small cell lung cancer patients with neoadjuvant immunotherapy. Cancer Sci 2023;114:4484-98.
13. Mauracher LM, Hell L, Moik F, et al. Neutrophils in lung cancer patients: Activation potential and neutrophil extracellular trap formation. Res Pract Thromb Haemost 2023;7:100126.
14. Sanchez-Cespedes M. B2M, JAK2 and MET in the genetic landscape of immunotolerance in lung cancer. Oncotarget 2018;9:35603-4.
15. Di Cristofaro J, Pelardy M, Loundou A, et al. HLA-E(*)01:03 Allele in Lung Transplant Recipients Correlates with Higher Chronic Lung Allograft Dysfunction Occurrence. J Immunol Res 2016;2016:1910852.
16. Ribatti D. A double-edged sword in tumor angiogenesis and progression. Dual roles of mast cells, macrophages, and neutrophils. Pathol Res Pract 2022;240:154167.
17. Kaplan MJ, Radic M. Neutrophil extracellular traps: double-edged swords of innate immunity. J Immunol 2012;189:2689-95.
18. Goldman M, Craft B, Hastie M, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. BioRxiv 2018. [Preprint]. doi: 10.1101/326470.
19. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013;41:D991-5.
20. Chen B, Khodadoust MS, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol 2018;1711:243-59.
21. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017;18:220.
22. Meylan M, Becht E, Sautès-Fridman C, et al. webMCP-counter: a web interface for transcriptomics-based quantification of immune and stromal cells in heterogeneous human or murine samples. BioRxiv 2020. [Preprint]. doi: 10.1101/2020.12.03.400754.
23. Wang P, Wang Y, Hang B, et al. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget 2016;7:55343-51.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
25. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
26. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.
27. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90-7.
28. Friedman J, Hastie T, Tibshirani R, et al. glmnet: Lasso and elastic-net regularized generalized linear models. R Package Version 1. 2009. Available online: https://cran.r-project.org/web/packages/glmnet/index.html
29. Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747-56.
30. Smok-Kalwat J, Mertowska P, Mertowski S, et al. The Importance of the Immune System and Molecular Cell

Signaling Pathways in the Pathogenesis and Progression of Lung Cancer. Int J Mol Sci 2023;24:1506.

31. Wang Q, Shao X, Zhang Y, et al. Role of tumor microenvironment in cancer progression and therapeutic strategy. Cancer Med 2023;12:11149-65.
32. Lecot P, Ardin M, Dussurgey S, et al. Gene signature of circulating platelet-bound neutrophils is associated with poor prognosis in cancer patients. Int J Cancer 2022;151:138-52.
33. Shaul ME, Eyal O, Guglietta S, et al. Circulating neutrophil subsets in advanced lung cancer patients exhibit unique immune signature and relate to prognosis. FASEB J 2020;34:4204-18.
34. Russo A, Russano M, Franchina T, et al. Neutrophil-to-Lymphocyte Ratio (NLR), Platelet-to-Lymphocyte Ratio (PLR), and Outcomes with Nivolumab in Pretreated Non-Small Cell Lung Cancer (NSCLC): A Large Retrospective Multicenter Study. Adv Ther 2020;37:1145-55.
35. Li H, Zhong R, He C, et al. Colony-stimulating factor CSF2 mediates the phenotypic plasticity of small-cell lung cancer by regulating the p-STAT3/MYC pathway. Oncol Rep 2022;48:122.
36. Cataldo LR, Gao Q, Argemi-Muntadas L, et al. The human batokine EPDR1 regulates β-cell metabolism and function. Mol Metab 2022;66:101629.
37. Liu F, Ou W, Tang W, et al. Increased AOC1 Expression Promotes Cancer Progression in Colorectal Cancer. Front Oncol 2021;11:657210.
38. Li X, Li H, Pei X, et al. CCDC68 Upregulation by IL-6 Promotes Endometrial Carcinoma Progression. J Interferon Cytokine Res 2021;41:12-9.
39. Liu G, Xu X, Geng H, et al. FGA inhibits metastases and induces autophagic cell death in gastric cancer via inhibiting ITGA5 to regulate the FAK/ERK pathway. Tissue Cell 2022;76:101767.
40. Malkomes P, Lunger I, Oppermann E, et al. Transglutaminase 2 is associated with adverse colorectal cancer survival and represents a therapeutic target. Cancer Gene Ther 2023;30:1346-54.
41. Wang CQ, Tang CH, Tzeng HE, et al. Impacts of RETN genetic polymorphism on breast cancer development. J Cancer 2020;11:2769-77.
42. Peng HH, Wang JN, Xiao LF, et al. Elevated Serum FGG Levels Prognosticate and Promote the Disease Progression in Prostate Cancer. Front Genet 2021;12:651647.
43. Vryza P, Fischer T, Mistakidi E, et al. Tumor mutation burden in the prognosis and response of lung cancer patients to immune-checkpoint inhibition therapies. Transl Oncol 2023;38:101788.
44. Choucair K, Morand S, Stanbery L, et al. TMB: a promising immune-response biomarker, and potential spearhead in advancing targeted therapy trials. Cancer Gene Ther 2020;27:841-53.
45. Deng L, Gyorffy B, Na F, et al. Association of PDCD1 and CTLA-4 Gene Expression with Clinicopathological Factors and Survival in Non-Small-Cell Lung Cancer: Results from a Large and Pooled Microarray Database. J Thorac Oncol 2015;10:1020-6.
46. Yang K, Halima A, Chan TA. Antigen presentation in cancer - mechanisms and clinical implications for immunotherapy. Nat Rev Clin Oncol 2023;20:604-23.
47. Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. Nat Rev Cancer 2021;21:298-312.
48. Habanjar O, Bingula R, Decombat C, et al. Crosstalk of Inflammatory Cytokines within the Breast Tumor Microenvironment. Int J Mol Sci 2023;24:4002.
49. Imai H, Wasamoto S, Tsuda T, et al. Using the neutrophil-to-lymphocyte ratio to predict the outcome of individuals with nonsquamous non-small cell lung cancer receiving pembrolizumab plus platinum and pemetrexed. Thorac Cancer 2023;14:2567-78.
50. Roberts E, Howell S, Evans DG. Polygenic risk scores and breast cancer risk prediction. Breast 2023;67:71-7.
51. Suh YS, Lee J, George J, et al. RNA expression of 6 genes from metastatic mucosal gastric cancer serves as the global prognostic marker for gastric cancer with functional validation. Br J Cancer 2024;130:1571-84.
52. He J, Tian Z, Yao X, et al. A novel RNA sequencing-based risk score model to predict papillary thyroid carcinoma recurrence. Clin Exp Metastasis 2020;37:257-67.
53. Wu Y, Deng J, Lai S, et al. A risk score model with five long non-coding RNAs for predicting prognosis in gastric cancer: an integrated analysis combining TCGA and GEO datasets. PeerJ 2021;9:e10556.

(English Language Editor: L. Huleatt)

Cite this article as: Wang Y, Li D, Li Q, Basnet A, Efird JT, Seki N. Neutrophil estimation and prognosis analysis based on existing lung squamous cell carcinoma datasets: the development and validation of a prognosis prediction model. Transl Lung Cancer Res 2024;13(8):2023-2037. doi: 10.21037/tlcr-24-411

Supplementary

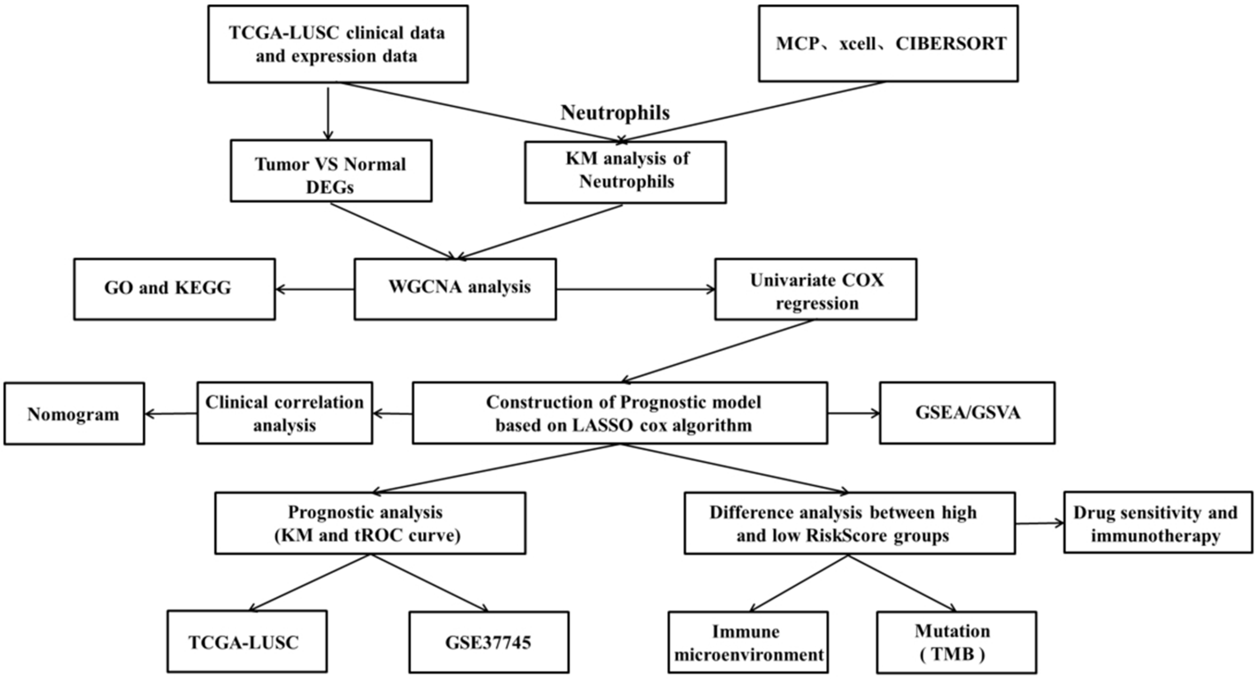

TCGA-LUSC clinical data and expression data
MCP、xcell、CIBERSORT
Neutrophils
Tumor VS Normal DEGs
KM analysis of Neutrophils
GO and KEGG
WGCNA analysis
Univariate COX regression
Nomogram
Clinical correlation analysis
Construction of Prognostic model based on LASSO cox algorithm
GSEA/GSVA
Prognostic analysis (KM and tROC curve)
Difference analysis between high and low RiskScore groups
Drug sensitivity and immunotherapy
TCGA-LUSC
GSE37745
Immune microenvironment
Mutation (TMB)

Figure S1 The workflow of this study. TCGA, The Cancer Genome Atlas; LUSC, lung squamous cell carcinoma; MCP, microenvironment cell populations; DEGs, differentially expressed genes; KM, Kaplan-Meier; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; WGCNA, weighted gene co-expression network analysis; LASSO, least absolute shrinkage and selection operator; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; tROC, time-dependent receiver operating characteristic; TMB, tumor mutation burden.

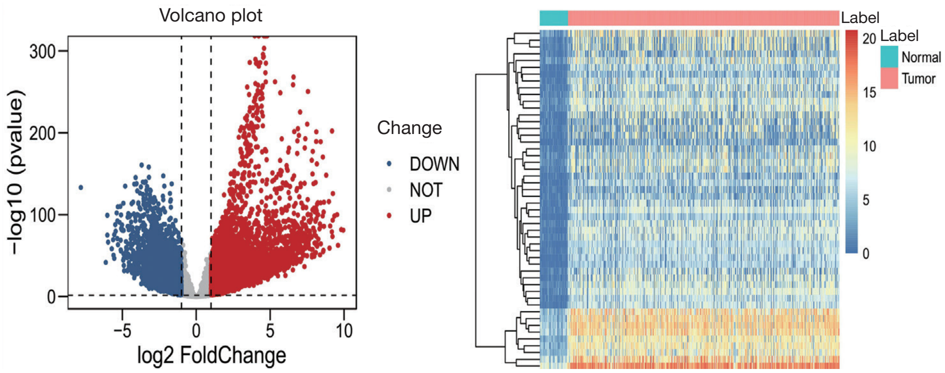


Figure S2 Comparison of neutrophil-related gene expression between lung cancer and normal tissue samples. Right: the adjacent color bar of the heat map indicates the expression level, with red indicating higher expression and blue indicating lower expression.

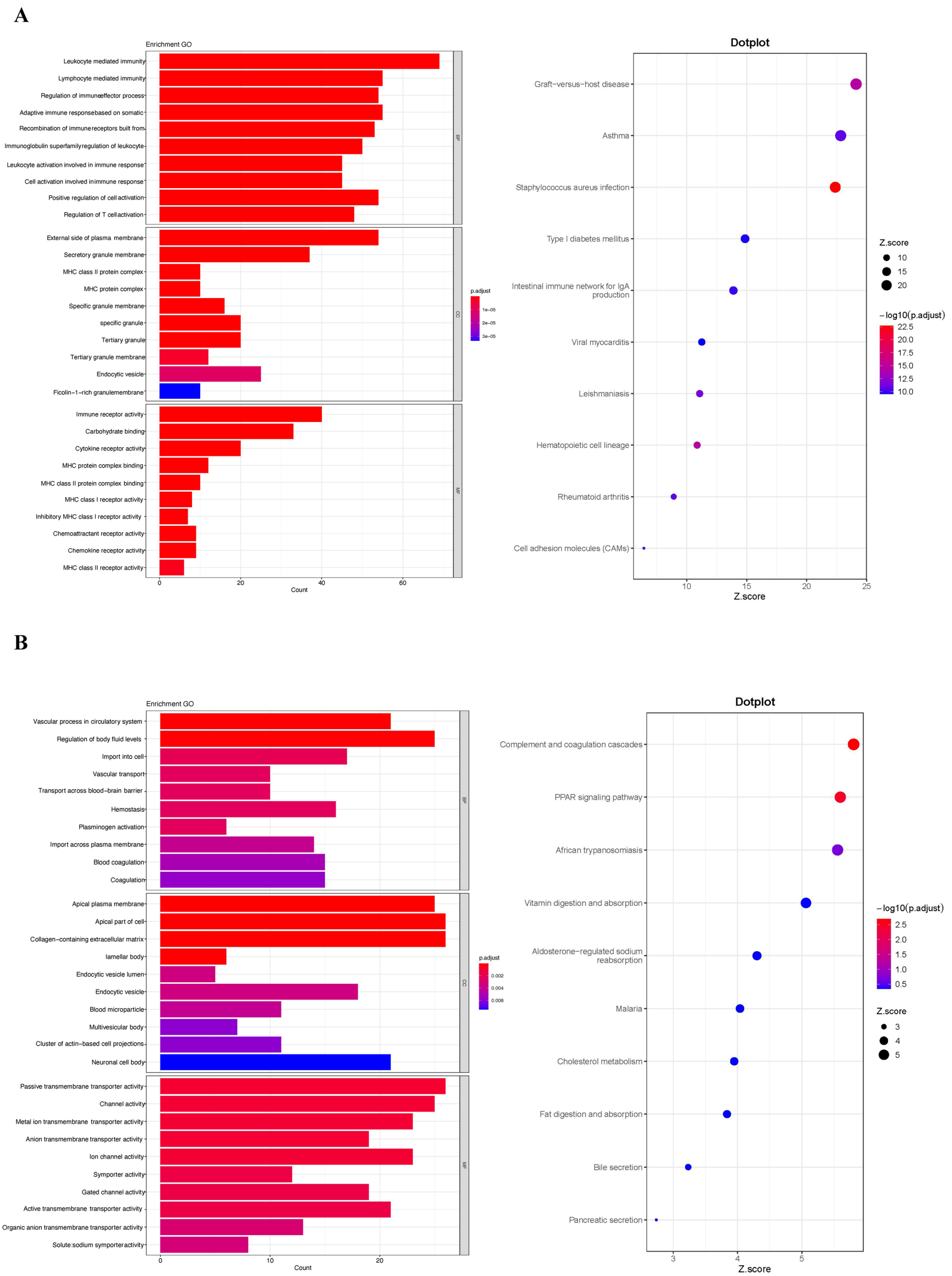


Figure S3 Annotation of the related path (GO BP) of the blue module (A) and yellow modules (B). GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function.

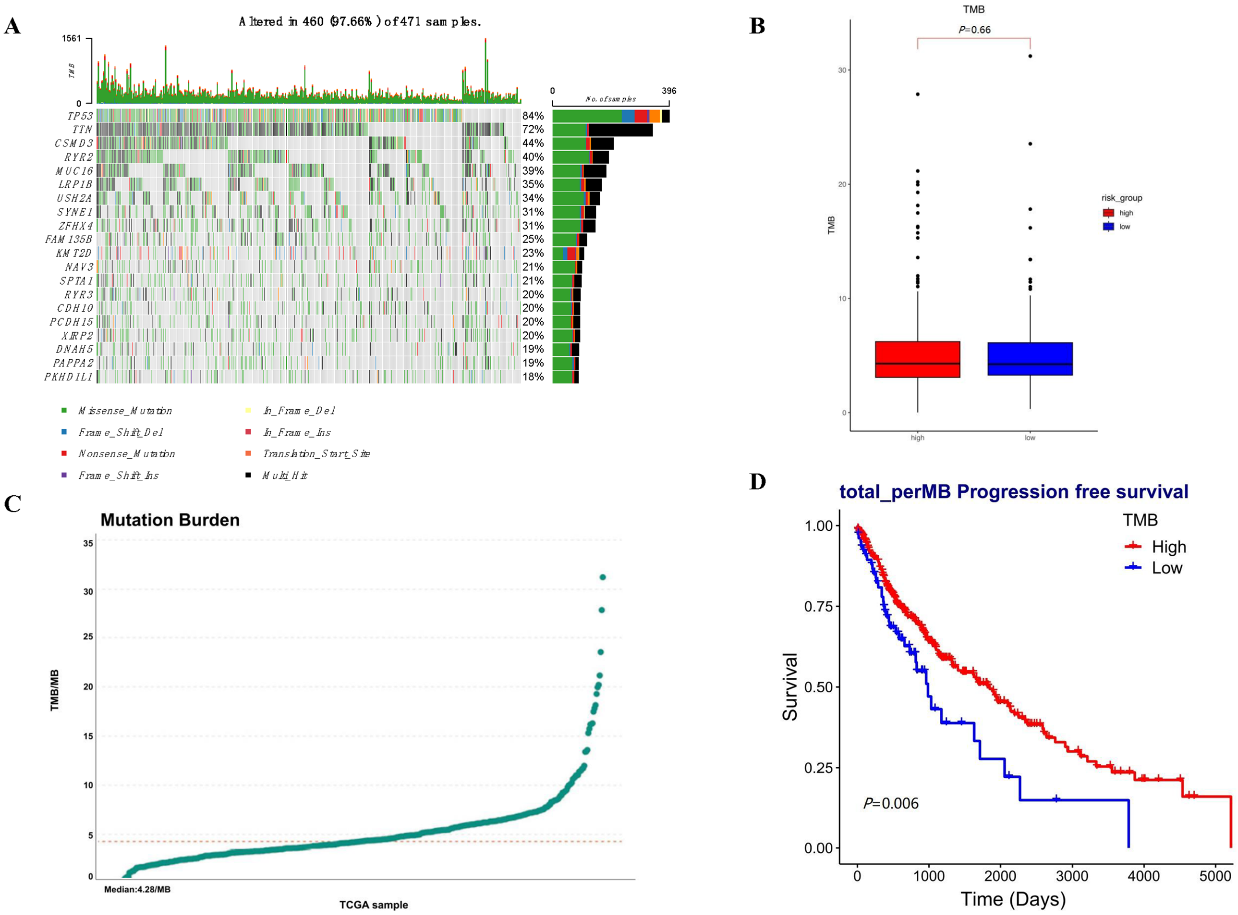


Figure S4 Analysis of the genetic mutations. (A) The 20 most frequently mutated genes. (B) Correlation between the risk group and TMB. (C) TMB sorting of all the samples. (D) Prognostic difference between the TMB-high and TMB-low groups. TMB, tumor mutation burden; TCGA, The Cancer Genome Atlas.

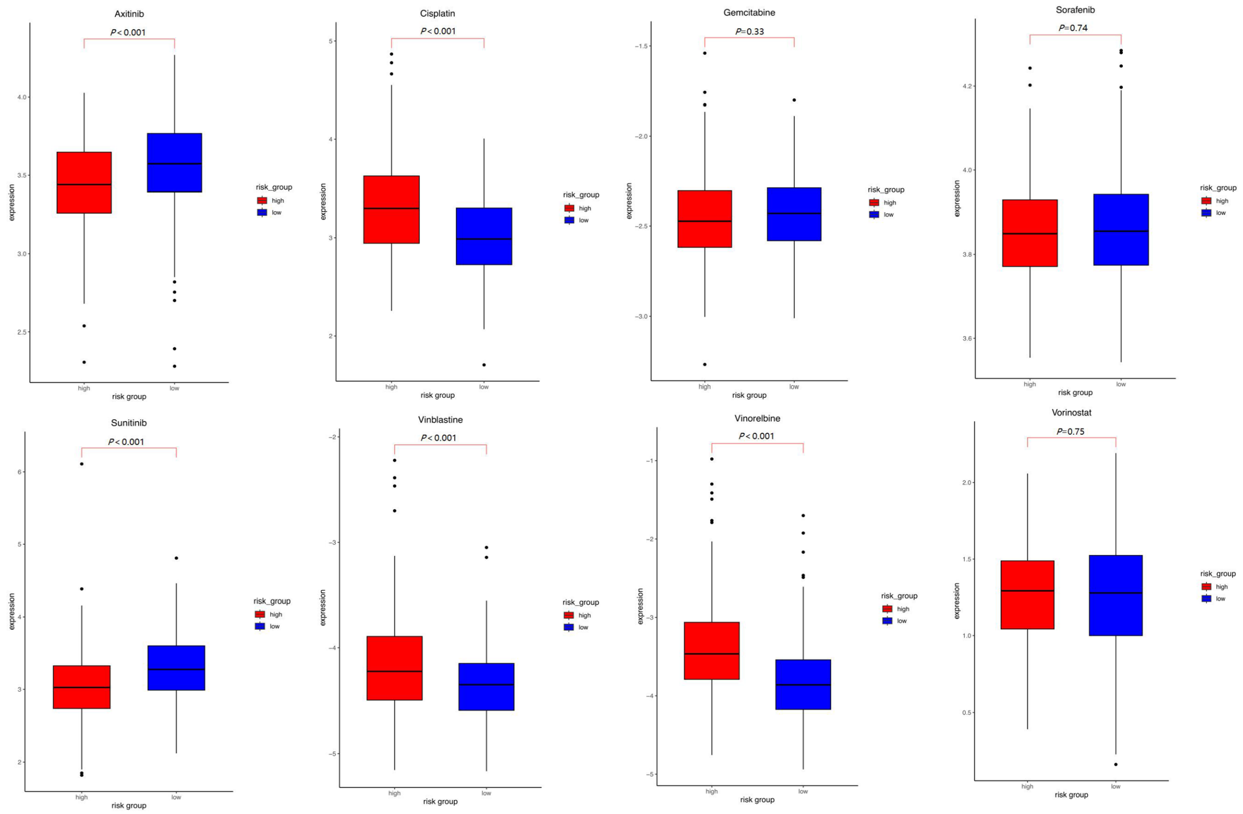


Figure S5 Difference in drug sensitivity between the high- and low-risk groups.

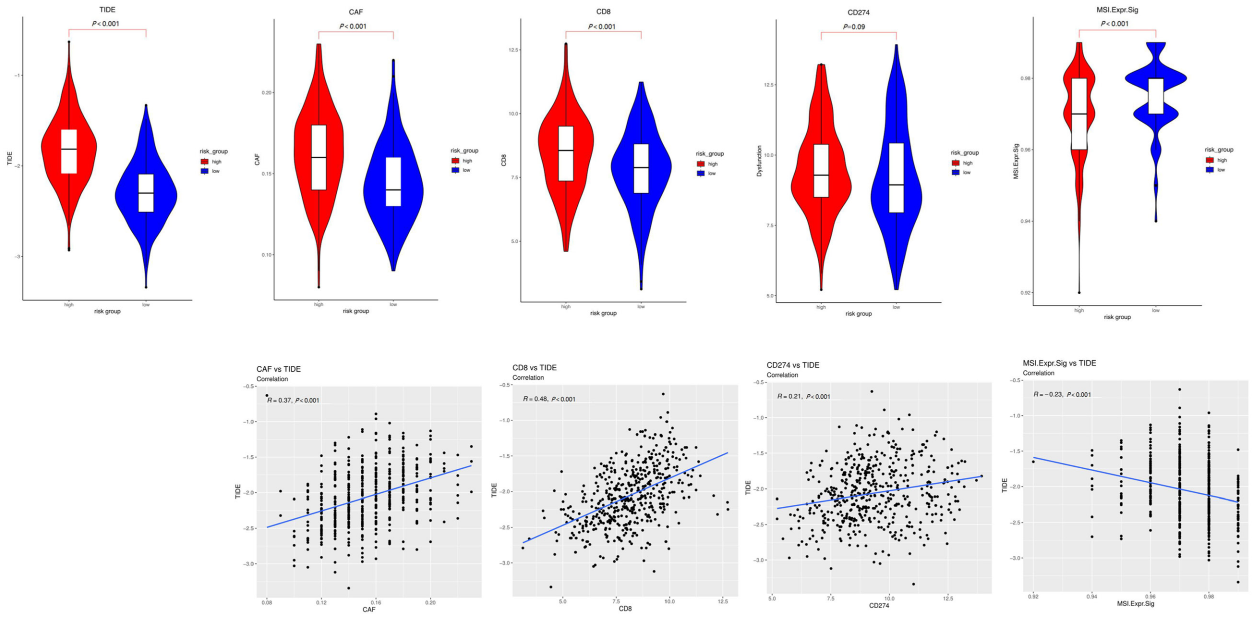


Figure S6 Analysis of TIDE. TIDE, Tumor Immune Dysfunction and Exclusion; CAF, cancer-associated fibroblasts; MSI, microsatellite instability; expr., expression; sig, significant.