Frontiers in Bioscience Landmark

Front. Biosci. (Landmark Ed) **2025**; 30(10): 44661
https://doi.org/10.31083/FBL44661

Original Research

Identification of Diagnostic Biomarkers Associated With M1 Macrophage in Lung Squamous Cell Carcinoma via Machine Learning

Huiting Deng[1,†], Zhenling Wang[2,†], Qiangzhe Zhang[3,*]

[1]Tianjin Key Laboratory of Extracorporeal Life Support for Critical Diseases, Tianjin Institute of Hepatobiliary Disease, Tianjin Third Central Hospital affiliated to Nankai University, 300170 Tianjin, China
[2]Department of Medicine, Tianjin Huanhu Hospital, School of Medicine, Nankai University, 300350 Tianjin, China
[3]State Key Laboratory of Medicinal Chemical Biology and College of Pharmacy, Tianjin Key Laboratory of Molecular Drug Research, Nankai University, 300350 Tianjin, China
*Correspondence: zhangqiangzhe@nankai.edu.cn (Qiangzhe Zhang)
[†]These authors contributed equally.
Academic Editor: Sukmook Lee
Submitted: 9 July 2025 Revised: 24 September 2025 Accepted: 11 October 2025 Published: 31 October 2025

Abstract

Background: Macrophage infiltration is prevalent in lung cancer tissues, significantly influencing disease progression and clinical outcomes. Lung squamous cell carcinoma (LUSC) is often diagnosed at advanced stages, resulting in poor prognosis. Identifying effective diagnostic biomarkers, particularly those associated with macrophage infiltration, is crucial for early detection and improved treatment outcomes. This study aims to identify diagnostic markers specifically linked to M1 macrophages in LUSC. **Methods**: Differential gene expression analysis and immune cell infiltration assessment were conducted using the limma and CIBERSORT packages. The WGCNA algorithm was then applied to identify genes in modules related to M1 macrophages. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were used to investigate the biological functions of M1 macrophage-related differentially expressed genes (DEGs). To identify M1 macrophage-associated biomarkers in LUSC, a diagnostic model was developed using four machine learning algorithms, with validation through nomogram visualization, calibration curves, and external datasets. Finally, immunohistochemical staining was performed to further confirm the expression of hub genes and the predictive accuracy of M1 macrophage-related biomarkers in LUSC. **Results**: A total of 143 M1 macrophage-related DEGs were identified, which were involved in regulating immune response pathways. The support vector machine (SVM) model based on these genes demonstrated exceptional performance, with area under the curve (AUC) values of 0.995 in the training cohort and 1.000 in three external validation datasets. Immunohistochemical analysis further confirmed the diagnostic accuracy of Matrix metalloproteinase-7 (MMP7), Reticulon-1 (RTN1), Zinc finger protein ZIC 2 (ZIC2), Killer cell lectin-like receptor subfamily B member 1 (KLRB1), and C-X-C motif chemokine 13 (CXCL13), yielding an AUC of 0.992. These results highlight the strong diagnostic capability of the 5 hub genes in LUSC. **Conclusion**: The study highlights the pivotal role of M1 macrophage-related DEGs in LUSC tumorigenesis. The newly identified 5 hub genes provide a highly accurate diagnostic tool for LUSC, offering potential improvements for both diagnostic and therapeutic strategies.

Keywords: carcinoma; non-small-cell lung; macrophage; biomarkers; machine learning; tumor microenvironment

1. Introduction

Lung cancer remains the leading cause of cancer incidence and mortality globally, affecting 185 countries [1]. In 2022, approximately 1,060,600 new lung cancer cases and 733,300 lung cancer-related deaths were reported in China [2]. Lung squamous cell carcinoma (LUSC) is the second most common subtype of non-small cell lung cancer (NSCLC), accounting for approximately 30% of all NSCLC cases [3]. Typically located in the central region of the lung and originating in the proximal bronchi, LUSC has a heightened risk of invading larger blood vessels [4]. The five-year survival rate for LUSC ranges from 73% at stage IA to 13% at stage IV [5]. Unfortunately, most patients are diagnosed at advanced stages [6,7]. Consequently, identifying novel biomarkers is essential to improving early diagnosis and treatment strategies for LUSC. Furthermore, understanding the pathogenesis of LUSC could contribute to improving survival outcomes.

Tumor-infiltrating immune cells (TICs) play a critical role in LUSC pathogenesis, profoundly influencing clinical characteristics and patient survival rates [8,9]. While immunotherapy has revolutionized cancer treatment, it still presents considerable clinical challenges. For patients with advanced LUSC, the median overall survival (OS) remains only 17.1 months, even with the use of PD-(L)1 inhibitors [10]. Although PD-L1 expression serves as a predictive marker for therapeutic benefit, it is not an infallible biomarker. Growing evidence suggests that TICs may serve as indicators of therapeutic response and prognosis in cancer patients, but their diagnostic value remains insufficiently explored.

Macrophages are highly adaptable cells capable of responding to environmental signals and polarizing into M1 or M2 phenotypes. Tumor-associated macrophages (TAMs), a major subset of TICs within the tumor microenvironment, play pivotal roles in the growth, invasion, and

metastasis of LUSC [11–13]. M2 macrophages typically promote tumor progression and suppress anti-tumor immune responses, whereas M1 macrophages can kill tumor cells through the release of inflammatory cytokines, reactive oxygen species, phagocytosis, and the activation of other immune cells [14]. In lung cancers, higher densities of M1 macrophages and lower densities of M2 macrophages are associated with more favorable clinical outcomes. Recent studies have further elucidated the complex regulation of M1 macrophages in LUSC. For instance, *SIRPG* promotes the differentiation of macrophage into M1 phentotype [15], while mutations in the *TTN* gene, a higher tumor mutational burden (TMB), and a more favorable prognosis, are linked to significant enrichment of M1 macrophages [16]. Conversely, *NKX2-1* expression is negatively correlated with M1 macrophages infiltration [17], and *CHRNA6* expression levels are also correlated with the proportion of M1 macrophages [18]. These findings collectively underscore M1 macrophages as a crucial anti-tumor component in LUSC. Accumulating evidence has demonstrated the link between TICs and hub genes involved in tumorigenesis. For instance, elevated Signal-regulatory protein alpha (SIRPα) and *SLC2A1* expression on TAMs correlates with poorer survival and reduced response to adjuvant immunochemotherapy [19,20]. Despite these insights, the diagnostic potential of macrophage-associated signature genes, particularly those regulating the beneficial M1 phenotype, remains underexplored.

This study analyzed seven Gene Expression Omnibus (GEO) datasets and identified 143 M1 macrophage-related differentially expressed genes (DEGs) using the limma, CIBERSORT, and WGCNA algorithms. These DEGs were then used to construct a 5-gene diagnostic signature, derived from the optimal predictive model selected from four machine learning algorithms: Support Vector Machine (SVM), Random Forest (RF), Generalized Linear Model (GLM), and Extreme Gradient Boosting (XGB). The predictive performance of this signature was validated through nomogram visualization, calibration curves, decision curve analysis (DCA), and external validation datasets. Ultimately, *MMP7*, *RTN1*, *ZIC2*, *KLRB1*, and *CXCL13* were identified as potential diagnostic biomarkers associated with LUSC-infiltrating M1 macrophages, highlighting their critical roles in LUSC pathogenesis.

2. Materials and Methods

2.1 Data Collection and Preprocessing

Fig. 1 illustrates the workflow of the bioinformatics analysis. Seven microarray datasets related to LUSC, including GSE19188, GSE2088, GSE30219, GSE33479, GSE33532, GSE21933, and GSE8569, were sourced from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), with detailed dataset information provided in **Supplementary Table 1**. The probe IDs in the series matrix files were annotated to gene symbols according to the corresponding platform files. For genes with multiple probes, the expression values were averaged and normalized using the "limma" package (Version 3.56.2, Bioconductor, USA, https://www.bioconductor.org/) in R software (Version 4.3.1, R Foundation for Statistical Computing, Vienna, Austria) to minimize discrepancies across datasets. Among these datasets, GSE19188, GSE2088, GSE30219, and GSE33479, comprising 150 patients with LUSC and 122 normal controls, were designated as the training dataset. External validation was performed using three additional GEO datasets: GSE33532, GSE21933, and GSE8569.

2.2 Differential Expression Analysis

To integrate the datasets (GSE19188, GSE2088, GSE30219, and GSE33479) into a unified training set, batch effects were corrected using the Combat function from the "sva" package (Version 3.48.0, Bioconductor, USA). DEGs between patients with LUSC and normal controls were identified using the "limma" package. Genes with a $|\log2FC|$ greater than 0.585 and an adjusted p-value below 0.05 were considered significantly differentially expressed. Volcano plots and heatmaps were generated to visualize the DEGs between patients with LUSC and normal controls.

2.3 Immune Cell Infiltration Analysis

The LM22 signature matrix, which includes 22 types of immune cells, was employed to estimate the relative abundances of infiltrating immune cells for each sample in the merged training dataset through the CIBERSORT algorithm (https://cibersort.stanford.edu). The total proportion of the 22 immune cell types in each sample was normalized to 1. The type and distribution of infiltrating immune cells were analyzed using the "ggplot2" package (Version 3.4.4, R Foundation for Statistical Computing, Vienna, Austria) in R software and presented as barplots and boxplots. Significant differences in the relative abundances of immune cell types between patients with LUSC and normal controls were assessed using the Wilcoxon test, with a p-value < 0.05. Spearman correlation analysis of the 22 immune cell types in 150 patients with LUSC was conducted using the "corrplot" package (Version 0.92, R Foundation for Statistical Computing, Vienna, Austria), with significance determined at a p-value < 0.05.

2.4 Weighted Gene Co-Expression Network Analysis (WGCNA)

To identify the significant co-expression module associated with macrophages, the top 50% of genes exhibiting the highest variance were selected for analysis using the WGCNA package (Version 1.72.1, Bioconductor, USA), ensuring robust results [21,22]. The optimal soft-thresholding power was determined using a weighted adjacency matrix, which was then converted into a topological overlap matrix (TOM). A minimum module size of 50 genes was set to ensure biologically relevant modules. Modules were assigned distinct colors based on their TOM dissim-

ilarity measure, calculated as 1 minus the TOM value, using hierarchical clustering. The eigengenes of each module were used to represent the overall gene expression profiles within the module. The significance of each module was evaluated for its association with corresponding immune cells. Additionally, gene significance was assessed by evaluating the strength of the relationship between individual genes and immune cells. Higher gene significance values indicated stronger correlations between specific genes and immune cells [22]. Ultimately, 567 key genes in the blue module associated with M1 macrophages were identified.

2.5 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Analyses

The R package "VennDiagram" was used to identify 142 overlapping genes between DEGs and blue module-associated genes. To investigate the biological functions of these genes, GO functional enrichment and KEGG pathway analyses were conducted using the R packages "clusterProfiler" (Version 4.8.3, Bioconductor, USA), "org.Hs.eg.db" (Version 3.17.0, Bioconductor, USA), and "enrichplot" (Version 1.20.3, Bioconductor, USA). Statistical significance was determined with an adjusted *p*-value threshold of less than 0.05. Results were visualized using bar plots and circle plots, generated with the "ggplot2" (Version 3.4.4, R Foundation for Statistical Computing, Vienna, Austria) and "circlize" (Version 0.4.15, R Foundation for Statistical Computing, Vienna, Austria) packages in R, respectively.

2.6 Constructing the Predictive Model Using Machine-Learning Methods

The 142 overlapping genes were then used as input features for four machine learning models: SVM, RF, GLM, and XGB, aimed at identifying significant predictive genes associated with LUSC. The 150 LUSC samples were randomly split into two datasets: a training dataset (60%, N = 90) and an internal validation dataset (40%, N = 60). These models were executed with default parameters, and the "caret" R package (Version 6.0.94, R Foundation for Statistical Computing, Vienna, Austria) was employed to optimize model parameters *via* grid search. The "DALEX" package (Version 2.4.3, R Foundation for Statistical Computing, Vienna, Austria) was subsequently used to analyze residual distributions and assess feature importance for the four models [23]. To evaluate the predictive accuracy of the models, the internal validation dataset was used for area under the curve (AUC) analysis *via* 5-fold cross-validation. Higher AUC values indicate superior model performance [24]. The most effective machine learning model was identified, and the top five variables from this model were determined as key predictive genes linked to LUSC. Finally, receiver operating characteristic (ROC) curves were generated using three external datasets (GSE33532, GSE21933, and GSE8569) to validate the diagnostic value of the signature derived from the best-performing machine learning model.

2.7 Development and Estimation of the Nomogram Model

A nomogram model was constructed using the "rms" package (Version 6.6.0, R Foundation for Statistical Computing, Vienna, Austria) to predict the likelihood of LUSC. Scores were assigned to five key genes, and their sum formed the "total score", which was used as an indicator of LUSC risk severity. Calibration curve analysis and DCA were also performed to assess the predictive accuracy and clinical utility of the nomogram model.

2.8 Validation Analysis Using Independent External Datasets

The discriminative ability of the predictive model to distinguish between patients with LUSC and healthy controls was validated using three independent external datasets (GSE33532, GSE21933, and GSE8569). ROC analysis was conducted, and the "pROC" R package (Version 1.18.5, R Foundation for Statistical Computing, Vienna, Austria) was employed to generate and visualize the ROC curves.

2.9 Immunohistochemical Validation of Hub Genes Expression and the Predictive Performance of M1 Macrophage-Associated Biomarkers

Forty-eight pairs of LUSC tissue samples were collected from Tianjin Medical University Cancer Institute and Hospital, with approval from the Ethics Committee (No. bc2023152). Informed consent was obtained from all subjects involved in the study. The study was carried out in accordance with the ethical guidelines of the Declaration of Helsinki. Immunohistochemical (IHC) analysis was performed on a cohort of 48 individuals diagnosed with LUSC following the manufacturer's protocol (#PV-9000, ZSGB-BIO, Beijing, China). Tissue sections were dewaxed and hydrated using a gradient ethanol approach. After antigen retrieval in citric acid buffer (pH 6.0) with microwave heating, endogenous peroxidase was blocked for 10 minutes. Sections were then incubated overnight at 4 °C with primary antibodies: ZIC2 rabbit polyclonal antibody (1:100, HA500195, Huabio, Hangzhou, Zhejiang, China), RTN1 rabbit polyclonal antibody (1:50, 15048-1-AP, Proteintech, Wuhan, Hubei, China), MMP7 rabbit polyclonal antibody (1:50, 10374-2-AP, Proteintech, Wuhan, Hubei, China), KLRB1 mouse monoclonal antibody (1:500, 67537-1-lg, Proteintech, Wuhan, Hubei, China), and CXCL13 rabbit monoclonal antibody (1:50, HA722117, Huabio, Hangzhou, Zhejiang, China). After washing, sections were treated with a reaction enhancer and enzyme-linked secondary antibody (sheep anti-rabbit/mouse IgG) (#PV-9000, ZSGB-BIO, Beijing, China). Staining was visualized using DAB chromogen, and counterstaining was performed with hematoxylin. Digital images at $\times$20 magnification were obtained using a BX51 microscope (OLYMPUS, Tokyo, Japan). The evaluation of ZIC2, RTN1, MMP7, KLRB1, and CXCL13 staining followed specific scoring criteria: a positive stained area $<$5% was scored 0; 5%–25% received

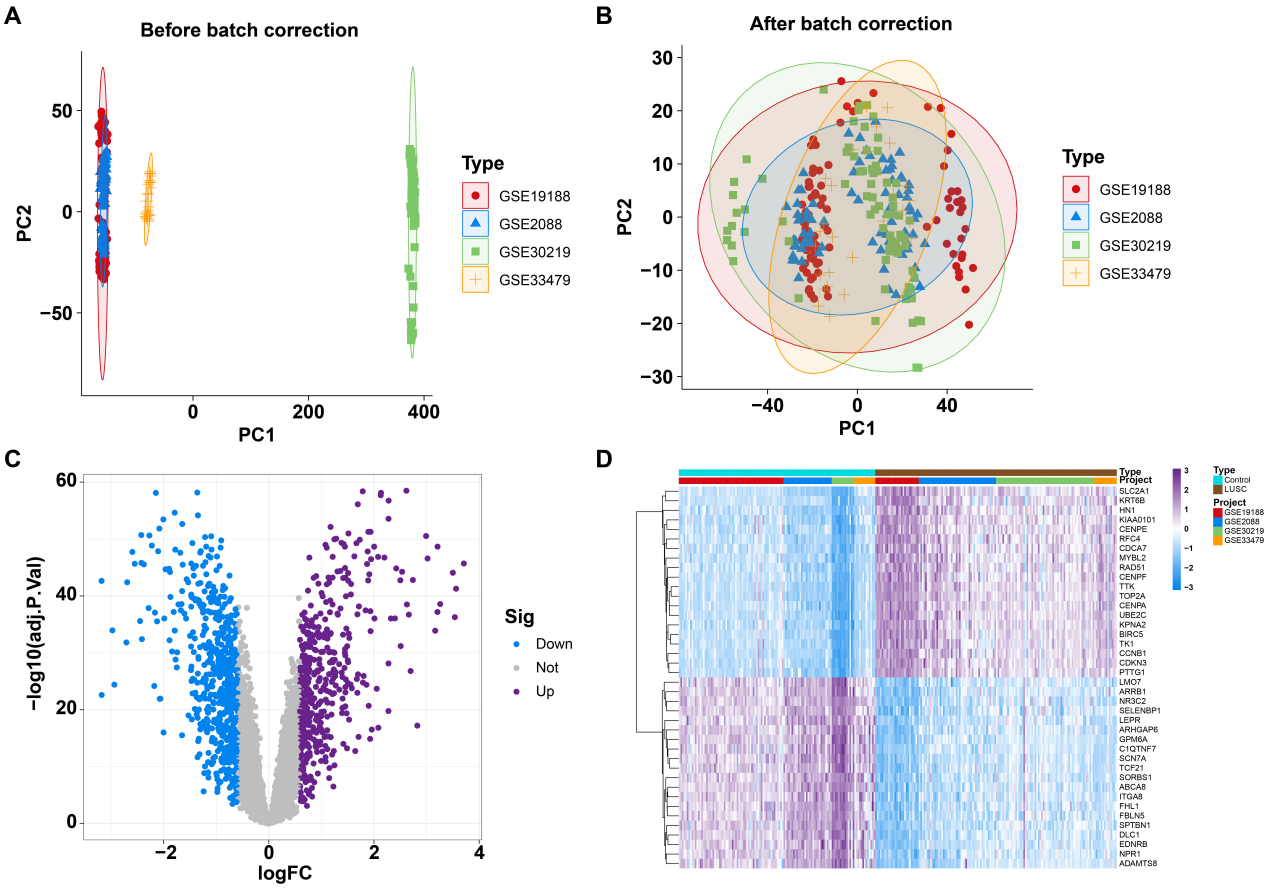


Fig. 1. Identified differentially expressed genes (DEGs) between normal and lung squamous cell carcinoma (LUSC) individuals from four Gene Expression Omnibus (GEO) datasets. (A) Principal component analysis (PCA) of four GEO datasets before batch correction. (B) PCA of four GEO datasets after batch correction. (C) Volcano plot illustrating DEGs between normal and LUSC patients. (D) Heatmap depicting DEGs between normal and LUSC patients.

1 point; 26%–50% equaled 2 points; 51%–75% was given 3 points; and areas exceeding 75% were assigned 4 points. For color intensity, no staining was scored 0 points; yellow staining received 1 point; brown staining was assigned 2 points; and tan staining received 3 points. The immunostaining score was calculated by multiplying the color intensity and positive stained area scores.

2.10 Statistical Analysis

Statistical analyses were performed using R software (Version 4.3.1). For comparisons between two groups, unpaired Student's *t*-tests were applied for normally distributed variables, and Wilcoxon rank-sum tests were used for non-normally distributed variables. Spearman correlation analysis was conducted to estimate correlation coefficients. Statistical significance was defined as $p < 0.05$ in two-tailed tests, unless otherwise specified.

3. Results

3.1 Identifying DEGs Between LUSC and Normal Individuals

To identify differential expression between lung squamous cell carcinoma (LUSC) and normal samples, four GEO datasets (GSE19188, GSE2088, GSE30219, and GSE33479) underwent batch correction and were subsequently merged into a large sample pool to form the training cohort. Principal component analysis (PCA) and boxplot analyses confirmed that batch effects were effectively minimized across the four original datasets (Fig. 1A,B; **Supplementary Fig. 1**). Differential expression analysis identified a total of 1032 DEGs, including 466 upregulated and 566 downregulated genes, between 122 control samples and 150 LUSC samples (Fig. 1C; **Supplementary Table 2**). Heatmaps depicted the top 20 genes with the most significant expression changes, either upregulated or downregulated (Fig. 1D).

3.2 Immune Characteristics in LUSC

Next, we explored the immune landscape in LUSC by assessing the abundance of 22 immune cell types using the CIBERSORT algorithm (Fig. 2A). Significant differences between control and LUSC samples were analyzed using the Wilcoxon rank-sum test. Boxplots and statistical results revealed that T cells CD4 memory resting, monocytes, mast cells resting, T cells CD4 memory activated, and M1 macrophages were the top five immune cell types showing

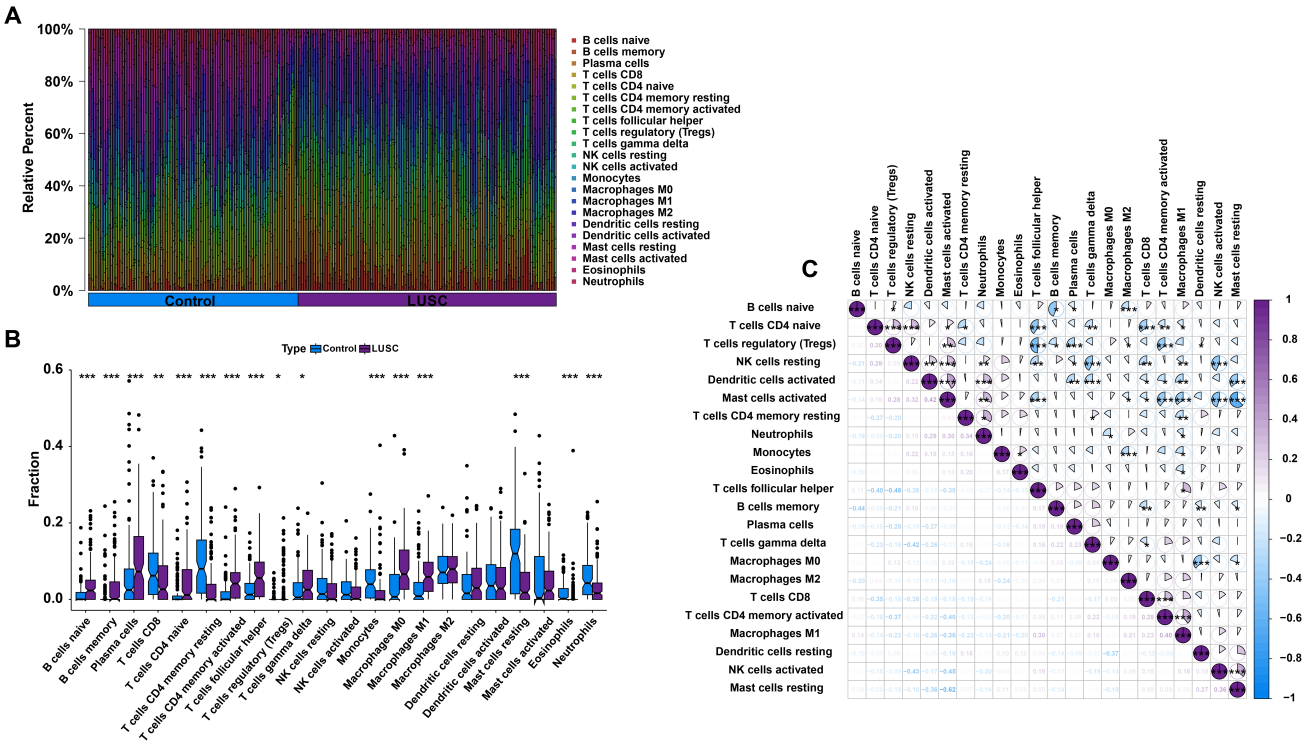


Fig. 2. Analysis of immune characteristics in LUSC. (A) Barplot showing the distribution of 22 types of infiltrating immune cells in control and LUSC samples. (B) Differential expression analysis of 22 types of infiltrating immune cells between control and LUSC samples. (C) Correlation analysis among 22 types of infiltrating immune cells in LUSC patients. $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.

significant differences between control and LUSC groups (Fig. 2B; **Supplementary Table 3**). Correlation analyses of the 22 types of infiltrating immune cells in patients with LUSC are shown in Fig. 2C.

3.3 Identification of M1 Macrophage-Related Genes

To gain insight into the genes associated with M1 macrophages, the WGCNA algorithm was applied to identify gene modules strongly associated with immune cell types. The optimal soft-thresholding parameters were determined as $\beta = 3$ and $R^2 = 0.9$ for constructing a scale-free network (Fig. 3A). Using the dynamic tree-cutting algorithm, seven distinct co-expression modules, each represented by a unique color, were generated (Fig. 3B). Genes from these modules were then evaluated for their co-expression and adjacency relative to the 22 types of immune cells. By integrating the results of WGCNA and immune cell differential analysis, the blue module, consisting of 567 genes, was found to have a significantly stronger association with M1 macrophages ($R = 0.37$, $p = 7.8 \times 10^{-20}$), and was selected for further investigation. These genes within the blue module were positively correlated with M1 macrophages (Fig. 3C,D).

3.4 Biological Function of M1 Macrophage-Related DEGs

To identify M1 macrophage-related DEGs, the DEGs were intersected with the genes from the WGCNA blue module. A Venn diagram revealed that 143 M1 macrophage-related DEGs were obtained (Fig. 4A; **Supplementary Table 4**). GO analysis indicated that these genes are involved in leukocyte- or lymphocyte-mediated immunity, antigen processing and presentation of peptide antigens, assembly of major histocompatibility complex (MHC) class I or II protein complexes, regulation of immune response signaling pathways, and positive regulation of immune cell-cell adhesion (Fig. 4B; **Supplementary Table 5**). KEGG pathway analysis revealed that these genes are associated with signaling pathways in rheumatoid arthritis, asthma, Fc epsilon RI signaling, chemokine signaling, and Tumor Necrosis Factor Signaling (TNF) signaling (Fig. 4C; **Supplementary Table 6**).

3.5 Development and Evaluation of Machine Learning Models

To explore the potential of these M1 macrophage-related DEGs for LUSC diagnosis, four machine learning models—SVM, XGB, RF, and GLM—were constructed based on the expression profiles of 143 M1 macrophage-related DEGs in the training dataset to identify diagnostic genes for LUSC. Residual distribution analysis revealed that SVM and XGB exhibited smaller residuals compared to RF and GLM (Fig. 5A,B). The top 10 feature variables for each model were ranked by their root mean square error (RMSE), and their corresponding feature importance

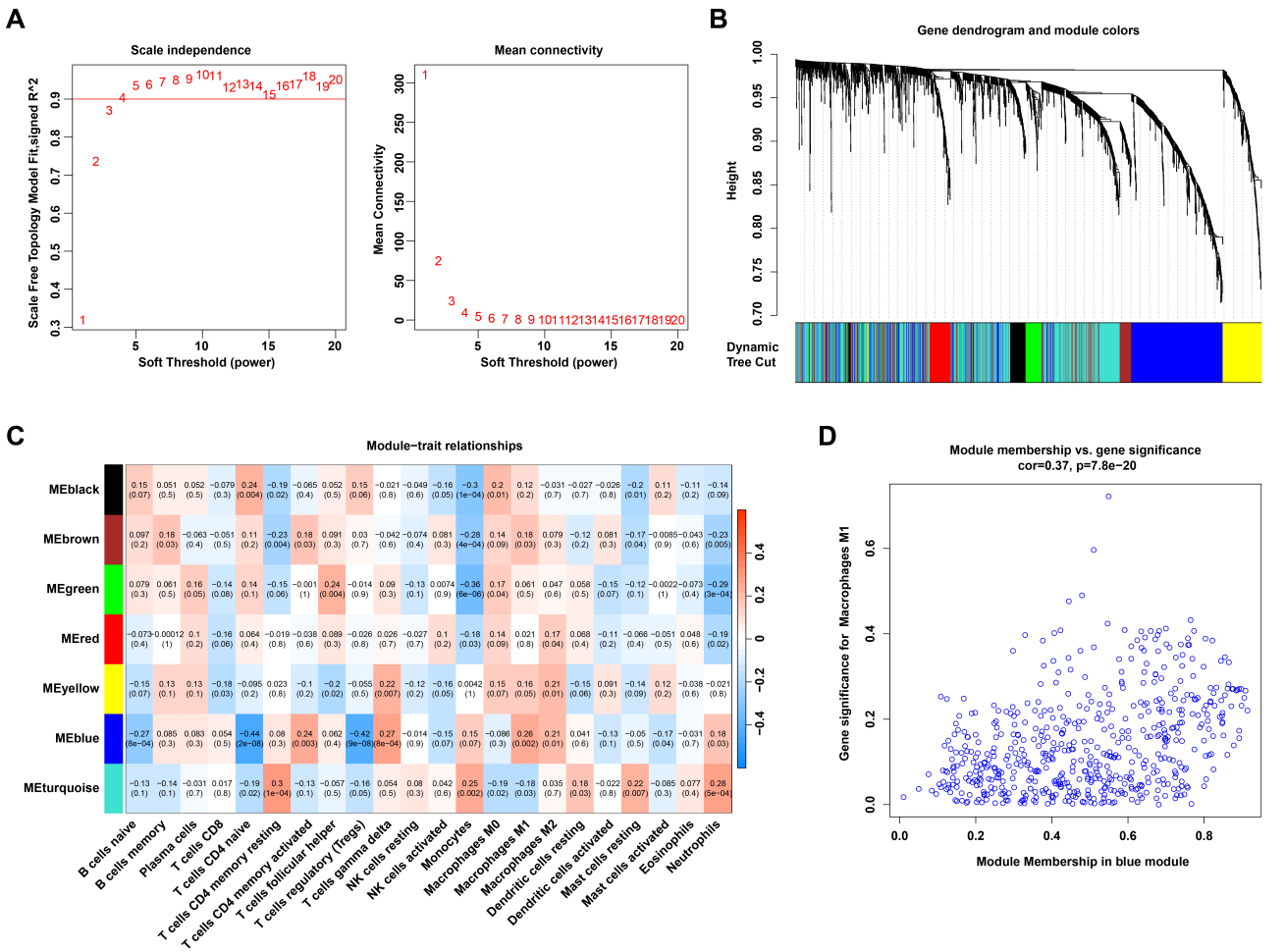


Fig. 3. Identification of M1 macrophage-related genes through WGCNA. (A) The process for selecting the soft threshold power. (B) Cluster dendrogram of genes grouped into co-expression modules, with various modules represented by distinct colors. (C) Correlation analysis between module eigengenes and 22 types of infiltrating immune cells. (D) Scatter plot showing the relationship between module membership in the blue module and gene significance for M1 macrophages.

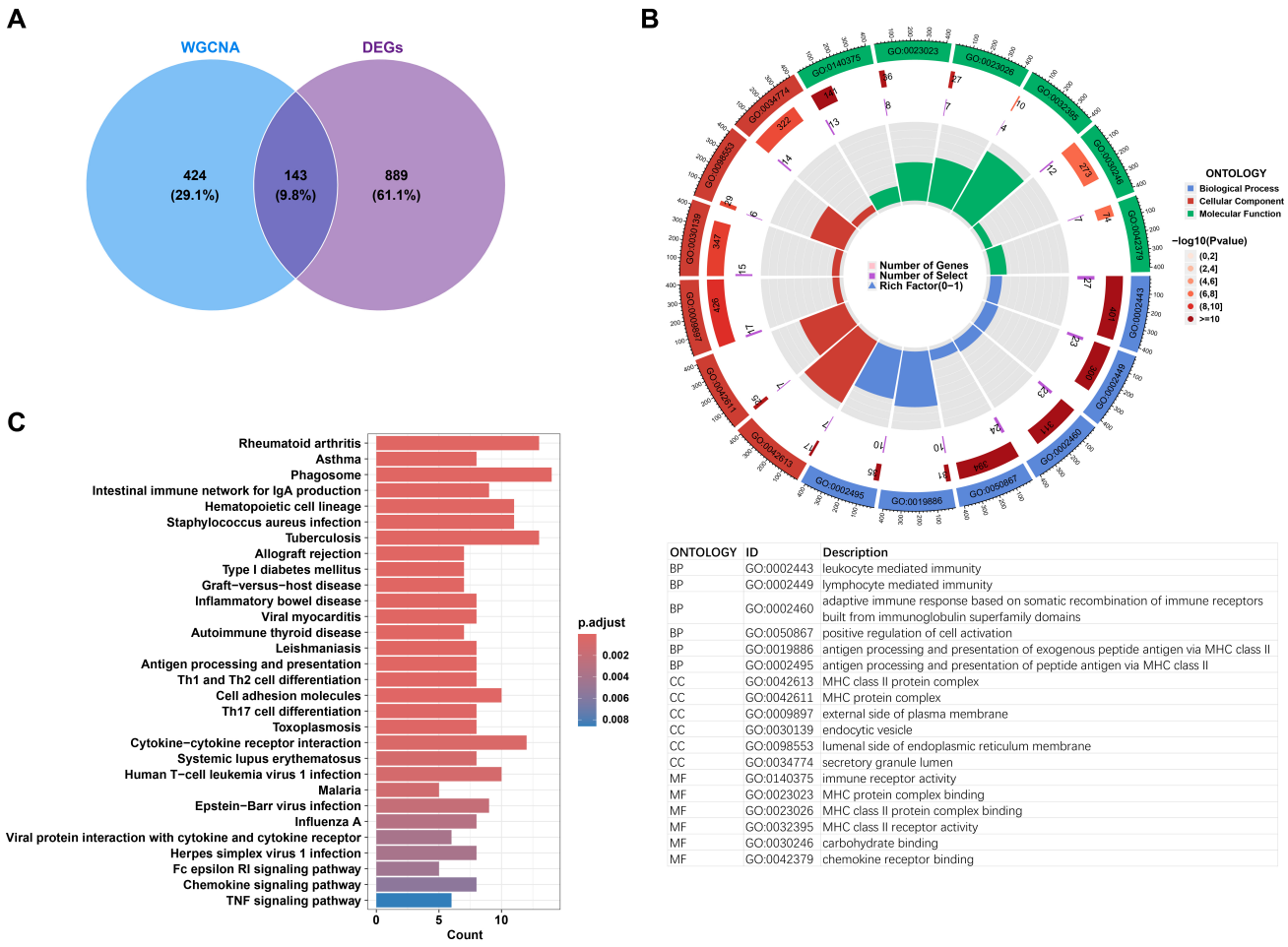
was visualized in Fig. 5C. Additionally, the performance of each machine learning model was evaluated using ROC curves generated through five-fold cross-validation on the testing datasets. The results showed that the SVM model achieved the highest AUC value ($AUC_{SVM} = 0.995$, $AUC_{RF} = 0.988$, $AUC_{XGB} = 0.983$, and $AUC_{GLM} = 0.622$) (Fig. 5D). These findings demonstrate that the SVM model was the most effective for distinguishing patients with LUSC. Consequently, *MMP7*, *RTN1*, ZIC2, *KLRB1*, and *CXCL13* were identified as novel predictive genes (5-gene signature) for LUSC due to their high significance in the SVM model.

3.6 Evaluation of Machine Learning Models in External Testing Databases

To assess the predictive power of the 5 hub genes, a nomogram was developed to estimate the risk associated with LUSC (Fig. 6A). The calibration curve showed excellent agreement between actual and predicted risks (Fig. 6B). DCA further validated the high accuracy and clinical utility of the nomogram (Fig. 6C). The predictive performance of the 5 hub genes was also evaluated in three independent external testing datasets. The AUC values for GSE33532, GSE21933, and GSE8569 were all 1.000, as demonstrated by the ROC curves (Fig. 6D–F). These results confirm that the diagnostic signature developed in this study exhibits robust performance in distinguishing LUSC from normal individuals across both internal and external validation cohorts.

3.7 Immunohistochemistry Validation of the Predictive Performance of M1 Macrophage-Associated Biomarkers

To validate the predictive performance of M1 macrophage-associated biomarkers, IHC experiments were performed to evaluate the protein expression levels of candidate genes in tissue samples. IHC results show ZIC2 is predominantly expressed in the nuclei of lung cancer cells, with elevated levels in cancer tissues. RTN1 is mainly found in the membrane of lung cancer cells, but its expression is downregulated in these tissues. MMP7, a secreted protein, is present in the nucleus, cytoplasm, and extracellular matrix, with elevated expression in lung cancer tissues. KLRB1, a transmembrane protein, is downregulated



ONTOLOGY	ID	Description
BP	GO:0002443	leukocyte mediated immunity
BP	GO:0002449	lymphocyte mediated immunity
BP	GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
BP	GO:0050867	positive regulation of cell activation
BP	GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II
BP	GO:0002495	antigen processing and presentation of peptide antigen via MHC class II
CC	GO:0042613	MHC class II protein complex
CC	GO:0042611	MHC protein complex
CC	GO:0009897	external side of plasma membrane
CC	GO:0030139	endocytic vesicle
CC	GO:0098553	lumenal side of endoplasmic reticulum membrane
CC	GO:0034774	secretory granule lumen
MF	GO:0140375	immune receptor activity
MF	GO:0023023	MHC protein complex binding
MF	GO:0023026	MHC class II protein complex binding
MF	GO:0032395	MHC class II receptor activity
MF	GO:0030246	carbohydrate binding
MF	GO:0042379	chemokine receptor binding

Fig. 4. Biological functional analysis of differentially expressed M1 macrophage-related genes. (A) Venn diagram illustrating the overlapping and unique differentially expressed M1 macrophage-related genes. (B) Circular visualization of Gene Ontology (GO) analysis highlighting the biological functions associated with differentially expressed M1 macrophage-related genes. (C) Barplot depicting the signaling pathways enriched in the differentially expressed M1 macrophage-related genes.

in lung cancer tissues. CXCL13 is expressed in the cytoplasm and extracellular matrix, with upregulation in lung cancer tissues. The results showed that the protein expression of those 5 hub genes was largely consistent with their corresponding mRNA expression levels in both the training dataset and the TCGA dataset (Fig. 7A,B; **Supplementary Fig. 2**; **Supplementary Table 2**). Furthermore, the predictive efficacy of the five hub genes was assessed based on the IHC results, yielding an AUC value of 0.992 in the ROC curve analysis (Fig. 7C). These results confirm the reliability of the data mining outcomes and highlight their potential for LUSC diagnosis.

4. Discussion

Macrophages are notably abundant in lung cancer tissues and play a pivotal role in determining disease progression and outcomes [25]. Traditionally, macrophages are categorized into two main types: classically activated (M1) and alternatively activated (M2) macrophages [26]. In LUSC, higher densities of M1-like macrophages and an increased M1/M2 ratio within tumor islets and stroma are strongly associated with improved patient survival [25]. Most patients with LUSC are diagnosed at advanced stages, leading to a low five-year survival rate [10,27]. This study utilized seven GEO microarray datasets, which were divided into training and testing cohorts, to identify novel diagnostic signatures for LUSC, with the goal of enhancing diagnostic accuracy and improving survival rates. A total of 143 M1 macrophage-related DEGs were identified using the limma, CIBERSORT, and WGCNA algorithms. These genes were significantly involved in antigen processing and presentation, MHC protein complex assembly, leukocyte- or lymphocyte-mediated immunity, and regulation of immune response signaling pathways—key processes influencing LUSC progression [28,29]. Through application of the optimal machine learning model, SVM, during both training and external validation, *MMP7*, *RTN1*, *ZIC2*, *KLRB1*, and *CXCL13* were identified as novel five hub genes associated with LUSC-infiltrating M1 macrophages, playing critical roles in the pathogenesis of LUSC.

TICs are closely linked to hub genes involved in tumorigenesis [30–33]. For instance, the expression of

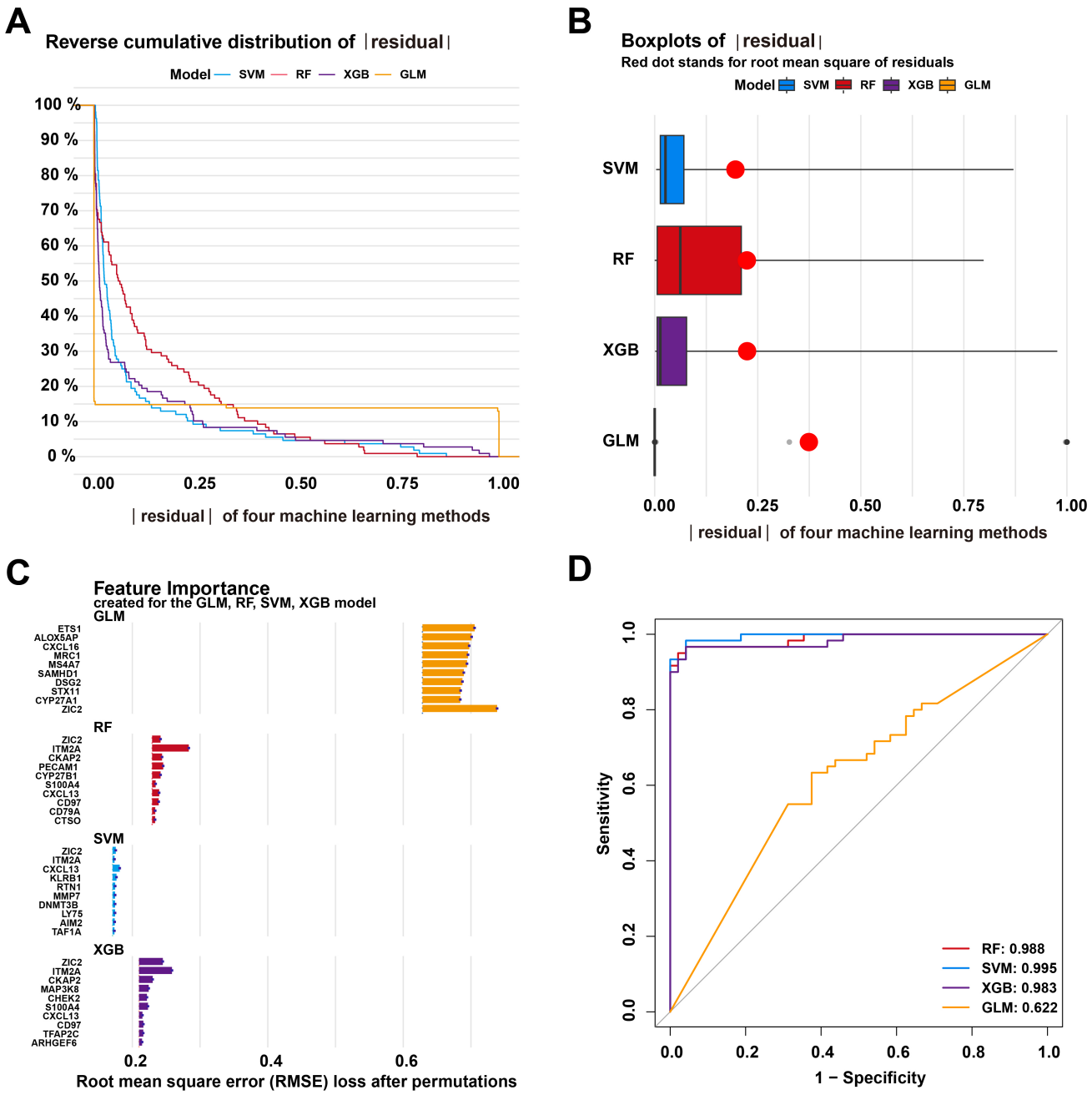


Fig. 5. Development and evaluation of the machine learning models based on 143 M1 macrophage-related genes. (A) The cumulative residual distribution of Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Random Forest (RF), and Generalized Linear Model (GLM) models. (B) The residual boxplots of the four machine learning models. (C) The important features of the four machine learning models. (D) Receiver operating characteristic (ROC) analysis of these models using 5-fold cross-validation in the internal testing dataset.

purinergic receptor P2RY13 was positively correlated with dendritic cell infiltration across various tumor types [31]. Zhang *et al.* [34] identified four key hub genes—*LAPTM5*, *C1QC*, *CSF1R*, and *SLCO2B1*—that promote the exhaustion of $CD8^+$ T cells in LUSC tumor tissues. Additionally, COL1A1, COL4A1, COL12A1, and PDGFRB were found to be potential prognostic biomarkers associated with M2 macrophage infiltration in gastric cancer [30]. In contrast, CXCL6 and POSTN were overexpressed in M0 macrophage-enriched clusters and served as poor prognostic factors in hepatocellular carcinoma [35]. Based on immune cell infiltration analysis in the training datasets, a significant increase in tumor-infiltrating M1 macrophages was observed in LUSC tissues compared to control samples. Furthermore, *MMP7*, *RTN1*, *ZIC2*, *KLRB1*, and *CXCL13* were found to be significantly correlated with M1 macrophage infiltration in LUSC.

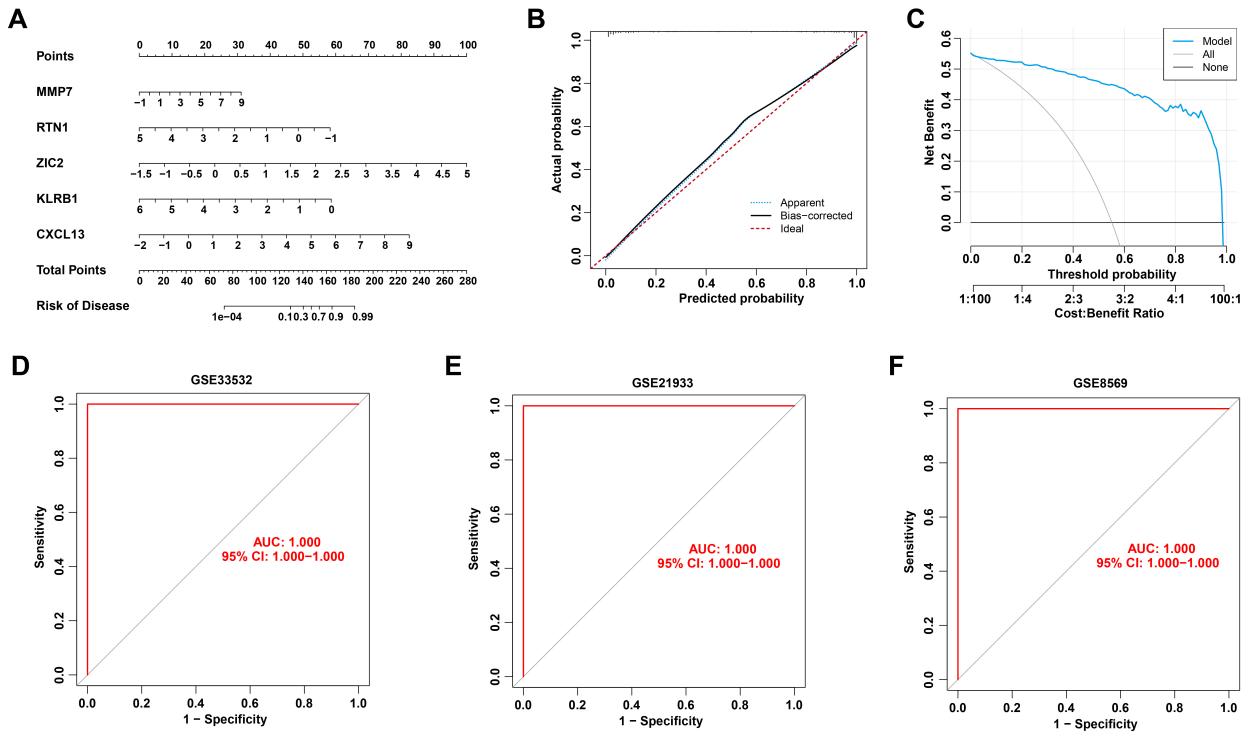


Fig. 6. Validation of the 5-gene diagnostic signature. (A) Nomogram of 5 hub genes for predicting the risk of LUSC. (B,C) Calibration curve (B) and decision curve analysis (DCA) (C) for validating the predictive accuracy of the nomogram. (D–F) ROC curve of the 5 hub genes in external GSE33532 (D), GSE21933 (E), and GSE8569 (F) datasets.

Our study focused on the application of machine learning to further refine diagnostic strategies. Four machine learning models—SVM, GLM, RF, and XGB—were employed to identify the optimal model by comparing their predictive performances. The SVM model outperformed the others, achieving the highest AUC value of 0.995. The 5-gene SVM model consistently demonstrated high accuracy in predicting LUSC across three externally validated datasets. Because our model is built on a highly refined set of 143 genes, carefully selected for their strong association with both DEGs in LUSC and M1 macrophages. This approach effectively reduces noise and enhances biological relevance. Additionally, the SVM algorithm is particularly adept at finding optimal boundaries in high-dimensional data, which is suited for this work. As a result, the AUC value has reached 1.000 in each dataset. These findings suggest that the model offers a novel diagnostic approach for LUSC. To our knowledge, no predictive signature based on M1 macrophage-related DEGs has been previously reported. Joon *et al.* [36] identified 40 genes using an XGBoost model that exhibited better predictive performance for LUSC than SVM, RF, k-Nearest Neighbor (kNN), and Decision Tree (DT) models. Ye *et al.* [37] identified ten feature miRNAs using the SVM model that could distinguish LUSC tissues from adjacent tissues. Therefore, our 5-gene SVM model offers the advantage of achieving similar predictive accuracy with fewer genes, while also providing insights into the relationship between M1 macrophages and LUSC.

Additionally, a nomogram model was developed for diagnosing LUSC based on the expression profiles of *MMP7*, *RTN1*, *ZIC2*, *KLRB1*, and *CXCL13*, with further validation of its predictive power and clinical utility. MMP7 exhibits anti-inflammatory effects [38] and is upregulated in various tumors, including LUSC [39]. Takayuki Shiomi *et al.* [40] have demonstrated CD151 may function as a possible docking molecule for proMMP-7 activation in LUAD. Loss of MMP7 resulted in M1 macrophage polarization within H. pylori-infected stomachs [41]. Reticulon 1 (RTN1), associated with the endoplasmic reticulum, is a specific marker for neurological diseases and cancers [42]. McGonigle *et al.* [43] demonstrated that RTN1 regulates macrophage differentiation and migration. Alterations in RTN1 expression can further lead to an increased proportion of immune effector cells, including $CD4^+$ T cells and $CD8^+$ T cells, which are often associated with a favorable prognosis. Importantly, *RTN1* expression was significantly correlated with the expression of multiple immune checkpoints, such as CD274 (PD-L1), CTLA4, HAVCR2, LAG3, PDCD1 (PD-1), PDCD1LG2 (PD-L2), TIGIT, and SIGLEC15, suggesting a potential immunomodulatory role that may involve macrophage-mediated regulation of the tumor immune microenvironment [44]. Zinc finger protein of cerebellum 2 (ZIC2) drives immunosuppressive M2 macrophage polarization in nasopharyngeal carcinoma by activating JUNB transcription and stimulating M-CSF secretion [45]. It also correlates with a suppressed tumor immune microenvironment, showing negative associations

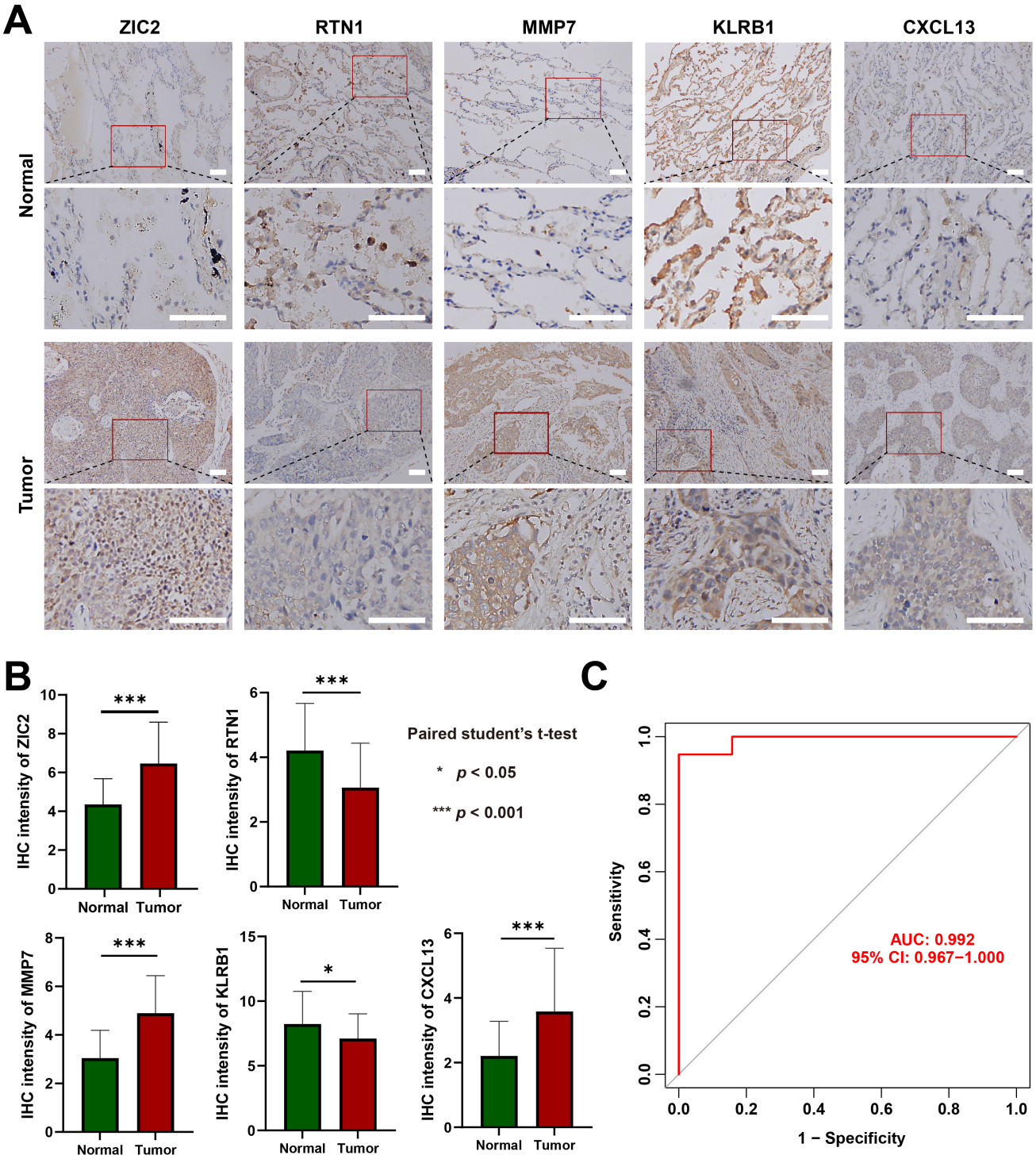


Fig. 7. Immunohistochemistry validation of the predictive performance of M1 macrophage-associated biomarkers. (A) Representative images of immunohistochemical staining for Zinc finger protein ZIC 2 (ZIC2), Reticulon-1 (RTN1), Matrix metalloproteinase-7 (MMP7), Killer cell lectin-like receptor subfamily B member 1 (KLRB1) and C-X-C motif chemokine 13 (CXCL13) in normal lung tissues and LUSC tissues (scale bar, 50 µm). (B) Statistical analysis of immunohistochemistry results in normal lung tissues and LUSC tissues. Data are mean ± SD. n = 48. *$p < 0.05$, and ***$p < 0.001$. (C) ROC curve of the 5-gene signature in the immunohistochemical (IHC) data.

with tumor-infiltrating lymphocytes and immune checkpoint expression in breast cancer [46]. Furthermore, ZIC2 enhances cancer stem cell traits by upregulating OCT4 expression in lung adenocarcinoma cells [47]. This study provides the first evidence linking ZIC2 to LUSC. KLRB1 (CD161), a C-type lectin-like receptor predominantly expressed on NK cells and CD8 T cells, has been identified together with its ligand LLT1 (CLEC2D) as an emerging

immune checkpoint [48]. Studies indicate that high *KLRB1* expression is associated with increased infiltration of M1 macrophages and CD8 T cells in the tumor microenvironment [49]. This effect is likely mediated indirectly through the secretion of pro-inflammatory cytokines such as Interferon gamma (IFN-γ) by activated KLRB1 immune cells, which promotes the polarization of macrophages toward an anti-tumor M1 phenotype. CXCL13 chemokine, originally discovered as a B-cell chemoattractant, is widely implicated in tumor development and progression [50]. CXCL13 secretion is a hallmark of M2-polarized macrophages and is widely associated with pro-tumor functions across multiple cancers. In lung and renal cell carcinoma models, $CXCL13^+$ macrophages promote tumor growth, invasion, and metastasis, with genetic deletion of CXCL13 or its receptor CXCR5 significantly suppressing cancer development [51]. Importantly, $CD68^+$ macrophages within the tumors were the origin of the observed increased CXCL13 levels [52].

While our study used comprehensive bioinformatics analyses and machine learning models, there are limitations that must be acknowledged. First, although machine learning models and comprehensive bioinformatics analyses were used and validated in both training and external validation datasets, the external validation dataset has a relatively small sample size and lacks standardized clinical phenotype information. This limitation may affect the generalizability of the model, suggesting that more detailed clinical data are needed to fully confirm its predictive performance. Second, further experimental studies are required to explore the functional relationship between the 5 hub genes and M1 macrophages.

These five genes are expected to be included in routine pathological diagnosis and liquid biopsy, such as detecting the expression levels of these genes in circulating immune cells or exosomes from blood samples for non-invasive early detection for LUSC. However, it still faces challenges such as technical standardization, clinical validation, and improvement of sensitivity. The prognostic and diagnostic value of these five genes must be validated in large-scale, multi-center clinical trials to ensure their applicability across diverse patient populations. With the development of big data, artificial intelligence and multi-omics technologies, it is expected that these barriers will be overcome in the future to achieve early diagnosis and precise treatment of LUSC.

5. Conclusion

This study identified 143 M1 macrophage-related DEGs. Among the machine learning models evaluated, the SVM model based on five genes was selected as the optimal approach after comparison with four other models. The genes *MMP7*, *RTN1*, *ZIC2*, *KLRB1*, and *CXCL13* exhibited high predictive accuracy for LUSC, potentially improving diagnostic precision and informing more effective treatment strategies for the disease.

Availability of Data and Materials

Publicly available datasets were analyzed in this study. These data can be found here: https://www.ncbi.nlm.nih.gov/geo/. The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

HTD, ZLW and QZZ collected, analyzed and generated data. HTD, ZLW and QZZ wrote the manuscript. QZZ and ZLW initiated the idea and supplied resources. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The clinical experiments were approved by the Tianjin Medical University Cancer Institute and Hospital Medical Research Ethics (No. bc2023152). All participators signed the written informed consent. The study was carried out in accordance with the guidelines of the Declaration of Helsinki.

Acknowledgment

We thank Wei Su (Tianjin Medical University Cancer Institute and Hospital) for providing clinical experiments for our study.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 82373028 and 82400730], and the Natural Science Foundation of Tianjin [grant numbers 21JCQNJC00130].

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/FBL44661.

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a Cancer Journal for Clinicians. 2024; 74: 229–263. https://doi.org/10.3322/caac.21834.

[2] Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, *et al*. Cancer incidence and mortality in China, 2022. Journal of the National Cancer Center. 2024; 4: 47–53. https://doi.org/10.1016/j.jncc.2024.01.006.

[3] Khodabakhshi Z, Mostafaei S, Arabi H, Oveisi M, Shiri I, Zaidi H. Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature. Computers in Biology and

Medicine. 2021; 136: 104752. https://doi.org/10.1016/j.compbiomed.2021.104752.
[4] Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet (London, England). 2021; 398: 535–554. https://doi.org/10.1016/S0140-6736(21)00312-3.
[5] Woodard GA, Jones KD, Jablons DM. Lung Cancer Staging and Prognosis. Cancer Treatment and Research. 2016; 170: 47–75. https://doi.org/10.1007/978-3-319-40389-2_3.
[6] Wang C, Zhang J, Wang H, Chen R, Lu M. Family with sequence similarity 83, member A (FAM83A) inhibits ferroptosis via the Wnt/β-catenin pathway in lung squamous cell cancer. Cell Death Discovery. 2024; 10: 332. https://doi.org/10.1038/s41420-024-02101-4.
[7] Hirsch FR, Kerr KM, Bunn PA, Jr, Kim ES, Obasaju C, Pérol M, *et al*. Molecular and Immune Biomarker Testing in Squamous-Cell Lung Cancer: Effect of Current and Future Therapies and Technologies. Clinical Lung Cancer. 2018; 19: 331–339. https://doi.org/10.1016/j.cllc.2018.03.014.
[8] Wang J, Zhou J, Zhou Q, Qi Y, Zhang P, Yan C, *et al*. Dysregulated Th1 cells in lung squamous cell carcinoma. Journal of Leukocyte Biology. 2022; 112: 1567–1576. https://doi.org/10.1002/JLB.1MA0422-208R.
[9] Chen M, Ma P, Zhang Y, Wang D, Yu Z, Fu Y, *et al*. Divergent tumor and immune cell reprogramming underlying immunotherapy response and immune-related adverse events in lung squamous cell carcinoma. Journal for Immunotherapy of Cancer. 2023; 11: e007305. https://doi.org/10.1136/jitc-2023-007305.
[10] Lau SCM, Pan Y, Velcheti V, Wong KK. Squamous cell lung cancer: Current landscape and future therapeutic options. Cancer Cell. 2022; 40: 1279–1293. https://doi.org/10.1016/j.ccell.2022.09.018.
[11] Chittezhath M, Dhillon MK, Lim JY, Laoui D, Shalova IN, Teo YL, *et al*. Molecular profiling reveals a tumor-promoting phenotype of monocytes and macrophages in human cancer progression. Immunity. 2014; 41: 815–829. https://doi.org/10.1016/j.immuni.2014.09.014.
[12] De Zuani M, Xue H, Park JS, Dentro SC, Seferbekova Z, Tessier J, *et al*. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. Nature Communications. 2024; 15: 4388. https://doi.org/10.1038/s41467-024-48700-8.
[13] Zheng X, Weigert A, Reu S, Guenther S, Mansouri S, Bassaly B, *et al*. Spatial Density and Distribution of Tumor-Associated Macrophages Predict Survival in Non-Small Cell Lung Carcinoma. Cancer Research. 2020; 80: 4414–4425. https://doi.org/10.1158/0008-5472.CAN-20-0069.
[14] Zhao CC, Han QJ, Ying HY, Gu XX, Yang N, Li LY, *et al*. TNFSF15 facilitates differentiation and polarization of macrophages toward M1 phenotype to inhibit tumor growth. Oncoimmunology. 2022; 11: 2032918. https://doi.org/10.1080/2162402X.2022.2032918.
[15] Mao G, Li J, Wang N, Yu H, Han S, Xiang M, *et al*. SIRPG promotes lung squamous cell carcinoma pathogenesis via M1 macrophages: a multi-omics study integrating data and Mendelian randomization. Frontiers in Oncology. 2024; 14: 1392417. https://doi.org/10.3389/fonc.2024.1392417.
[16] Zou S, Ye J, Hu S, Wei Y, Xu J. Mutations in the *TTN* Gene are a Prognostic Factor for Patients with Lung Squamous Cell Carcinomas. International Journal of General Medicine. 2022; 15: 19–31. https://doi.org/10.2147/IJGM.S343259.
[17] Lin H, Wang J, Shi Q, Wu M. Significance of *NKX2-1* as a biomarker for clinical prognosis, immune infiltration, and drug therapy in lung squamous cell carcinoma. PeerJ. 2024; 12: e17338. https://doi.org/10.7717/peerj.17338.
[18] Zhang F, Zhang M, Yuan X, Tao Y, Wang J. Involvement of CHRNA6 in the Immune Response in Lung Squamous Cell Carcinoma and its Potential as a Drug Target for the Disease. Current Pharmaceutical Design. 2023; 29: 2091–2100. https://doi.org/10.2174/1381612829666230901143203.
[19] Nagano T, Takada K, Narutomi F, Kinoshita F, Akamine T, Kohno M, *et al*. Clinical Significance of SIRPα Expression on Tumor-Associated Macrophages in Patients with Lung Squamous Cell Carcinoma. Annals of Surgical Oncology. 2024; 31: 6309–6319. https://doi.org/10.1245/s10434-024-15649-3.
[20] Hao B, Dong H, Xiong R, Song C, Xu C, Li N, *et al*. Identification of SLC2A1 as a predictive biomarker for survival and response to immunotherapy in lung squamous cell carcinoma. Computers in Biology and Medicine. 2024; 171: 108183. https://doi.org/10.1016/j.compbiomed.2024.108183.
[21] Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of Statistical Software. 2012; 46: i11.
[22] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559. https://doi.org/10.1186/1471-2105-9-559.
[23] Wang Z, Zhang L, Huang T, Yang R, Cheng H, Wang H, *et al*. Developing an explainable machine learning model to predict the mechanical ventilation duration of patients with ARDS in intensive care units. Heart & Lung: the Journal of Critical Care. 2023; 58: 74–81. https://doi.org/10.1016/j.hrtlng.2022.11.005.
[24] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12: 77. https://doi.org/10.1186/1471-2105-12-77.
[25] Wang X, Wu Y, Gu J, Xu J. Tumor-associated macrophages in lung carcinoma: From mechanism to therapy. Pathology, Research and Practice. 2022; 229: 153747. https://doi.org/10.1016/j.prp.2021.153747.
[26] Funes SC, Rios M, Escobar-Vera J, Kalergis AM. Implications of macrophage polarization in autoimmunity. Immunology. 2018; 154: 186–195. https://doi.org/10.1111/imm.12910.
[27] Luo YH, Luo L, Wampfler JA, Wang Y, Liu D, Chen YM, *et al*. 5-year overall survival in patients with lung cancer eligible or ineligible for screening according to US Preventive Services Task Force criteria: a prospective, observational cohort study. The Lancet. Oncology. 2019; 20: 1098–1108. https://doi.org/10.1016/S1470-2045(19)30329-8.
[28] Hao B, Zhang Z, Lu Z, Xiong J, Fan T, Song C, *et al*. Single-cell RNA sequencing analysis revealed cellular and molecular immune profiles in lung squamous cell carcinoma. Translational Oncology. 2023; 27: 101568. https://doi.org/10.1016/j.tranon.2022.101568.
[29] Gómez-López S, Whiteman ZE, Janes SM. Mapping lung squamous cell carcinoma pathogenesis through in vitro and in vivo models. Communications Biology. 2021; 4: 937. https://doi.org/10.1038/s42003-021-02470-x.
[30] Liu B, Ma X, Ha W. Identification of Potential Prognostic Biomarkers Associated With Macrophage M2 Infiltration in Gastric Cancer. Frontiers in Genetics. 2022; 12: 827444. https://doi.org/10.3389/fgene.2021.827444.
[31] Wang J, Shi W, Miao Y, Gan J, Guan Q, Ran J. Evaluation of tumor microenvironmental immune regulation and prognostic in lung adenocarcinoma from the perspective of purinergic receptor P2Y13. Bioengineered. 2021; 12: 6286–6304. https://doi.org/10.1080/21655979.2021.1971029.
[32] Zhu G, Xiong Z, Chen W, Zhu Z, Wang W. Identification of key biomarkers and related immune cell infiltration in cervical cancer tissue based on bioinformatics analysis. Scientific Reports. 2023; 13: 10121. https://doi.org/10.1038/s41598-023-37346-z.
[33] Xu J, Yang Z, Xie W, Wan R, Li C, Fei K, *et al*. A prognostic and immunotherapeutic predictive model based on the cell-originated characterization of tumor microenvironment in lung adenocarcinoma. iScience. 2023; 26: 106616. https://doi.org/10.1016/j.isci.2023.106616.
[34] Zhang T, Yang H, Sun B, Yao F. Four hub genes regulate tumor infiltration by immune cells, antitumor immunity in the tumor

microenvironment, and survival outcomes in lung squamous cell carcinoma patients. Aging. 2021; 13: 3819–3842. https://doi.org/10.18632/aging.202351.

[35] Farha M, Jairath NK, Lawrence TS, El Naqa I. Characterization of the Tumor Immune Microenvironment Identifies M0 Macrophage-Enriched Cluster as a Poor Prognostic Factor in Hepatocellular Carcinoma. JCO Clinical Cancer Informatics. 2020; 4: 1002–1013. https://doi.org/10.1200/CCI.20.00077.

[36] Joon HK, Thalor A, Gupta D. Machine learning analysis of lung squamous cell carcinoma gene expression datasets reveals novel prognostic signatures. Computers in Biology and Medicine. 2023; 165: 107430. https://doi.org/10.1016/j.compbiomed.2023.107430.

[37] Ye Z, Sun B, Xiao Z. Machine learning identifies 10 feature miRNAs for lung squamous cell carcinoma. Gene. 2020; 749: 144669. https://doi.org/10.1016/j.gene.2020.144669.

[38] Vandenbroucke RE, Vanlaere I, Van Hauwermeiren F, Van Wonterghem E, Wilson C, Libert C. Pro-inflammatory effects of matrix metalloproteinase 7 in acute inflammation. Mucosal Immunology. 2014; 7: 579–588. https://doi.org/10.1038/mi.2013.76.

[39] Liao HY, Da CM, Liao B, Zhang HH. Roles of matrix metalloproteinase-7 (MMP-7) in cancer. Clinical Biochemistry. 2021; 92: 9–18. https://doi.org/10.1016/j.clinbiochem.2021.03.003.

[40] Shiomi T, Inoki I, Kataoka F, Ohtsuka T, Hashimoto G, Nemori R, *et al*. Pericellular activation of proMMP-7 (promatrilysin-1) through interaction with CD151. Laboratory Investigation; a Journal of Technical Methods and Pathology. 2005; 85: 1489–1506. https://doi.org/10.1038/labinvest.3700351.

[41] Krakowiak MS, Noto JM, Piazuelo MB, Hardbower DM, Romero-Gallo J, Delgado A, *et al*. Matrix metalloproteinase 7 restrains Helicobacter pylori-induced gastric inflammation and premalignant lesions in the stomach by altering macrophage polarization. Oncogene. 2015; 34: 1865–1871. https://doi.org/10.1038/onc.2014.135.

[42] Di Sano F, Fazi B, Tufi R, Nardacci R, Piacentini M. Reticulon-1C acts as a molecular switch between endoplasmic reticulum stress and genotoxic cell death pathway in human neuroblastoma cells. Journal of Neurochemistry. 2007; 102: 345–353. https://doi.org/10.1111/j.1471-4159.2007.04479.x.

[43] McGonigle TA, Dwyer AR, Greenland EL, Scott NM, Carter KW, Keane KN, *et al*. Reticulon-1 and Reduced Migration toward Chemoattractants by Macrophages Differentiated from the Bone Marrow of Ultraviolet-Irradiated and Ultraviolet-Chimeric Mice. Journal of Immunology (Baltimore, Md.: 1950). 2018; 200: 260–270. https://doi.org/10.4049/jimmunol.1700760.

[44] Zhu S, Zu L, Xu S. The Expression of RTN1 in Lung Adenocarcinoma and Its Effect on Immune Microenvironment. Zhongguo Fei Ai Za Zhi. 2022; 25: 385–395. (In Chinese)

[45] Liu Q, Yang T, Zhang Y, Hu ZD, Liu YM, Luo YL, *et al*. ZIC2 induces pro-tumor macrophage polarization in nasopharyngeal carcinoma by activating the JUNB/MCSF axis. Cell Death & Disease. 2023; 14: 455. https://doi.org/10.1038/s41419-023-05983-x.

[46] Wu Q, Tao X, Luo Y, Zheng S, Lin N, Xie X. A novel super-enhancer-related gene signature predicts prognosis and immune microenvironment for breast cancer. BMC Cancer. 2023; 23: 776. https://doi.org/10.1186/s12885-023-11241-2.

[47] Wei-Hua W, Ning Z, Qian C, Dao-Wen J. ZIC2 promotes cancer stem cell traits via up-regulating OCT4 expression in lung adenocarcinoma cells. Journal of Cancer. 2020; 11: 6070–6080. https://doi.org/10.7150/jca.44367.

[48] Hu X, Dong Y, Xie S, Song Y, Yu C, He Y, *et al*. Immune checkpoint CD161/LLT1-associated immunological landscape and diagnostic value in oral squamous cell carcinoma. The Journal of Pathology. Clinical Research. 2024; 10: e353. https://doi.org/10.1002/cjp2.353.

[49] Zeng C, Wang J, Zhao S, Wei Y, Qi Y, Liu S, *et al*. Multi-cohort validation of a lipid metabolism and ferroptosis-associated index for prognosis and immunotherapy response prediction in hormone receptor-positive breast cancer. International Journal of Biological Sciences. 2025; 21: 3968–3992. https://doi.org/10.7150/ijbs.113213.

[50] Kazanietz MG, Durando M, Cooke M. CXCL13 and Its Receptor CXCR5 in Cancer: Inflammation, Immune Response, and Beyond. Frontiers in Endocrinology. 2019; 10: 471. https://doi.org/10.3389/fendo.2019.00471.

[51] Beider K, Voevoda-Dimenshtein V, Zoabi A, Rosenberg E, Magen H, Ostrovsky O, *et al*. CXCL13 chemokine is a novel player in multiple myeloma osteolytic microenvironment, M2 macrophage polarization, and tumor progression. Journal of Hematology & Oncology. 2022; 15: 144. https://doi.org/10.1186/s13045-022-01366-5.

[52] Wang GZ, Cheng X, Zhou B, Wen ZS, Huang YC, Chen HB, *et al*. The chemokine CXCL13 in lung cancers associated with environmental polycyclic aromatic hydrocarbons pollution. eLife. 2015; 4: e09419. https://doi.org/10.7554/eLife.09419.