



# A Machine-Learning Approach to Developing a Predictive Signature Based on Transcriptome Profiling of Ground-Glass Opacities for Accurate Classification and Exploring the Immune Microenvironment of Early-Stage LUAD

## OPEN ACCESS

### Edited by:

Weiyi Fang,  
Southern Medical University, China

### Reviewed by:

Richeng Jiang,  
Tianjin Medical University, China  
Yongbin Chen,  
Key Laboratory of Animal Models and  
Human Disease Mechanisms,  
Kunming Institute of Zoology (CAS),  
China

### \*Correspondence:

Xiang Wang  
wangxiang@csu.edu.cn  
Muyun Peng  
pengmuyu@csu.edu.cn  
Fenglei Yu  
yufenglei@csu.edu.cn

†These authors share first authorship

### Specialty section:

This article was submitted to  
Cancer Immunity  
and Immunotherapy,  
a section of the journal  
Frontiers in Immunology

Received: 09 February 2022

Accepted: 27 April 2022

Published: 26 May 2022

### Citation:

Zhao Z, Yin W, Peng X, Cai Q,  
He B, Shi S, Peng W, Tu G, Li Y,  
Li D, Tao Y, Peng M, Wang X  
and Yu F (2022) A Machine-  
Learning Approach to Developing  
a Predictive Signature Based on  
Transcriptome Profiling of Ground-  
Glass Opacities for Accurate  
Classification and Exploring the  
Immune Microenvironment  
of Early-Stage LUAD.  
*Front. Immunol.* 13:872387.  
doi: 10.3389/fimmu.2022.872387

Zhenyu Zhao<sup>1,2†</sup>, Wei Yin<sup>1,2†</sup>, Xiong Peng<sup>1,2</sup>, Qidong Cai<sup>1,2</sup>, Boxue He<sup>1,2</sup>, Shuai Shi<sup>1,2</sup>,  
Weilin Peng<sup>1,2</sup>, Guangxu Tu<sup>1,2</sup>, Yunping Li<sup>3</sup>, Dateng Li<sup>4</sup>, Yongguang Tao<sup>1,2,5,6</sup>,  
Muyun Peng<sup>1,2\*</sup>, Xiang Wang<sup>1,2\*</sup> and Fenglei Yu<sup>1,2\*</sup>

<sup>1</sup> Department of Thoracic Surgery, The Second Xiangya Hospital of Central South University, Changsha, China, <sup>2</sup> Hunan Key Laboratory of Early Diagnosis and Precise Treatment of Lung Cancer, The Second Xiangya Hospital of Central South University, Changsha, China, <sup>3</sup> Department of Ophthalmology, The Second Xiangya Hospital of Central South University, Changsha, China, <sup>4</sup> White Plains, NY, United States, <sup>5</sup> Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education, Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, <sup>6</sup> National Health Council (NHC) Key Laboratory of Carcinogenesis (Central South University), Cancer Research Institute and School of Basic Medicine, Central South University, Changsha, China

Screening for early-stage lung cancer with low-dose computed tomography is recommended for high-risk populations; consequently, the incidence of pure ground-glass opacity (pGGO) is increasing. Ground-glass opacity (GGO) is considered the appearance of early lung cancer, and there remains an unmet clinical need to understand the pathology of small GGO (<1 cm in diameter). The objective of this study was to use the transcriptome profiling of pGGO specimens <1 cm in diameter to construct a pGGO-related gene risk signature to predict the prognosis of early-stage lung adenocarcinoma (LUAD) and explore the immune microenvironment of GGO. pGGO-related differentially expressed genes (DEGs) were screened to identify prognostic marker genes with two machine learning algorithms. A 15-gene risk signature was constructed from the DEGs that were shared between the algorithms. Risk scores were calculated using the regression coefficients for the pGGO-related DEGs. Patients with Stage I/II LUAD or Stage IA LUAD and high-risk scores had a worse prognosis than patients with low-risk scores. The prognosis of high-risk patients with Stage IA LUAD was almost identical to that of patients with Stage II LUAD, suggesting that treatment strategies for patients with Stage II LUAD may be beneficial in high-risk patients with Stage IA LUAD. pGGO-related DEGs were mainly enriched in immune-related pathways. Patients with high-risk scores and high tumor mutation burden had a worse prognosis and may benefit from immunotherapy. A nomogram was constructed to facilitate the clinical application of

the 15-gene risk signature. Receiver operating characteristic curves and decision curve analysis validated the predictive ability of the nomogram in patients with Stage I LUAD in the TCGA-LUAD cohort and GEO datasets.

**Keywords:** GGO (ground-glass opacity), LUAD, TCGA, GEO, prognosis

## INTRODUCTION

Screening for early-stage lung cancer with low-dose computed tomography (LDCT) is recommended for high-risk populations; consequently, the incidence of pulmonary ground-glass opacity (GGO) is increasing (1). On CT, GGO appears as hazy opacities that do not obscure underlying pulmonary vessels or bronchial structures (2). GGO can manifest as benign or malignant lesions, including inflammation, preinvasive lesions, or adenocarcinomas (1). Typically, early lung adenocarcinomas (LUADs) *in situ* appear as pure ground-glass opacities (pGGOs), while advanced adenocarcinoma may appear as mixed ground-glass opacities (mGGOs) (3). pGGO and mGGO have significantly different prognoses, and solid LUADs are associated with shorter overall survival (OS) and recurrence-free survival compared to lesions with a GGO component (4, 5).

Guidelines on the management of GGO have been published (6–8); however, differentiating malignant and benign GGOs and clinical decision-making on the need for and timing of surgical resection are controversial (6, 9, 10). Persistent GGOs may represent premalignant conditions. Surgery involving wedge resection or segmentectomy, with or without regional lymph node dissection, is the most effective therapy for these patients (5, 11, 12). Most patients with GGO have satisfactory 5-year OS after appropriate therapy (4, 5); however, GGO may grow or demonstrate malignancy in approximately 20% of patients with pGGO and 40% of patients with mGGO (10, 13). A better understanding of the natural history of GGO, improved technology for diagnosis and follow-up, and establishing a precise size threshold for intervention may advance the management of patients with GGO (6, 13–17).

There is an unmet clinical need to understand the pathology of small GGO (<10 mm in diameter) (18–20), the factors associated with GGO growth and progression (21), and how evolving technology, including next-generation sequencing (NGS) combined with clinicopathological information, can facilitate a more accurate diagnosis of early-stage lung cancer (22). The objective of this study was to use the transcriptome profiling of pGGO specimens <1 cm in diameter to 1) construct a 15-gene risk signature to predict the prognosis of early-stage LUAD and 2) explore the immune microenvironment of GGO. Findings may inform a new classification strategy for early-stage lung cancer and improve diagnosis, follow-up, and treatment strategies.

## METHODS

### Specimen Collection

All specimens were collected from patients undergoing surgery in the Second Xiangya Hospital of Central South University from

May 2020 to May 2021. Specimens were stored at  $-80^{\circ}\text{C}$  until analysis. Inclusion criteria were 1) pGGO < 1 cm in diameter detected with high-resolution CT (HRCT), 2) the patient underwent surgical resection and pathological analysis for clinical decision-making, and tumors were staged according to the American Joint Commission on Cancer (AJCC) 8th edition TNM staging system, and 3) postoperative pathological diagnosis confirmed LUAD. Finally, 30 paired samples of pGGO and adjacent normal tissue were sent to BGI Tech SOLUTIONS (Hongkong) for high-throughput transcriptome sequencing. The clinical characteristics of the patients with pGGO are summarized in **Table 1**.

### Data Standardization and Differential Gene Expression Between pGGO and Adjacent Normal Tissue

The high-throughput transcriptome sequencing dataset was normalized using the “edgeR” package in R (23). Differentially expressed genes (DEGs) between pGGO and adjacent normal tissue samples were obtained using the “Limma” package in R ( $|\log \text{FC}| > 1$ , FDR  $P < 0.05$ ) (24). pGGO-related DEGs were

**TABLE 1** | Clinical characteristics of patients.

#### Clinical characteristics of patients

<b>Patients (n = 30)</b>	
NSCLC patients	30 (50.0%)
Non-cancer controls	30 (50.0%)
<b>Genders (n = 30)</b>	
Male	8 (26.7%)
Female	22 (73.3%)
<b>Age (n = 30)</b>	
≤ 60	25 (83.3%)
> 60	5 (16.7%)
<b>Sampling methods (n = 30)</b>	
Bronchoscopy	24 (80.0%)
Lobectomy	6 (20.0%)
<b>Smoking status (n = 30)</b>	
Smoker	10 (33.3%)
Non-smoker	20 (66.7%)
<b>TNM stage (n = 30)</b>	
I–II	30 (100%)
III–IV	0 (0%)
<b>pathological type</b>	
Adenocarcinoma <i>in situ</i>	6 (20.0%)
Minimally invasive adenocarcinoma	7 (23.3%)
Poorly-differentiated adenocarcinoma	4 (13.3%)
Moderately-differentiated adenocarcinoma	5 (16.7%)
Well-differentiated adenocarcinoma	8 (26.7%)
<b>GGO type</b>	
Single pure GGO	20 (66.7%)
Multiple pure GGO	10 (33.3%)

exhibited in a heatmap and volcano plot, which were generated by the “pheatmap, ggrepel, dplyr” package in R (25). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted to identify the function of the DEGs.

## Non-Negative Matrix Factorization to Identify Molecular Subtypes

Next, we extracted the expression data of the pGGO-related DEGs in TCGA-LUAD stage I-II datasets. We also performed the non-negative matrix factorization (NMF) method to cluster the LUAD stage I-II patients. NMF is an unsupervised learning technique for dimension reduction that decomposes a large measurement matrix into two low-rank non-negative matrices (26). The cophenetic correlation coefficient, based on the consensus matrix and proposed by Brunet et al. (2004), is used to measure the stability of clusters (27). NMF can classify samples better than consensus clustering. mRNA expression profiles and the clinical data for 497 patients with LUAD (Stages I-II, n=347; Stages III-IV, n=121) were downloaded from The Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>); follow-up information was available for each patient. The ID of the pGGO-related DEGs was used to extract expression data from the TCGA-LUAD cohort. The expression of the pGGO-related DEGs was verified in the TCGA-LUAD cohort. NMF was used to cluster patients with Stage I-II LUAD in the TCGA cohort. The clustering effect was evaluated with progression-free survival (PFS) and overall survival (OS).

## Identification of pGGO-Related DEGs and Differences in the Tumor Microenvironment Between the NMF Subgroups

Differences in pGGO-related DEGs between NMF subgroups were obtained using the “Limma” package in R. DEGs that were differently expressed between pGGOs and normal adjacent tissues, as well as differentially expressed between NMF subgroups, were identified. GO and KEGG pathway enrichment analyses were conducted to identify the function of the DEGs (28). The immune cell content of each NMF subgroup was analyzed using the “MCPcounter” package in R (29). HLA expression was compared between NMF subgroups. Based on a previous study, six immune subtypes were defined according to immune infiltrates. The relationship between the six immune subtypes and NMF subgroups was explored (30).

## Identification of pGGO-Related DEGs With Prognostic Value Using LASSO Cox Regression and Support Vector Machine—Recursive Feature Elimination

pGGO-related DEGs were screened with LASSO cox regression and support vector machine—recursive feature elimination (SVM-RFE) to identify prognostic marker genes (31). Feature selection for multiclass classification problems is challenging in machine learning. Existing multi-class gene selection algorithms are often not Pareto optimal. In this study, two machine learning

algorithms, LASSO cox regression, and SVM-RFE, were used to achieve Pareto optimality (32, 33). LASSO cox regression is used for data dimensionality reduction and feature selection. The regression coefficient is penalized by L1, some coefficients are shrunk to zero, features with non-zero regression coefficients are selected, and 10-fold cross-validation is used to evaluate the prediction model (34). SVM-RFE is often used for gene selection. SVM-RFE ranks features from most important to least, and least important features are iteratively eliminated (35, 36). Venn analysis was used to identify 15 pGGO-related DEGs that were shared between the LASSO cox regression and SVM-RFE machine learning algorithms.

## Validation of the pGGO-Related DEGs by the Quantitative Real-Time Reverse Transcription-Polymerase Chain Reaction

Quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR) was used to validate the expression of the pGGO-related DEGs with a prognostic value. RNA was extracted from 24 pGGO samples, and qRT-PCR was performed for 7 pGGO-related DEGs. Primer sequences were designed from Primer3web (<https://primer3.ut.ee>) (Table 2). qRT-PCR was performed with a SYBR Green super-mix reagent, and  $\beta$ -actin was the internal reference gene. The relative change in gene expression was calculated using the  $2^{-\Delta\Delta Ct}$  method. Results are presented as the mean of 3 replicates.

## Establishment and Validation of the pGGO-Related DEG Signature

Multivariate Cox regression was used to calculate regression coefficients for the pGGO-related DEGs with prognostic values. A prognostic signature was constructed using the following formula: risk score = coefficient(gene1) \* exp(gene1) + ... + coefficient (gene n) \* exp (gene n). A deviation plot was constructed to show the expression profile of the 15 pGGO-related DEGs with prognostic values. Patients in the TCGA-LUAD cohort were stratified into a high-risk or low-risk group using the median risk score as a cut-off. The 15-gene risk signature was verified in the GSE50081 and GSE72094 datasets, which were downloaded from the Gene Expression

**TABLE 2** | The primer sequence of the DEGs in risk signature.

Primer	Primer sequence (5'to 3')
PCP2-F	GAGAAGACGGAGGAAGGCTC
PCP2-R	CTCTGGCTCTTGGTGGTCTG
DKK1-F	CCATTGACAACACCAGCCG
DKK1-R	TTTTGCAGTAATTCCCGGGG
KCNV1-F	CGGGAATTCTTGTCTTGGCC
KCNV1-R	CTCCATGATACTCCGGGCAT
FAIM2-F	AGCTTCCAGACCAAGTTCTGA
FAIM2-R	TGTAATAACACCCGCTCCCA
FGF5-F	AGTGGTATGTGGCCCTGAAT
FGF5-R	TGGCTTGATAGGGCTAGGTG
NPAS1-F	CTTGTGAGAGCAGAGTCAGC
NPAS1-R	CTGCAGCCAACGGTAGTAAC
LINC00563-F	ATCTGGGATCATCTGGGTGG
LINC00563-R	CTTCTGCATTCCCTTCGCTC

Omnibus–NCBI (<https://www.ncbi.nlm.nih.gov/geo/>). Kaplan–Meier survival curves, decision curve analysis (DCA), risk plots, and time-dependent receiver operating characteristic (ROC) curves were used to investigate the predictive power of the prognostic signature in the TCGA-LUAD cohort and GEO datasets (37, 38). The relationship between clinicopathologic factors, immune scores, clusters, and risk scores was examined using Pearson’s chi-squared test (39).

## Clinical Application of the pGGO-Related DEG Signature

Univariate and multivariable Cox proportional hazard models were used to analyze independent associations between clinical outcomes. A nomogram based on clinicopathologic factors and risk scores was constructed. A calibration curve displayed the nomogram’s predictive power. Gene set enrichment analyses (GSEAs) in low-risk and high-risk groups were performed using the “clusterProfiler, enrichplot, DOSE, org.Hs.eg.db” package in R (40). Enrichment scores (ESs) that represent the degree to which a gene set is overrepresented at the top or bottom of a ranked list were calculated (<https://www.gsea-msigdb.org/gsea/index.jsp>). Tumor mutation burden (TMB) data were retrieved for the TCGA-LUAD cohort. The relationship between TMB and the 15-gene risk signature was evaluated using the “reshape2” package in R. The immunophenoscore (IPS) in patients with LUAD was downloaded from the Cancer Immunome Database (TCIA) (<https://tcia.at/home>) (41). The relationship between IPS and the 15-gene risk signature was evaluated with Pearson correlation analysis.

## Statistical Validation

All statistical analyses were conducted with R software (Version 4.0.1).

NMF algorithms were performed using the “NMF, survival” package (42). Kaplan–Meier survival curves were plotted using the “survminer” package. SVM-RFE was performed using the “e1071, kernlab, caret” package (43). The nomogram was constructed using the “rms” package. The predictive ability of the pGGO-related DEGs signature was evaluated using the “SurvivalROC” package. The deviation plot was constructed using the “ggpubr” package. Correlation analysis was performed using the Pearson method. Differences between subgroups were evaluated using the Wilcoxon rank-sum test. mRNA profiles from the GEO datasets were normalized using the “sva, Limma” package. All figures were plotted using the “ggplot” package. For the GO and KEGG analyses, a P-value <0.05 and FDR q-value <0.25 were considered statistically significant.

## RESULTS

### Identification of DEGs Between pGGO and Adjacent Normal Tissue and NMF Clustering

A total of 1,734 DEGs ( $|\log FC| > 2$ ,  $P < 0.05$ ) between pGGO and adjacent normal tissue samples were identified; of these, 648 DEGs

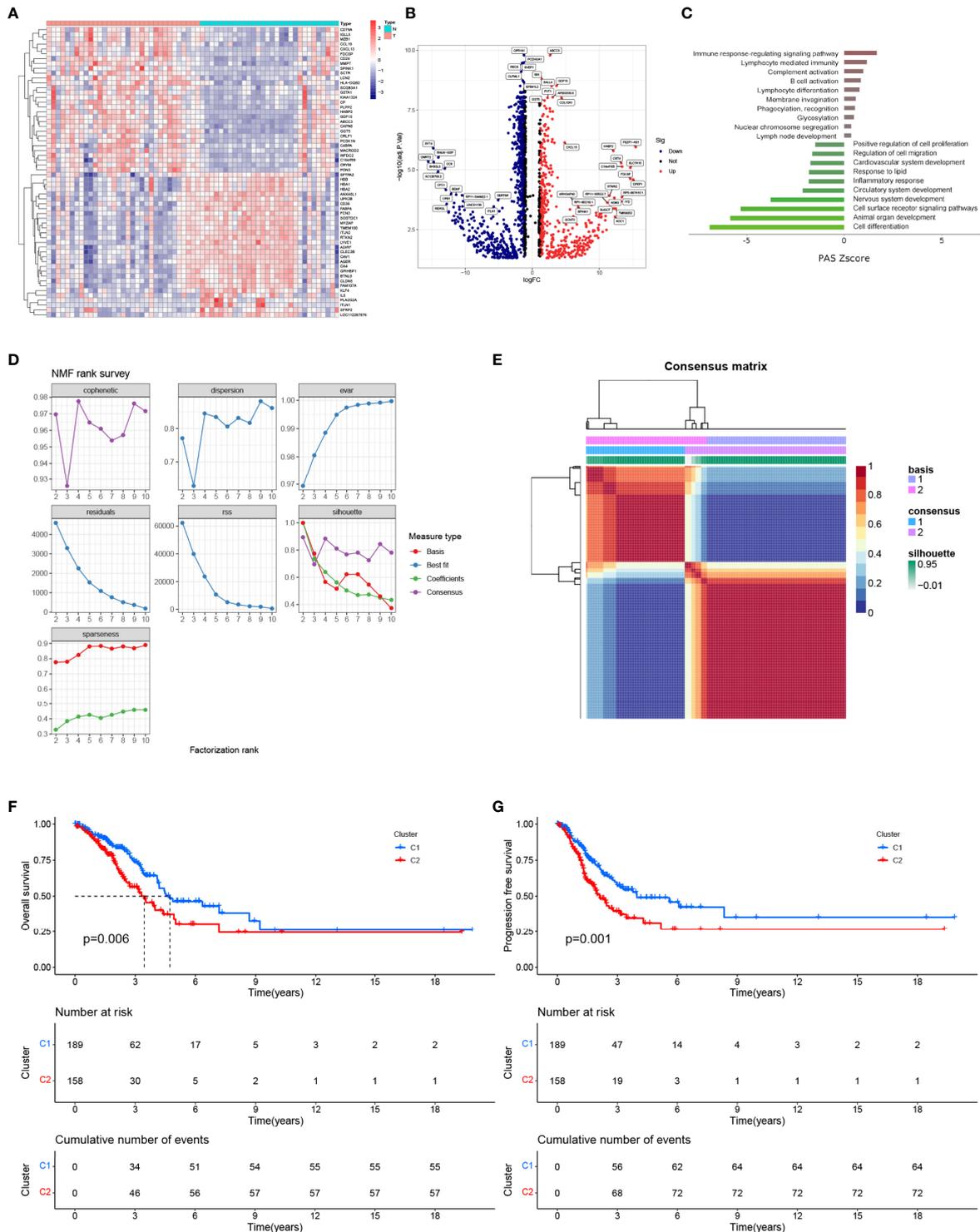
were upregulated, and 1,086 DEGs were downregulated (Figures 1A, B and Table S1). GO and KEGG pathway enrichment analyses suggested that the DEGs were mainly involved in immune-related pathways, such as immune response-regulating signaling pathways and lymphocyte-mediated immunity pathways (Figure 1C). Next, we extract the mRNA expression data of the pGGO-related DEGs in TCGA-LUAD stage IA datasets and perform the NMF consensus cluster analysis. The best cluster number was chosen as the coexistence correlation coefficient  $K$  value = 2 (Figure 1D); therefore, patients with Stage I–II LUAD in the TCGA cohort ( $n = 347$ ) were divided into two clusters (Figure 1E and Table S2). Patients in Cluster 1 had better PFS and OS compared to patients in Cluster 2 (Figures 1F, G).

### Identification of pGGO-Related DEGs and Differences in the Tumor Microenvironment Between the NMF Subgroups

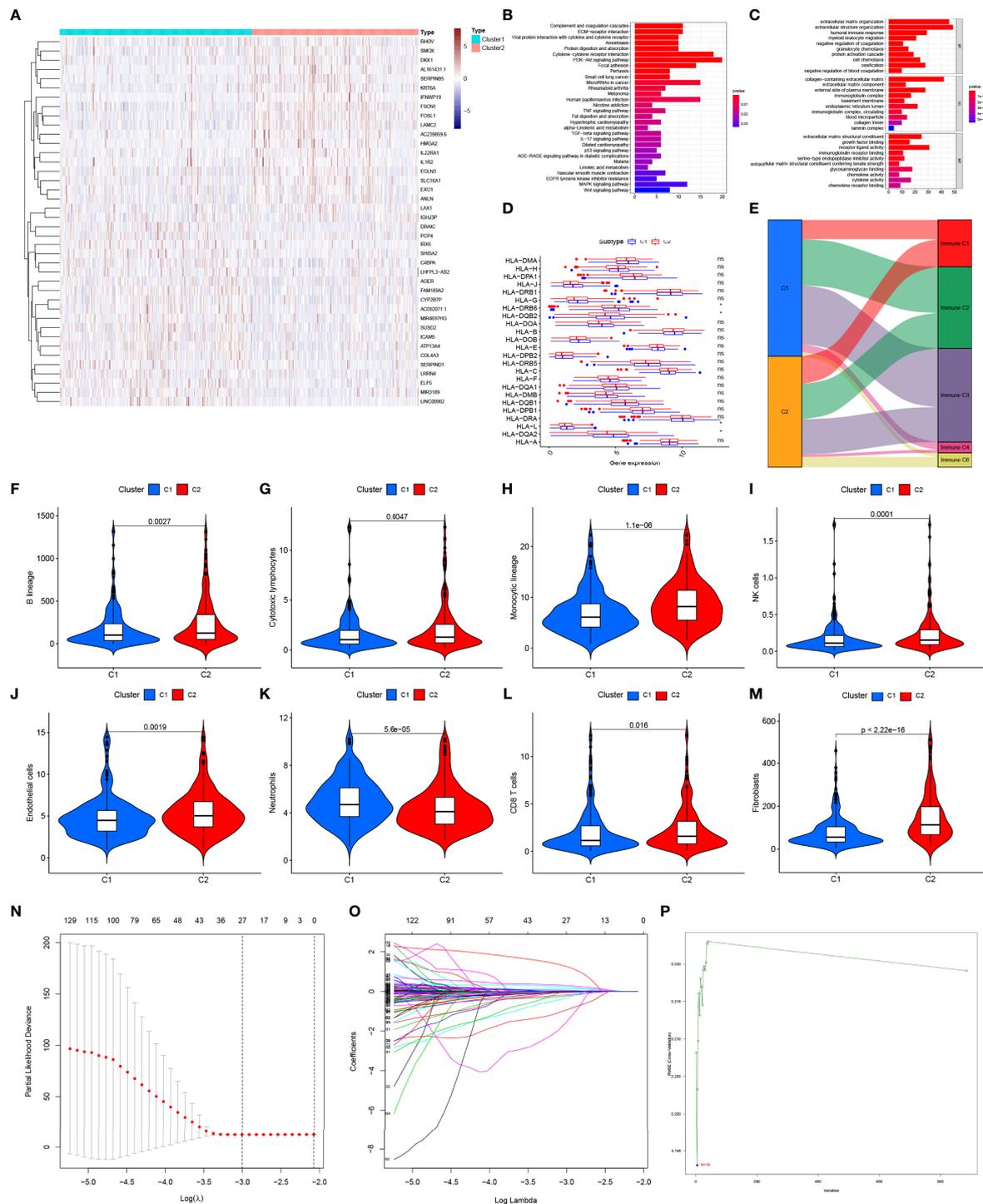
A total of 208 pGGO-related DEGs between Cluster 1 and Cluster 2 were identified; of these, 33 pGGO-related DEGs were upregulated, and 175 pGGO-related DEGs were downregulated (Figure 2A and Table S3). KEGG pathway enrichment analysis suggested that the pGGO-related DEGs were mainly enriched in immune-related pathways, including the cytokine–cytokine receptor interaction pathway and IL-17 signaling pathway (Figure 2B). GO function analysis suggested that the pGGO-related DEGs were mainly enriched in immune-related processes, including the humoral immune response process and immunoglobulin receptor-binding process (Figure 2C). HLA gene expression data showed that HLA-L, HLA-DQA2, HLA-DQB2, and HLA-DRB6 were highly expressed in Cluster 1 (Figure 2D). A Sankey diagram showed a relationship between the cluster subtype and six immune subtypes defined in a previous study (Figure 2E and Table S4). Violin plots suggested that Cluster 2 was characterized by a high expression of B-cell lineages, endothelial cells, cytotoxic lymphocytes, CD8 T cells, monocyte-lineage cells, fibroblasts, and NK cells; meanwhile, Cluster 1 was characterized by a high expression of neutrophils (Figures 2F–M).

Verification of pGGO-Specific DEGs Using Machine Learning Algorithms.

Lasso cox regression and SVM-RFE identified the 23 (Figures 2N, O and Table S5) and 19 (Figure 2P and Table S5) most representative prognostic pGGO-related DEGs, respectively, from among the 208 pGGO-related DEGs between Cluster 1 and Cluster 2. Venn analysis was used to identify 15 DEGs that were shared between the machine learning algorithms (Figure 3A). These included *DKK1*, *NPAS1*, *AL357143.1*, *KCNV1*, *AC068228.1*, *AC239859.6*, and *FGF5* which were highly expressed in pGGO samples, and *AC087763.1*, *PCP2*, *FAIM2*, *AL357143.1*, *AC022148.1*, *AC021678.2*, *LSP1P2*, and *LINC00563* that were highly expressed in adjacent normal tissue samples (Figure 3B). This expression profile was validated by qRT-PCR (Figure S1). Regression coefficients were used to identify a 15-gene risk signature. A forest plot showed that *FGF5*, *AC239859.6*,



**FIGURE 1** | Identification of DEGs between pGGO and adjacent normal tissue and NMF clustering. **(A)** Heatmap of the DEGs ( $|\log FC| > 1$ ,  $FDR P < 0.05$ ); **(B)** Volcano map of the DEGs ( $|\log FC| > 1$ ,  $FDR P < 0.05$ ); **(C)** KEGG pathway enrichment analysis of the DEGs; **(D, E)** NMF clustering of patients with Stage I–II LUAD in the TCGA-LUAD cohort (**D** shows the NMF rank survey and **E** shows a heatmap of the consensus matrix; the best cluster number was chosen as the coexistence correlation coefficient  $K$  value = 2); **(F, G)** Kaplan–Meier survival curves for the NMF subgroups (**F** shows OS and **G** shows PFS). DEGs, differentially expressed genes; GGO, ground-glass opacity; KEGG: Kyoto Encyclopedia of Genes and Genomes; NMF, non-negative matrix factorization; LUAD, lung adenocarcinoma; PFS, progression-free survival; OS, overall survival.



**FIGURE 2** | Identification of pGGO-related DEGs between Cluster 1 and Cluster 2 **(A)** Heatmap of the pGGO-related DEGs ( $|\log FC| > 1$ , FDR  $P < 0.05$ ); **(B, C)** KEGG **(B)** and GO **(C)** enrichment pathways analysis of the pGGO-related DEGs; **(D)** HLA gene expression data for Cluster 1 and Cluster 2; **(E)** Sankey diagram showing the relationship between the cluster subtype and six immune subtypes defined in a previous study; **(F–M)** Violin plots showing the expression of the immune cells in Cluster 1 and Cluster 2; **(N–P)** Selection of prognostic pGGO-DEGs in Cluster 1 and Cluster 2 using Lasso cox regression **(N, O)** and SVM-RFE (threshold value = 19) **(P)**. DEGs, differentially expressed genes; GGO, ground-glass opacity; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; NMF, non-negative matrix factorization; SVM-RFE: support vector machine—recursive feature elimination. ns represent non significant, \* represent  $P \leq 0.05$ .

*NPASI*, *KCNV1*, *AC068228.1*, and *AL353746.1* were risk factors in LUAD; while *DKK1*, *LINC00563*, *AC021678.2*, *AL357143.1*, *AC022148.1*, *LSP1P2*, *AC087763.1*, *PCP2*, and *FAIM2* were protective factors in LUAD (Figure 3C).

The TCGA-LUAD cohort was stratified into a high-risk group and low-risk group ( $n=234$ ) based on their median risk score (Table S6). Kaplan–Meier survival analysis showed that patients in the high-risk group had worse OS than patients in the low-risk group ( $P<0.001$ ) (Figure 3D). The 15-gene risk signature was predictive of patients with Stage IA LUAD in the TCGA-LUAD cohort. This was expected as pGGO is recognized as a component of TNM Stage IA LUAD according to the AJCC 8th edition TNM staging system. Patients with Stage IA LUAD and high-risk scores in the TCGA-LUAD cohort had worse OS than patients with Stage IA LUAD and low-risk scores ( $P<0.001$ ) (Figure 3E). There was no significant difference in OS between patients with Stage IA LUAD and high-risk scores and patients with Stage II LUAD (Figure 3F). To confirm the feasibility of the 15-gene risk signature, information for patients with LUAD from the GSE50081 and GSE72094 datasets was downloaded, combined, and stratified into a high-risk group ( $n=263$ ) and low-risk group ( $n=262$ ) based on the median risk score from the TCGA-LUAD cohort (Table S7). All patients, and patients with Stage IA LUAD, in the high-risk group had worse OS than patients in the low-risk group (Figures 3G, H); there was no significant difference in OS between patients with Stage IA LUAD and high-risk scores and patients with Stage II LUAD (Figure 3I). Time-dependent ROC curve analysis at 1, 3, and 5 years showed that the 15-gene risk signature had better predictive performance than other clinical traits in the TCGA-LUAD cohort (1 year:  $AUC=0.811$ ; 3 years:  $AUC=0.779$ ; 5 years:  $AUC=0.780$ ; Figures 3J–L). DCA showed that the 15-gene risk signature had more clinical benefits than other clinical traits (Figures 3M–O).

## Clinical Application of the 15-Gene Risk Signature

A heatmap showed that the TNM stage, immune scores, NMF subgroup, gender, T stage, and N stage were significantly associated with risk scores in TCGA datasets (Figure 4A,  $P<0.05$ ). Most patients with high-risk scores were men (Figure 4A,  $P<0.001$ ) and had higher-grade tumors (Figure 4B). Among the 15 pGGO-related DEGs in the gene risk signature, *DKK1*, *NPASI*, *AL357143.1*, *KCNV1*, *AC068228.1*, *AC239859.6*, and *FGF5* were highly expressed in the high-risk group, and *AC087763.1*, *PCP2*, *FAIM2*, *AL357143.1*, *AC022148.1*, *AC021678.2*, *LSP1P2*, and *LINC00563* were highly expressed in the low-risk group (Figure 4C). Risk curves based on a per-sample risk score also validated the predictive power of the 15-gene risk signature (Figures 4D, E). Univariate and multivariate COX regression analyses showed that the 15-gene risk signature is an independent prognostic factor in patients with Stage I–II LUAD (Figures 4F, G: univariate Cox regression analyses:  $P<0.001$ ,  $HR = 1.041$ ; 95% CI: 1.029–1.054; multivariate Cox regression analyses:  $P<0.001$ ,  $HR = 1.038$ ; 95% CI: 1.024–1.052).

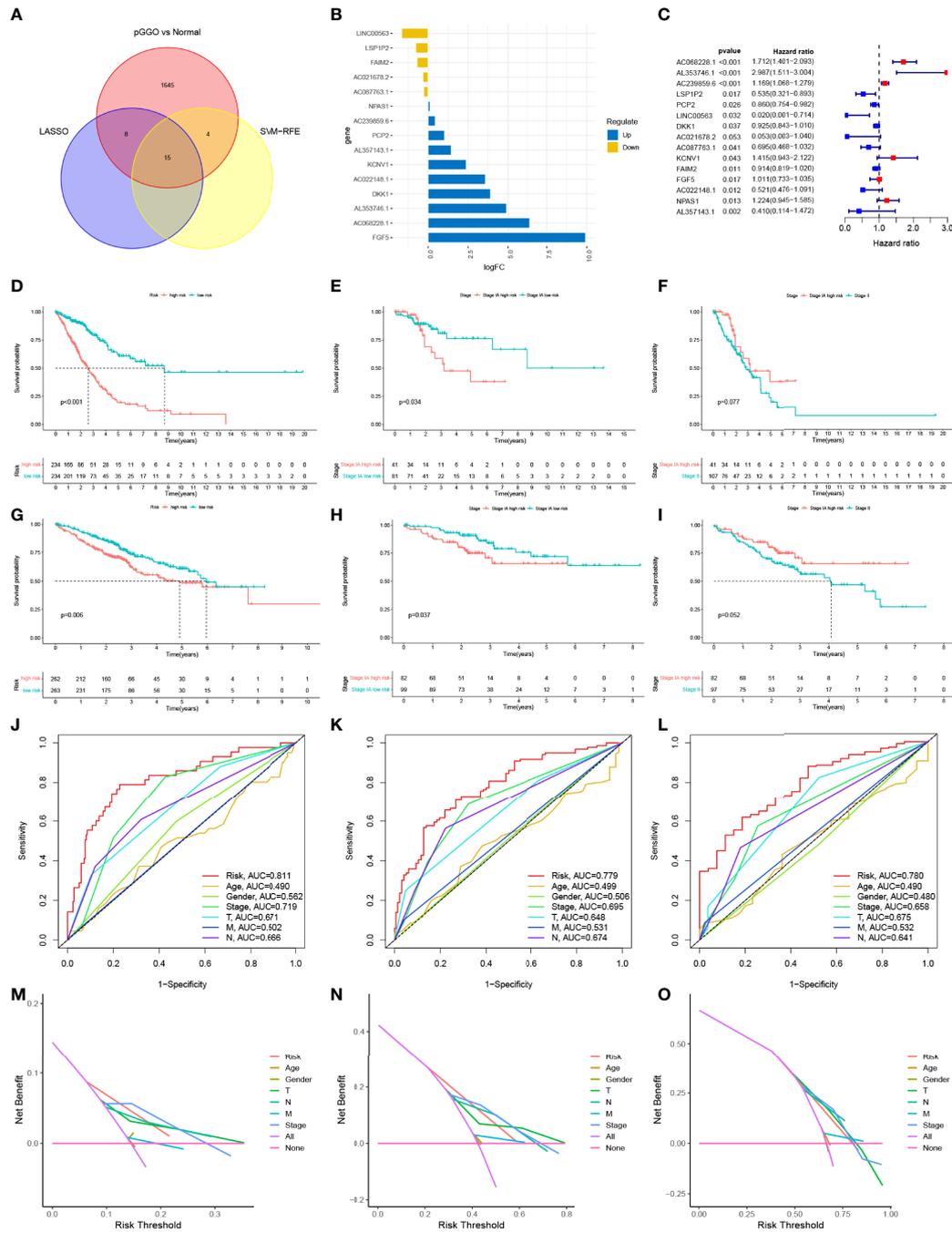
A nomogram incorporating clinical factors and the 15-gene risk signature was constructed for visualization and convenient clinical application (Figure 5A). Calibration curves validated the ability of the 15-gene risk signature to predict OS in the TCGA-LUAD cohort and LUAD dataset obtained from the GEO (Figures 5B, C). The ROC curve and DCA analysis validated that the nomogram had better predictive performance than the 15-gene risk signature alone at 1, 3, and 5 years in patients with Stage I LUAD from the TCGA-LUAD cohort (Figures 5D–I) and GEO (Figures 5J–O).

## The Relationship Between the 15-Gene Risk Signature, TMB, and the IPS

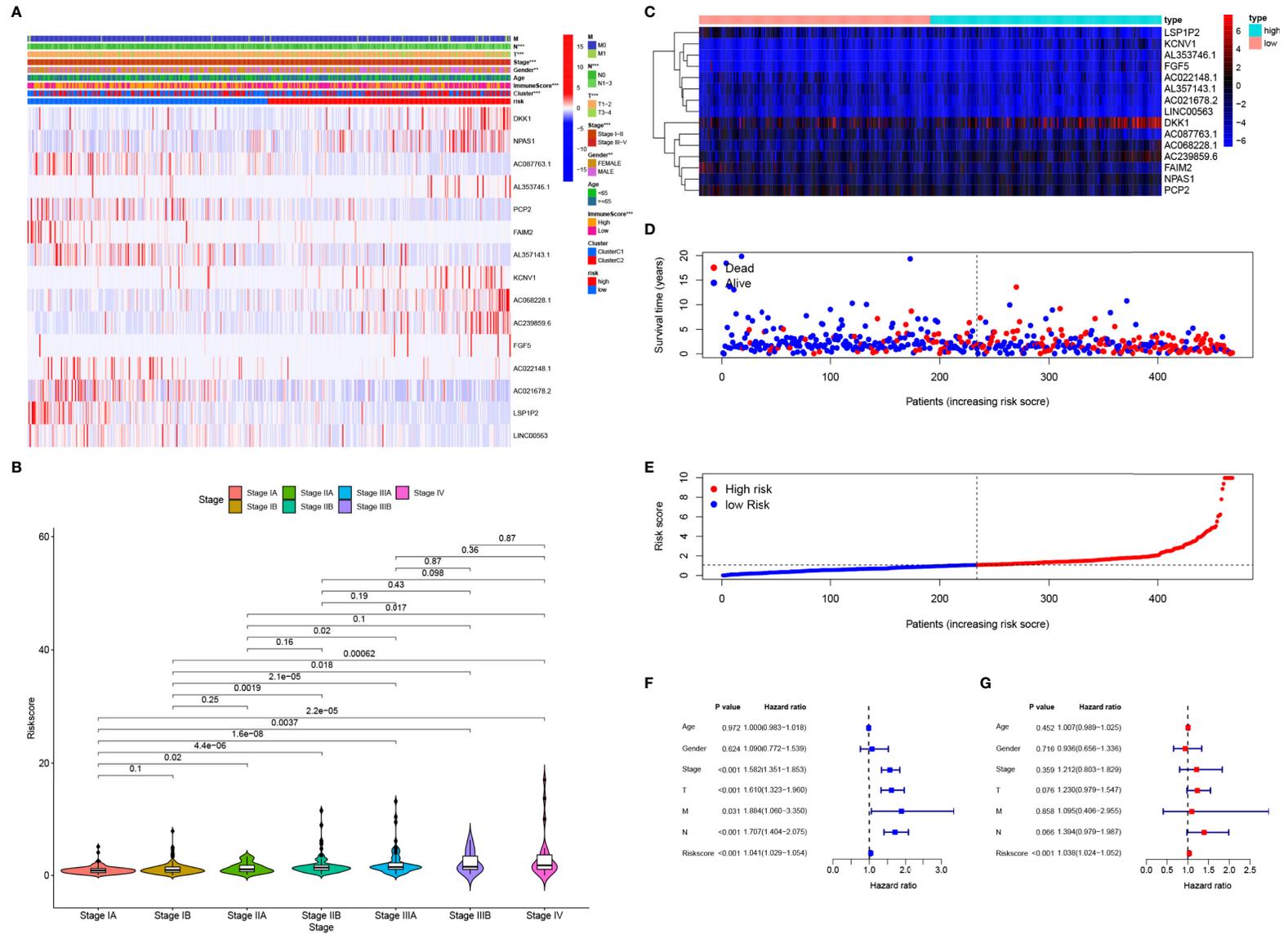
In the TCGA cohort, the GSEA suggested that the cell cycle, DNA replication, Parkinson's disease, pyrimidine metabolism, and ribosome pathways were activated in patients with high-risk scores; meanwhile, allograft rejection, asthma, the intestinal immune network for IgA production, primary immunodeficiency, and systemic lupus erythematosus pathways were activated in patients with low-risk scores (Figures 6A, B). High-risk patients had higher TMB than low-risk patients (Figure 6C and Table S8). Patients were stratified into a high-TMB group and a low-TMB group based on the median TMB value. There was no significant difference in OS between patients in the high-TMB group and low-TMB group; however, patients with a high-risk score and a high TMB had the worse OS (Figure 6D). Patients with a low-risk score and high TMB had the best OS (Figure 6E). Patients with high-risk scores always had low levels of immune infiltration (Figure 6F). Patients with a low-risk score and tumors that were CTLA4 positive and PD1 negative had a high IPS, suggesting that these patients may benefit from immunotherapy (Figure 6G, Table S9).

## DISCUSSION

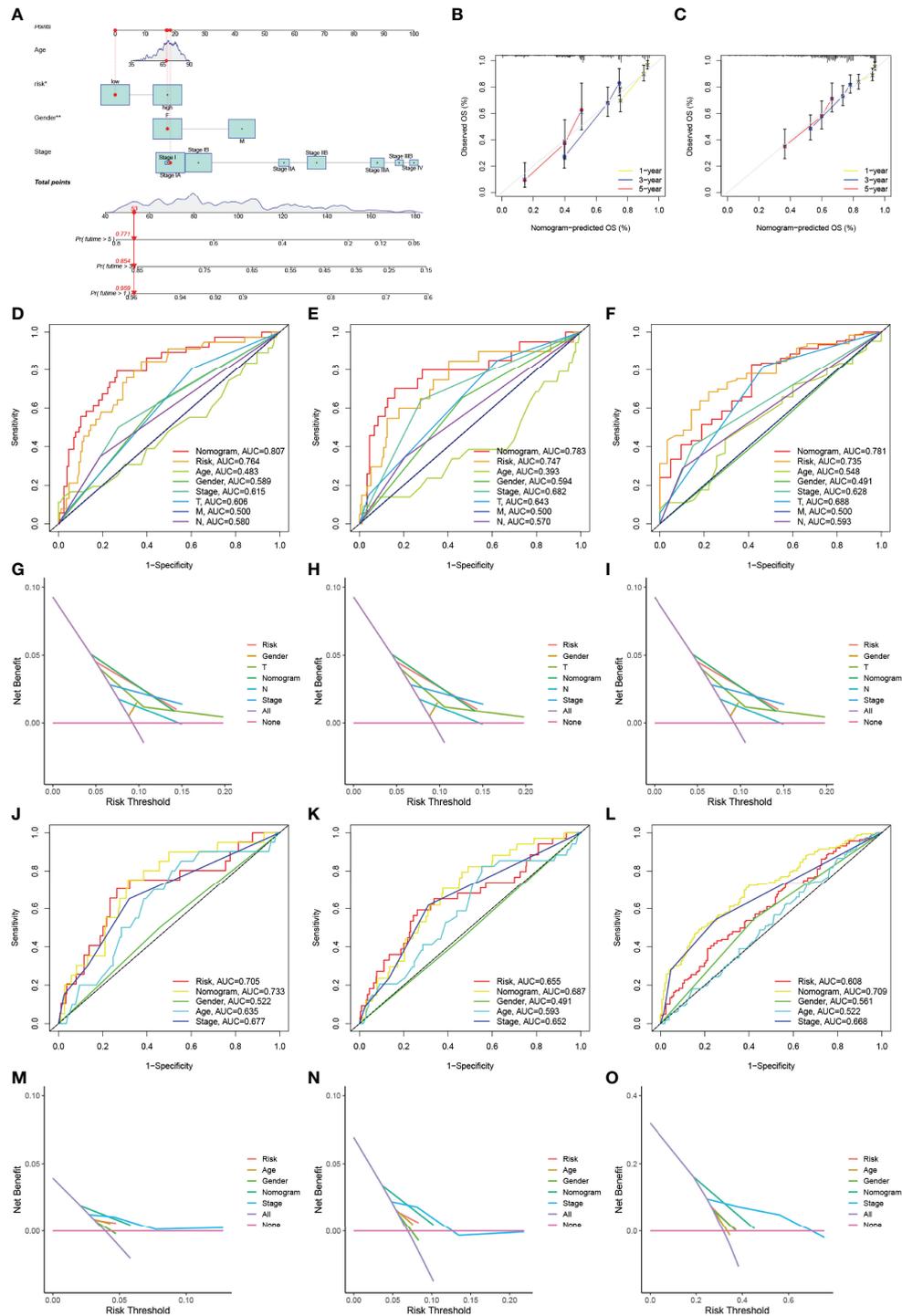
The understanding of the pathogenesis of early-stage LUAD has increased with the advent of high-throughput transcriptome sequencing technology; however, the knowledge of the etiology and natural progression of pGGO is limited, especially for pGGO  $<1$  cm in diameter (44). In this study, we identified DEGs between pGGO  $<1$  cm in diameter and the adjacent normal tissue. Functional analysis of the DEGs suggested that the immune microenvironment plays an important role in LUAD tumorigenesis. NMF identified the 2 subgroups (Cluster 1 and Cluster 2) of patients with Stage I–II LUAD in the TCGA-LUAD cohort. The pGGO-specific DEGs between Cluster 1 and Cluster 2 mainly participated in immune-related pathways. Immune infiltrates in Cluster 2 were characterized by a high expression of B-cell lineages, endothelial cells, cytotoxic lymphocytes, CD8 T cells, monocyte-lineage cells, fibroblasts, and NK cells, while Cluster 1 was characterized by a high expression of neutrophils. Survival analysis showed that patients in Cluster 1 had a better prognosis than patients in Cluster 2. Taken together, these data suggest that a higher level of immune infiltrates may indicate a poor prognosis in patients with Stage I–II LUAD.



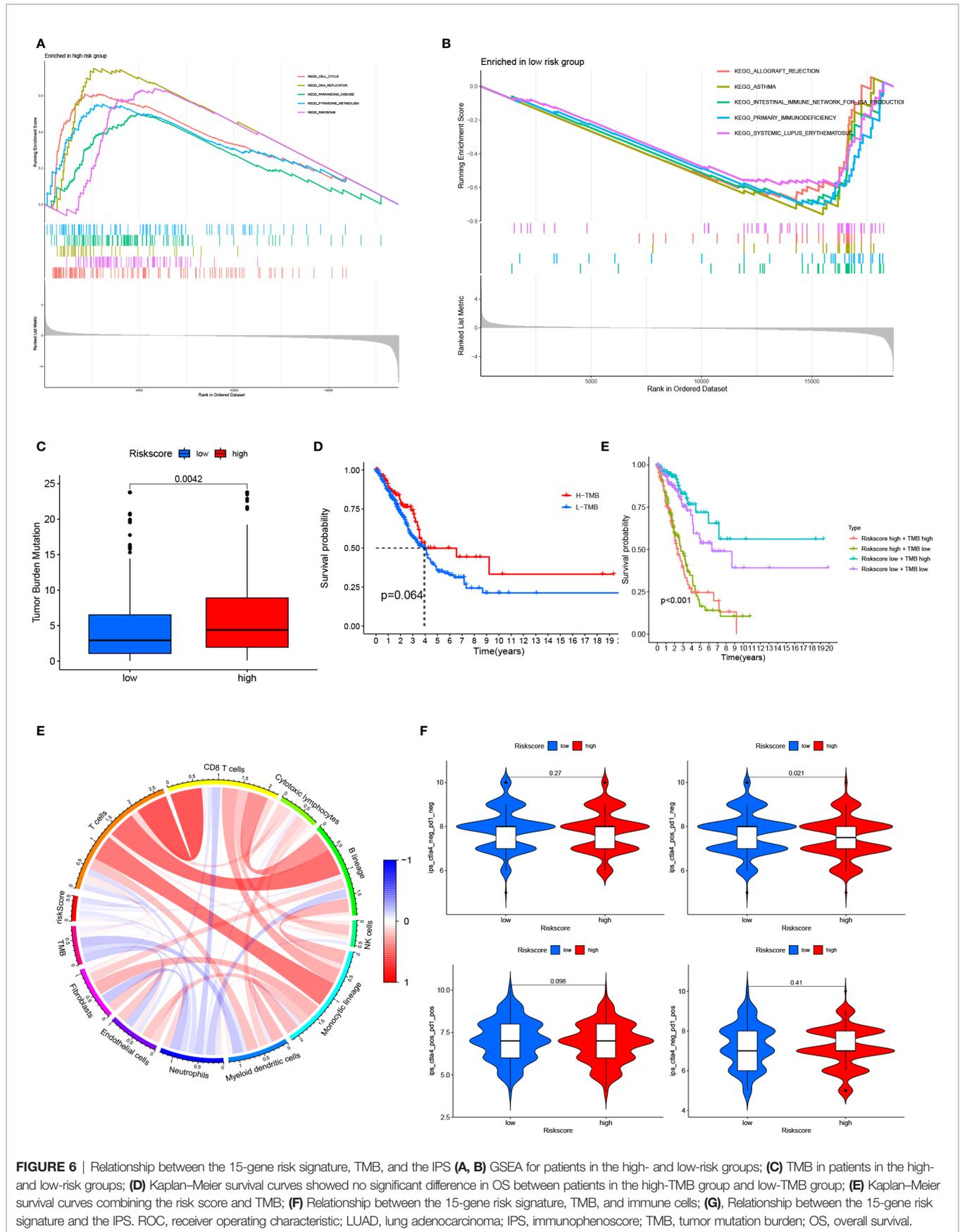
**FIGURE 3** | The construction and verification of the risk signature. **(A)** Venn analysis was used to identify 15 DEGs that were shared between the machine learning algorithms (Lasso cox regression and SVM-RFE); **(B)** Expression data of the 15 prognostic pGGO-related DEGs; **(C)** Forest plot of the 15 prognostic pGGO-related DEGs (red: risk factors; blue: protective factors); **(D–I)** Kaplan–Meier survival curves. The TCGA-LUAD cohort was stratified into a high-risk group and low-risk group. Patients in the high-risk group had a worse prognosis than the patients in the low-risk group in the overall TCGA-LUAD cohort **(D)** and patients with Stage IA LUAD **(E)**. There was no significant difference in OS between patients with Stage IA LUAD and high-risk scores and patients with Stage II LUAD **(F)**. Patients with LUAD from the GSE50081 and GSE72094 datasets were stratified into a high-risk group and low-risk group. Patients in the high-risk group had a worse prognosis than the patients in the low-risk group in the overall GEO-LUAD dataset **(G)** and patients with Stage IA LUAD **(H)**. There was no significant difference in OS between patients with Stage IA LUAD and high-risk scores and patients with Stage II LUAD **(I)**. **(J–L)** R Time-dependent ROC curve analysis at 1 **(J)**, 3 **(K)**, and 5 years **(L)** verified the predictive performance of the 15-gene risk signature in the TCGA-LUAD cohort. **(M–O)** DCA at 1 **(M)**, 3 **(N)**, and 5 years **(O)** verified the predictive performance of the 15-gene risk signature in the TCGA-LUAD cohort. DEGs, differentially expressed genes; GGO, ground-glass opacity; ROC, receiver operating characteristic; LUAD, lung adenocarcinoma; SVM-RFE, support vector machine–recursive feature elimination; DCA, decision curve analysis.



**FIGURE 4** | Clinical application of the 15-gene risk signature. **(A)** Heatmap showing that the TNM stage, immune scores, NMF subgroup, gender, T stage, and N stage were significantly associated with risk scores, and the expression levels of the 15 DEGs were different between the high- and low-risk groups. **(B)** Violin plots showing the relationship between the risk score and TNM stage. **(C–E)** Heatmap of the mRNA expression of the risk signature **(C)** and risk curves in the TCGA-LUAD cohort **(D, E)**. **(F, G)** Univariate **(F)** and multivariate **(G)** Cox regression analyses suggested that the 15-gene risk signature was the independent prognostic factor in the TCGA-LUAD cohort. ROC, receiver operating characteristic curve; LUAD, lung adenocarcinoma; DCA, decision curve analysis. \* represent  $P \leq 0.05$ , \*\* represent  $P \leq 0.01$ , \*\*\* represent  $P \leq 0.001$ , and \*\*\*\* represent  $P \leq 0.0001$ .



**FIGURE 5** | Construction and verification of the nomogram. **(A)** The nomogram was constructed using the TNM stage, risk signature, T stage, N stage, and gender. **(B, C)** Calibration curves verifying the performance of the nomogram at 1, 3, and 5 years in the TCGA-LUAD cohort **(B)** and GEO datasets **(C)**. **(D–F)** ROC curve analysis verifying the performance of the nomogram at 1 **(D)**, 3 **(E)**, and 5 years **(F)** in patients with Stage I LUAD in the TCGA cohort. **(D–F)** DCA verifying the performance of the nomogram at 1 **(G)**, 3 **(H)**, and 5 years **(I)** in patients with Stage I LUAD in the TCGA cohort; **(J–L)** ROC curve analysis verifying the performance of the nomogram at 1 **(J)**, 3 **(K)**, and 5 years **(L)** in patients with Stage I LUAD in the GEO datasets. **(M–O)** DCA verifying the performance of the nomogram at 1 **(M)**, 3 **(N)**, and 5 years **(O)** in patients with Stage I LUAD in the GEO datasets. ROC, receiver operating characteristic curve; LUAD, lung adenocarcinoma; DCA, decision curve analysis. \* represent  $P \leq 0.05$ , \*\* represent  $P \leq 0.01$ , \*\*\* represent  $P \leq 0.001$ , and \*\*\*\* represent  $P \leq 0.0001$ .



**FIGURE 6** | Relationship between the 15-gene risk signature, TMB, and the IPS (**A, B**) GSEA for patients in the high- and low-risk groups; (**C**) TMB in patients in the high- and low-risk groups; (**D**) Kaplan–Meier survival curves showed no significant difference in OS between patients in the high-TMB group and low-TMB group; (**E**) Kaplan–Meier survival curves combining the risk score and TMB; (**F**) Relationship between the 15-gene risk signature, TMB, and immune cells; (**G**), Relationship between the 15-gene risk signature and the IPS. ROC, receiver operating characteristic; LUAD, lung adenocarcinoma; IPS, immunophenoscore; TMB, tumor mutation burden; OS, overall survival.

pGGO-related DEGs were screened to identify prognostic marker genes with two machine learning algorithms. A 15-gene risk signature was constructed from the DEGs that were shared between the algorithms. Risk scores were calculated using the regression coefficients for those pGGO-related DEGs. As pGGO is recognized as a component of TNM Stage IA LUAD according to the AJCC 8th edition TNM staging system, we evaluated the predictive ability of the 15-gene risk signature in patients with Stage IA LUAD. Patients with Stage IA LUAD and high-risk scores had worse OS than patients with Stage IA LUAD and low-risk scores, but the prognosis of high-risk patients with Stage IA LUAD was almost identical to that of patients with Stage II LUAD. These data suggest that treatment strategies for patients with Stage II LUAD may be beneficial in high-risk patients with Stage IA LUAD. The clinical application of the 15-gene risk signature was verified in two GEO datasets (GSE50081 and GSE72094). Findings showed no significant difference in OS between patients with Stage IA LUAD and high-risk scores and patients with Stage II LUAD. ROC curve analysis and DCA confirmed the risk signature's ability to predict Stage I LUAD in the TCGA cohort. An enhanced nomogram was constructed to facilitate the clinical application of the risk signature. The predictive ability of the nomogram was verified in patients with Stage I LUAD in the TCGA cohort and GEO datasets. ROC curve analysis and DCA indicated that the nomogram combining clinicopathologic characteristics and the 15-gene risk signature had better predictive performance than the 15-gene risk signature alone.

GSEA was used to further explore the 15-gene risk signature. Cell cycle, DNA replication, Parkinson's disease, pyrimidine metabolism, and ribosome pathways were activated in patients with high-risk scores, while allograft rejection, asthma, the intestinal immune network for IgA production, primary immunodeficiency, and systemic lupus erythematosus pathways were activated in patients with low-risk scores. High-risk patients had higher TMB than low-risk patients, and patients with high-risk scores and high TMB values had a poor prognosis. These data imply that the tumor immune microenvironment may be a prognostic factor in patients with LUAD and that patients with low-risk scores and CTLA4-positive and PD1-negative tumors may benefit from immunotherapy.

Some of the DEGs in the gene risk signature have been characterized. *DKK1* is a specific inhibitor of the canonical Wnt pathway (45). *DKK1* may reduce tumor cell migration and invasion by inhibiting the expression of  $\beta$ -catenin (46). The downregulation of *DKK1* may allow tumor cells to escape NK-cell-mediated cytotoxicity. *FAIM2* has been identified as an antiapoptotic protein that may protect cells from Fas-induced apoptosis. *FAIM2* may promote bone metastasis through the Wnt signaling pathway in patients with non-small-cell lung cancer (47, 48). *FGF5* is involved in many biological processes, including embryonic development, mitosis, and cell growth by regulating the cell cycle and VEGF pathway (49, 50). *PCP2* is a member of the R2B subfamily and is considered a tumor suppressor that influences the development of many cancers (51). *PCP2* can regulate the proliferation and differentiation of

megakaryocyte cells (51). The data characterizing the other DEGs in the 15-gene risk signature are limited.

To the authors' knowledge, the present study is the first to focus on the prognostic significance of pGGO-related DEGs in early-stage LUAD. As GGO is considered the appearance of early lung cancer and an important prognostic parameter in early-stage LUAD, our pGGO-related gene signature may contribute to patient classification, and have a clinical value in the diagnosis of patients with early-stage LUAD, and inform individualized treatment decisions. Patients with early-stage LUAD have a relatively good prognosis; however, the current staging system is imprecise for prognostic prediction. There is a need for novel prognostic signatures that identify high-risk patients with early-stage LUAD and guide clinical practice. Several reports have described robust gene risk signatures in LUAD, but only a few have focused on early-stage LUAD. Krzystanek et al. analyzed the gene expression from seven published LUAD cohorts and developed a 7-gene prognostic signature to enable better stratification and treatment of patients with Stage I LUAD. Wu et al. used public LUAD cohorts to establish a 21-immune-related gene prognostic signature for estimating OS in early-stage LUAD, recognizing the importance of the immune system in cancer initiation and progression. Peng et al. identified DE lncRNAs in individual cancer patients by comparing the disrupted ordering of expression levels of lncRNAs to stable normal ordering. They developed two lncRNAs' (*C1orf132* and *TMPO-AS1*) prognostic signatures for patients with Stage I-II LUAD who had not received adjuvant therapy.

In conclusion, we constructed a 15-pGGO-related DEG risk signature to predict the prognosis of early-stage LUAD. Risk scores were calculated using the regression coefficients for these pGGO-related DEGs. Patients with Stage IA LUAD and high-risk scores had poor prognoses with mortality approaching patients with Stage II LUAD. Therefore, treatment strategies for patients with Stage II LUAD may be beneficial in high-risk patients with Stage IA LUAD. A prospective randomized clinical trial is needed to confirm these findings.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

ZZ and WY conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft. DL analyzed the data. BH, QC, WP, XP, GT, YL, YT, MP, FY, and XW conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

## FUNDING

This work was supported by the National Natural Science Foundation of China (81972195, FY; 82072594, YT), Fundamental Research Funds for the Central Universities of Central South University (2021zzts0385), the Hunan Provincial Key Area R&D Programmes (2019SK2253, FY, XW; 2020K53424, 2021SK2013), and the Scientific Research Program of Hunan Provincial Health Commission (Grant number: 20201047).

## ACKNOWLEDGMENTS

The authors thank Professor Yongguang Tao (Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education,

Department of Pathology, Xiangya Hospital, Central South University, Hunan, China) for his comments and suggestions throughout the writing process. The authors thank Dateng Li Ph.D. for statistical analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.872387/full#supplementary-material>

**Supplementary Figure 1** | RT-PCR.

## REFERENCES

- Gao J-W, Rizzo S, Ma LH, Qiu XY, Warth A, Seki N, et al. Pulmonary Ground-Glass Opacity: Computed Tomography Features, Histopathology and Molecular Pathology. *Trans Lung Cancer Res* (2017) 6(1):68–75. doi: 10.21037/tlcr.2017.01.02
- Zhang Y, Fu F, Chen H. Management of Ground-Glass Opacities in the Lung Cancer Spectrum. *Ann Thorac Surg* (2020) 110:1796–804. doi: 10.1016/j.athoracsur.2020.04.094
- Kobayashi Y, Mitsudomi T. Management of Ground-Glass Opacities: Should All Pulmonary Lesions With Ground-Glass Opacity Be Surgically Resected? *Trans Lung Cancer Res* (2013) 2:354. doi: 10.3978/j.issn.2218-6751.2013.09.03
- Hattori A, Hirayama S, Matsunaga T, Hayashi T, Takamochi K, Oh S, et al. Distinct Clinicopathologic Characteristics and Prognosis Based on the Presence of Ground Glass Opacity Component in Clinical Stage IA Lung Adenocarcinoma. *J Thorac Oncol* (2019) 14(2):265–75. doi: 10.1016/j.jtho.2018.09.026
- Fu F, et al. Distinct Prognostic Factors in Patients With Stage I Non-small Cell Lung Cancer With Radiologic Part-Solid or Solid Lesions. *J Thorac Oncol* (2019) 14:2133–42. doi: 10.1016/j.jtho.2019.08.002
- MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* (2017) 284(1):228–43. doi: 10.1148/radiol.2017161659
- Baldwin DR, Callister ME. The British Thoracic Society Guidelines on the Investigation and Management of Pulmonary Nodules. *Thorax* (2015) 70:794–8. doi: 10.1136/thoraxjnl-2015-207221
- Aokage K, Saji H, Suzuki K, Mizutani T, Katayama H, Shibata T, et al. A Non-Randomized Confirmatory Trial of Segmentectomy for Clinical T1N0 Lung Cancer With Dominant Ground Glass Opacity Based on Thin-Section Computed Tomography (JCOG1211). *Gen Thorac Cardiovasc Surg* (2017) 65(2):267–72. doi: 10.1007/s11748-016-0741-1
- Ettinger DS, Wood DE, Aisner DS, Akerley W, Bauman JR, Bharat A, et al. NCCN Guidelines Insights: Non-small Cell Lung Cancer, Version 1.2020: Featured Updates to the NCCN Guidelines. *J Natl Compr Cancer Network* (2021) 19(3):254–66. doi: 10.6004/jnccn.2019.0059
- Qin Y, Xu Y, Ma D, Tian Z, Huang C, Zhou X, et al. Clinical Characteristics of Resected Solitary Ground-Glass Opacities: Comparison Between Benign and Malignant Nodules. *Thorac Cancer* (2020) 11(10):2767–74. doi: 10.1111/1759-7714.13575
- Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, et al. Lung Cancer: Current Therapies and New Targeted Treatments. *Lancet* (2017) 389(10066):299–311. doi: 10.1016/S0140-6736(16)30958-8
- Migliore M, Fornito M, Palazzolo M, Criscione A, Gangemi M, Borrata F, et al. Ground Glass Opacities Management in the Lung Cancer Screening Era. *Ann Trans Med* (2018) 6(5):90. doi: 10.21037/atm.2017.07.28
- Robbins HA, Katki HA, Cheung LC, Landy R, Berg CD. Insights for Management of Ground-Glass Opacities From the National Lung Screening Trial. *J Thorac Oncol* (2019) 14:1662–5. doi: 10.1016/j.jtho.2019.05.012
- Sakurai H, Asamura H. Sublobar Resection for Early-Stage Lung Cancer. *Trans Lung Cancer Res* (2014) 3:164. doi: 10.3978/j.issn.2218-6751.2014.06.11
- Kobayashi Y, Ambrogio C, Mitsudomi T. Ground-Glass Nodules of the Lung in Never-Smokers and Smokers: Clinical and Genetic Insights. *Trans Lung Cancer Res* (2018) 7:487. doi: 10.21037/tlcr.2018.07.04
- Huang C, Wang C, Wang Y, Liu J, Bie F, Wang Y, et al. The Prognostic Significance of Pure Ground Glass Opacities in Lung Cancer Computed Tomographic Images. *J Cancer* (2019) 10(27):6888–95. doi: 10.7150/jca.33132
- Oh J-Y, Kwon SY, Yoon HI, Lee SM, Yim JJ, Lee JH, et al. Clinical Significance of a Solitary Ground-Glass Opacity (GGO) Lesion of the Lung Detected by Chest CT. *Lung Cancer* (2007) 55(1):67–73. doi: 10.1016/j.lungcan.2006.09.009
- Mimae T, Tsutani Y, Miyata Y, Imai K, Ito H, Nakayama T, et al. Solid Tumor Size of 2 Cm Divides Outcomes of Patients With Mixed Ground Glass Opacity Lung Tumors. *Ann Thorac Surg* (2020) 109(5):1530–6. doi: 10.1016/j.athoracsur.2019.12.008
- Han SJ, Jeon JH, Jung W, Seong YW, Cho S, Kim K, et al. Do Ground-Glass Opacity-Dominant Features Have Prognostic Significance in Node-Negative Adenocarcinomas With Invasive Components of Similar Sizes? *Eur J Cardio-Thoracic Surg* (2020) 57(6):1189–94. doi: 10.1093/ejcts/ezaa016
- Hiramatsu M, Inagaki T, Inagaki T, Matsui Y, Satoh Y, Okumura S, et al. Pulmonary Ground-Glass Opacity (GGO) Lesions—large Size and a History of Lung Cancer Are Risk Factors for Growth. *J Thorac Oncol* (2008) 3(11):1245–50. doi: 10.1097/JTO.0b013e318189f526
- Chen K, Bai J, Reuben A, Zhao H, Kang G, Zhang C, et al. Multiomics Analysis Reveals Distinct Immunogenomic Features of Lung Cancer with Ground-Glass Opacity. *Am J Respir Crit Care Med* (2021) 204(10):1180–92. doi: 10.1164/rccm.202101-0119OC
- Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A Compendium of Mutational Cancer Driver Genes. *Nat Rev Cancer* (2020) 20(10):555–72. doi: 10.1038/s41568-020-0290-x
- Gardini E, Giorgi FM, Decherchi S, Cavalli A. Spathial: An R Package for the Evolutionary Analysis of Biological Data. *Bioinf (Oxford England)* (2020) 36:4664–7. doi: 10.1093/bioinformatics/btaa273
- Ritchie ME, Phipson B, Wu D, Hu ME, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
- Mangiola S, Doyle MA, Papenfuss AT. Interfacing Seurat with the R tidy Universe. *Bioinf (Oxford England)* (2021) 24:btab404. doi: 10.1093/bioinformatics/btab404
- Rosales RA, Drummond RD, Valieris R, Dias-Neto E, da Silva IT. signeR: An Empirical Bayesian Approach to Mutational Signature Discovery. *Bioinf (Oxford England)* (2017) 33:8–16. doi: 10.1093/bioinformatics/btw572
- Sharma G, Colantuoni C, Goff LA, Fertig EJ, Stein-O'Brien G. ProjectR: an R/Bioconductor Package for Transfer Learning. *Via PCA NMF Correlation Clustering Bioinf (Oxford England)* (2020) 36:3592–3. doi: 10.1093/bioinformatics/btaa183
- Xu M, Li Ys, Li W, Zhao Q, Zhang Q, Le K, et al. Immune and Stroma Related Genes in Breast Cancer: A Comprehensive Analysis of Tumor

- Microenvironment Based on the Cancer Genome Atlas (TCGA) Database. *Front Med* (2020) 7:64. doi: 10.3389/fmed.2020.00064
29. Dienstmann R, Villacampa G, Sveen A, Mason MJ, Niedzwiecki D, Nesbakken A, et al. Relative Contribution of Clinicopathological Variables, Genomic Markers, Transcriptomic Subtyping and Microenvironment Features for Outcome Prediction in Stage II/III Colorectal Cancer. *Ann Oncol Off J Eur Soc Med Oncol* (2019) 30(10):1622–9. doi: 10.1093/annonc/mdz287
  30. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48(4):812–30.e814. doi: 10.1016/j.immuni.2018.03.023
  31. Nedaie A, Najafi AA. Support Vector Machine With Dirichlet Feature Mapping. *Neural Networks Off J Int Neural Netw Soc* (2018) 98:87–101. doi: 10.1016/j.neunet.2017.11.006
  32. Tang J, Tian Y, Liu X, Li D, Lv J, Kou G, et al. Improved Multi-View Privileged Support Vector Machine. *Neural Netw Off J Int Neural Netw Soc* (2018) 106:96–109. doi: 10.1016/j.neunet.2018.06.017
  33. Yu J, Zhu M, Lv M, Wu X, Zhang X, Zhang Y, et al. Characterization of a Five-microRNA Signature as a Prognostic Biomarker for Esophageal Squamous cell carcinoma. *Sci Rep* (2019) 9(1):19847. doi: 10.1038/s41598-019-56367-1
  34. Tu Z, Wu L, Wang P, Hu Q, Tao C, Li K, et al. N6-Methyladenosine-Related lncRNAs Are Potential Biomarkers for Predicting the Overall Survival of Lower-Grade Glioma Patients. *Front Cell Dev Biol* (2020) 8:642. doi: 10.3389/fcell.2020.00642
  35. Wang C, Ye Q, Luo P, Ye N, Fu L. Robust Capped L1-Norm Twin Support Vector Machine. *Neural Networks Off J Int Neural Netw Soc* (2019) 114:47–59. doi: 10.1016/j.neunet.2019.01.016
  36. Chen Q, Cao F. Distributed Support Vector Machine in Master-Slave Mode. *Neural Networks Off J Int Neural Netw Soc* (2018) 101:94–100. doi: 10.1016/j.neunet.2018.02.006
  37. Jarmolinska AI, Zhou Q, Sulkowska JI, Morcos F. DCA-MOL: A PyMOL Plugin To Analyze Direct Evolutionary Couplings. *J Chem Inf Modeling* (2019) 59:625–9. doi: 10.1021/acs.jcim.8b00690
  38. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-Dependent ROC Curve Analysis in Medical Research: Current Methods and Applications. *BMC Med Res Method* (2017) 17:53. doi: 10.1186/s12874-017-0332-6
  39. Zhang L, Chen J, Ning D, Liu Q, Wang C, Zhang Z, et al. FBXO22 Promotes the Development of Hepatocellular Carcinoma by Regulating the Ubiquitination and Degradation of p21. *J Exp Clin Cancer Res CR* (2019) 38(1):101. doi: 10.1186/s13046-019-1058-6
  40. Powers RK, Goodspeed A, Pielke-Lombardo H, Tan AC, Costello JC. GSEA-InContext: Identifying Novel and Common Patterns in Expression Experiments. *Bioinf (Oxford England)* (2018) 34:i555–64. doi: 10.1093/bioinformatics/bty271
  41. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
  42. Zhao Y, Liu X, Xiao K, Wang L, Li Y, Kan M, et al. Clinicopathological Value of Long Non-Coding RNA Profiles in Gastrointestinal Stromal Tumor. *PeerJ* (2021) 9:e11946. doi: 10.7717/peerj.11946
  43. Golpour P, Ghayour-Mobarhan M, Saki A, Esmaily H, Taghipour A, Tajfard M, et al. Comparison of Support Vector Machine, Naïve Bayes and Logistic Regression for Assessing the Necessity for Coronary Angiography. *Int J Environ Res Public Health* (2020) 17(18):6449. doi: 10.3390/ijerph17186449
  44. Qian J, Massion PP. Next-Generation Molecular Therapy in Lung Cancer. *Trans Lung Cancer Res* (2018) 7:S31–4. doi: 10.21037/tlcr.2018.01.03
  45. Stewart DJ. Wnt Signaling Pathway in Non-Small Cell Lung Cancer. *J Natl Cancer Institute* (2014) 106:djt356. doi: 10.1093/jnci/djt356
  46. Niu J, Li XM, Wang X, Liang C, Zhang YD, Li HY, et al. DKK1 Inhibits Breast Cancer Cell Migration and Invasion Through Suppression of  $\beta$ -catenin/MMP7 Signaling Pathway. *Cancer Cell Int* (2019) 19:168. doi: 10.1186/s12935-019-0883-1
  47. She K, Yang W, Li M, Xiong W, Zhou M. FAIM2 Promotes Non-Small Cell Lung Cancer Cell Growth and Bone Metastasis by Activating the Wnt/ $\beta$ -Catenin Pathway. *Front Oncol* (2021) 11:690142. doi: 10.3389/fonc.2021.690142
  48. Planells-Ferrer L, Urresti J, Coccia E, Galenkamp KM, Calleja-Yagüe L, López-Soriano J, et al. Fas Apoptosis Inhibitory Molecules: More Than Death-Receptor Antagonists in the Nervous System. *J Neurochem* (2016) 139(1):11–21. doi: 10.1111/jnc.13729
  49. Zhou Y, Yu Q, Chu Y, Zhu X, Deng J, Liu Q, et al. Downregulation of Fibroblast Growth Factor 5 Inhibits Cell Growth and Invasion of Human Nonsmall-Cell Lung Cancer Cells. *J Cell Biochem* (2018). doi: 10.1002/jcb.28107
  50. Itoh N, Nakayama Y, Konishi M. Roles of FGFs As Paracrine or Endocrine Signals in Liver Development, Health, and Disease. *Front Cell Dev Biol* (2016) 4:30. doi: 10.3389/fcell.2016.00030
  51. Craig SE, Brady-Kalnay SM. Regulation of Development and Cancer by the R2B Subfamily of RPTPs and the Implications of Proteolysis. *Semin Cell Dev Biol* (2015) 37:108–18. doi: 10.1016/j.semcdb.2014.09.004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhao, Yin, Peng, Cai, He, Shi, Peng, Tu, Li, Li, Tao, Peng, Wang and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.