



## OPEN ACCESS

## EDITED BY

Ramcés Falfán-Valencia,  
Instituto Nacional de Enfermedades  
Respiratorias-México (INER), Mexico

## REVIEWED BY

Ming Zheng,  
Academy of Military Medical Sciences,  
China  
Katsunori Shigehara,  
Sapporo Medical University, Japan  
Roshni Roy,  
National Institute on Aging (NIH),  
United States  
Mathieu Garand,  
Washington University in St. Louis,  
United States

## \*CORRESPONDENCE

Jun Liu

✉ liuyu8566@126.com

Shan-Qiang Zhang

✉ zhangshanqiang@stu.edu.cn

†These authors have contributed  
equally to this work

## SPECIALTY SECTION

This article was submitted to  
Systems Immunology,  
a section of the journal  
Frontiers in Immunology

RECEIVED 03 August 2022

ACCEPTED 06 December 2022

PUBLISHED 19 December 2022

## CITATION

Yin Y-Q, Peng F, Situ H-J, Xie J-L,  
Tan L, Wei J, Jiang F-f, Zhang S-Q and  
Liu J (2022) Construction of prediction  
model of inflammation related genes  
in idiopathic pulmonary fibrosis and its  
correlation with immune  
microenvironment.

*Front. Immunol.* 13:1010345.

doi: 10.3389/fimmu.2022.1010345

## COPYRIGHT

© 2022 Yin, Peng, Situ, Xie, Tan, Wei,  
Jiang, Zhang and Liu. 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.

# Construction of prediction model of inflammation related genes in idiopathic pulmonary fibrosis and its correlation with immune microenvironment

Ying-Qiu Yin<sup>1</sup>, Feng Peng<sup>1</sup>, Hui-Jing Situ<sup>2</sup>, Jun-Ling Xie<sup>1</sup>,  
Liming Tan<sup>1</sup>, Jie Wei<sup>1</sup>, Fang-fang Jiang<sup>1</sup>, Shan-Qiang Zhang<sup>3\*†</sup>  
and Jun Liu<sup>3\*†</sup>

<sup>1</sup>Department of Respiratory Medicine, Yue Bei People's Hospital, Shantou University Medical College, Shaoguan, Guangdong, China, <sup>2</sup>Department of Radiotherapy, Yue Bei People's Hospital, Shantou University Medical College, Shaoguan, Guangdong, China, <sup>3</sup>Medical Research Center, Yue Bei People's Hospital, Shantou University Medical College, Shaoguan, Guangdong, China

**Background:** The role of inflammation in the formation of idiopathic pulmonary fibrosis (IPF) has gained a lot of attention recently. However, the involvement of genes related to inflammation and immune exchange environment status in the prognosis of IPF remains to be further clarified. The objective of this research is to establish a new model for the prediction of the overall survival (OS) rate of inflammation-related IPF.

**Methods:** Gene Expression Omnibus (GEO) was employed to obtain the three expression microarrays of IPF, including two from alveolar lavage fluid cells and one from peripheral blood mononuclear cells. To construct the risk assessment model of inflammation-linked genes, least absolute shrinkage and selection operator (lasso), univariate cox and multivariate stepwise regression, and random forest method were used. The proportion of immune cell infiltration was evaluated by single sample Gene Set Enrichment Analysis (ssGSEA) algorithm.

**Results:** The value of genes linked with inflammation in the prognosis of IPF was analyzed, and a four-genes risk model was constructed, including *tpbg*, *Myc*, *ffar2*, and *CCL2*. It was highlighted by Kaplan Meier (K-M) survival analysis that patients with high-risk scores had worse overall survival time in all training and validation sets, and univariate and multivariate analysis highlighted that it has the potential to act as an independent risk indicator for poor prognosis. ROC analysis showed that the prediction efficiency of 1-, 3-, and 5-year OS time in the training set reached 0.784, 0.835, and 0.921, respectively. Immune infiltration analysis showed that Myeloid-Derived Suppressor Cells (MDSC), macrophages, regulatory T cells, cd4+ t cells, neutrophils, and dendritic cells were more infiltrated in the high-risk group than in the low-risk group.

**Conclusion:** Inflammation-related genes can be well used to evaluate the IPF prognosis and impart a new idea for the treatment and follow-up management of IPF patients.

#### KEYWORDS

idiopathic pulmonary fibrosis, inflammation, immune microenvironment, prognosis, ssGSEA

## Introduction

Idiopathic pulmonary fibrosis (IPF) is one of the most frequent malignant diseases associated with interstitial pneumonia (1). IPF is chronic, progressive, and often occurs in men after the age of 50, but its specific pathogenesis remains to be studied further (2, 3). IPF is mainly caused by chronic inflammatory stimulation and pulmonary interstitial fibrosis proliferation, which often causes dyspnea in patients, and in severe cases, induces respiratory failure, leading to death (4). After the onset of IPF, its progress is relatively rapid, the prognosis is poor, and the average median survival time is < 4 years (5). Due to the deterioration of the environment and the increase in the aging population, the incidence rate of IPF also shows an increasing trend every year (6, 7). IPF has high heterogeneity, and there are great differences in the rate of disease progression between different patients (2). At present, although pirfenidone and nidanib have been approved for the clinical treatment of IPF, their clinical efficacy is still not ideal, and clinical studies show that they are unable to reduce mortality (8–10). Research shows that monitoring the disease process of IPF patients in order to implement precise and individualized management and supportive care helps improve the survival time and quality of life of IPF patients (11). However, effective prognostic indicators for IPF are still needed.

Transcriptomics has been paid increasing attention by researchers in the study of disease occurrence and development (12, 13). Presently with the advancement made in the field of genomics, a bulk of accessible genomic and clinical data is available, providing greater convenience and possibilities for the study of the occurrence of diseases. Recently, the biological processes and potential biomarkers related to the occurrence of diseases have been reported by many researchers through public databases (14, 15). However, the role of the genes related to inflammation in the prognosis of IPF needs to be studied in further detail. Compared with lung biopsy, bronchoalveolar lavage fluid (BALF) has better patient compliance and wider application prospects in the diagnosis of lung diseases (16, 17). In individuals with IPF in BALF samples, researchers observed a remarkable increase in the amount of platelet-derived growth factor (PDGF) (18). PDGF has been

proved to be related to pulmonary angiogenesis and pulmonary hypertension and is significantly involved in the early stage of pulmonary fiber formation (19). In addition, TGF- $\beta$  was also found to be significantly overexpressed in BALF of IPF patients (20). Studies have shown that TGF- $\beta$  is expressed in alveolar epithelial cells and macrophages, and its overexpression can lead to pulmonary fibrosis in rats (21). These results suggest that BALF can be used to analyze the occurrence of pulmonary fibrosis and its development.

In this study, the mRNA expression microarray datasets of BALF and PBMC were obtained from Gene Expression Omnibus (GEO) database. This study investigates the involvement of genes linked with inflammation in the prognosis of IPF, screens and develops a risk model for the prognosis evaluation of IPF with the help of a variety of machine learning methods, and verifies its robustness. This research gives a new insight to understand the role that inflammation plays in the prognosis of IPF and develops a risk model of genes related to inflammation that can be employed for the prediction of the overall survival.

## Methods and materials

### Acquisition and processing of microarray data

The expression microarray data of bronchoalveolar lavage (BAL) cells (GSE70866), were obtained from the GEO database. Among them, there were 20 individuals from the control group and 112 individuals with IPF patients. A total of 132 mRNA microarray data were obtained from Agilent-028004 SurePrint G3 Human GE 8x60K Microarray, GPL14550 platform. The microarray data of 64 IPF patients was obtained through Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray, GPL17077 detection platform. In addition, the gene expression microarray data of peripheral blood mononuclear cell (PBMC), GSE28042, was also obtained from the GEO database. This dataset contains 75 patients with IPF, and their microarray information detection was done by Agilent-014850 Whole Human Genome Microarray 4x44K

G4112F platform. The clinical data and OS follow-up information of all patients with IPF were also obtained from the GEO database. The “SVA” function of the “limma” package was used to correct the background, standardize and remove the batch effect of three groups of original data sets. Among them, 132 patients from the GSE70866 dataset GPL14550 were used as the discovery set, 64 IPF patients from the GPL17077 platform, and 75 IPF patients from the GSE28042 queue were used as the independent validation set.

## Acquisition and analysis of inflammation-related genes

The molecular signatures database (MSigDB, <http://www.GSEa-msigdb.org/GSEa/index.jsp>) was employed to obtain the inflammatory response-related gene set. To determine the differentially expressed inflammation-linked genes the “limma” package was utilized, and  $p < 0.05$  and False Discovery Rate (FDR)  $< 0.05$  were considered as cut-off values. The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to enrich the differentially expressed genes (DEGs) using the “clusterProfiler” package.

## Stratification of IPF patients

The identification of genes linked with the prognosis of IPF was made by Univariate Cox regression. The least absolute shrinkage and selection operator (Lasso) regression is often used to determine characteristic variables, which can well retain valuable variables and avoid overfitting. In this study, Lasso regression was employed for the identification of characteristic genes related to OS. Stochastic forest analysis is an analytical method that uses a decision tree to evaluate the importance of variables. The random forest was utilized to rank the involvement of genes in OS, and then the intersection of the genes identified by the two methods was taken. Finally, the gene set independently related to OS was identified through multivariate stepwise Cox regression, and a risk stratification model was constructed.

## Immune correlation analysis

Through the single sample Gene Set Enrichment Analysis (ssGSEA) algorithm of the “GSVA” R package, the level of immune cell infiltration was evaluated in the discovery set and validation set, respectively, in accordance with the gene expression array. CIBERSORT algorithm (<https://cibersortx.stanford.edu/>) was also used to evaluate the immune cell infiltration.

## Bioinformatics analysis and statistical analysis

R software was employed for the complete data analysis of this study. The “ggplot2” package was used to perform chart drawing. “Survival” and “Survminer” packages were used to draw Kaplan Meier (K-M) curves, and for the evaluation of the statistical significance of survival time distribution, the log-rank test was employed. The univariate and multivariate analysis was performed using the “survival” package. Wilcoxon test was employed to evaluate the variation between non-normal distribution variables, and the value of  $p < 0.05$  was deemed statistically significant. The receiver operating characteristic (ROC) curve was drawn employing the ‘survivalROC’ package, and the area under the curve (AUC) was also calculated using the package. The “RMS” package was employed to draw the calibration of nomograms. Pearson correlation analysis was performed to evaluate the relationship between immune cell proportion and risk score.

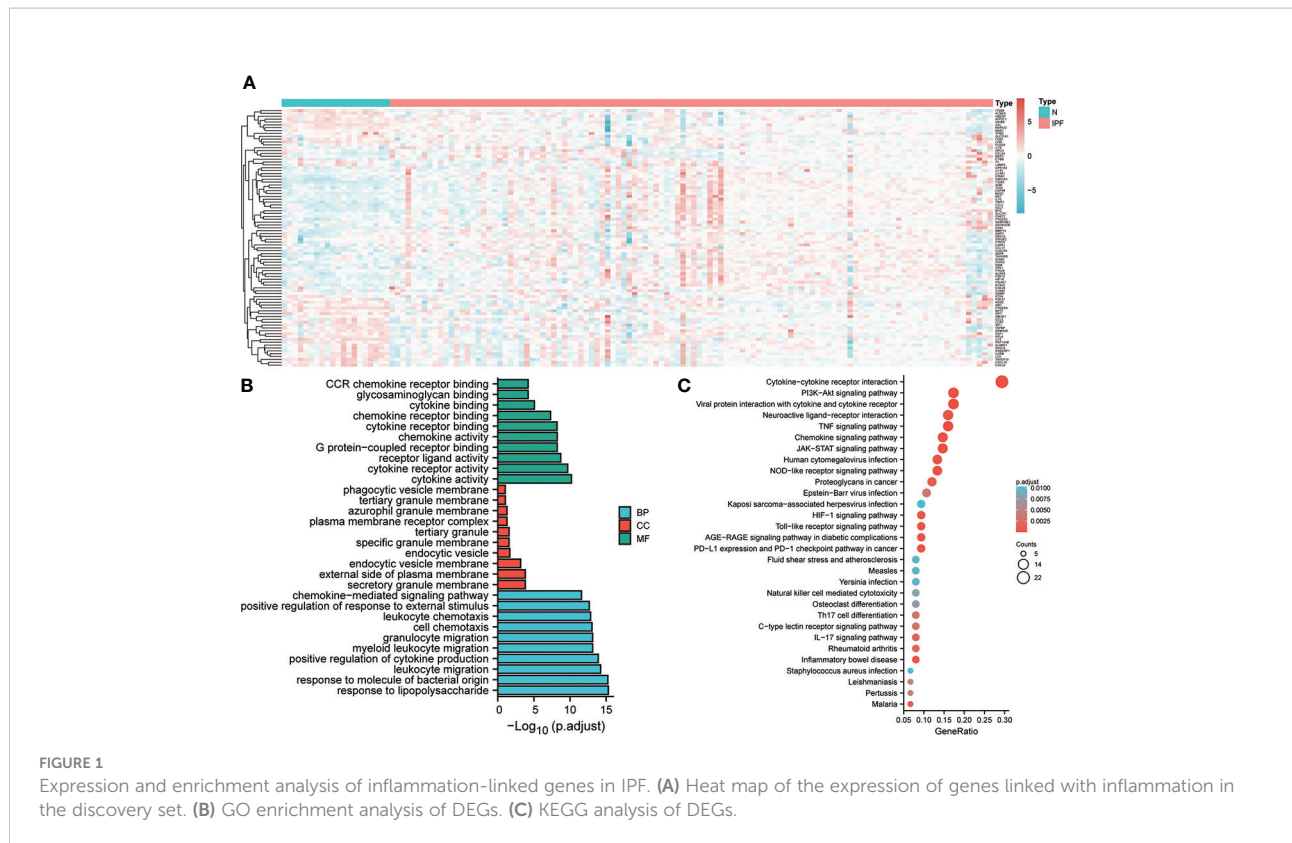
## Results

### Differential analysis of inflammation-related genes

In the alveolar lavage fluid of 20 healthy controls and 112 individuals with IPF, a total number of 90 differentially expressing inflammation-related genes were identified, and their expression in the two groups is shown in the heat map (Figure 1A). Subsequently, Gene Ontology (GO) enrichment analysis revealed that the Biological Process (BP) to which these 90 differential genes were mainly enriched in response to lipopolysaccharide, leucocyte migration, and positive regulation of cytokine production. The enriched Cellular Component (CC) was the granular secretory membrane, the external side of the plasma membrane, and the endocytic vesicle membrane. The Molecular Function (MF) enriched were mainly cytokine activity, cytokine receptor activity, and receptor-ligand activity (Figure 1B). These DEGs were mainly linked to the PI3K-Akt signaling pathway, Cytokine-cytokine receptor interaction, neuroactive ligand-receptor interaction, TNF signaling, and other pathways, as shown by KEGG enrichment analysis. (Figure 1C).

### Construction of inflammation-related gene model

Initially, the identification of 46 genes associated with the poor prognosis of individuals with IPF from 90 DEGs was carried out by univariate Cox regression, including 10 negative



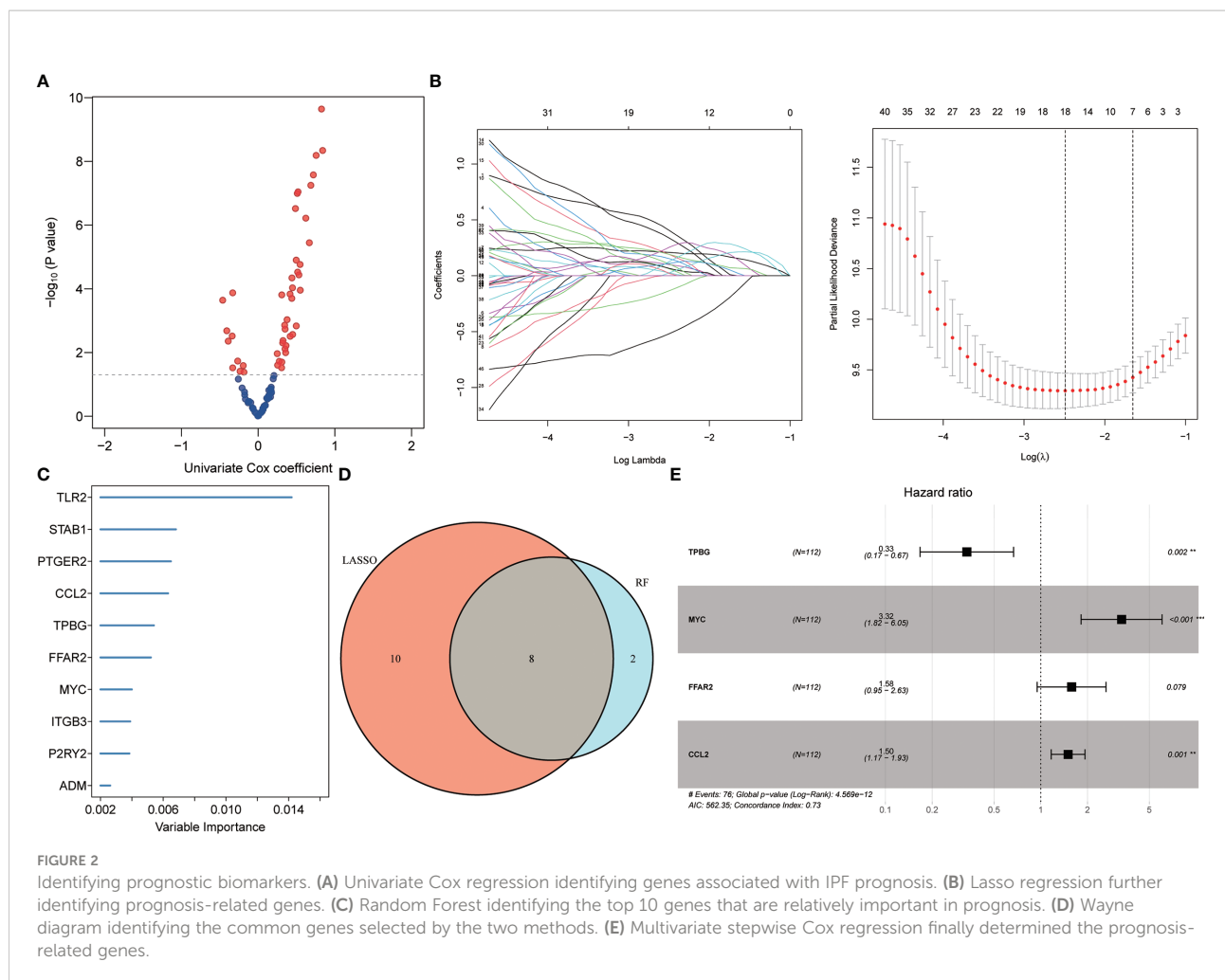
**FIGURE 1** Expression and enrichment analysis of inflammation-linked genes in IPF. **(A)** Heat map of the expression of genes linked with inflammation in the discovery set. **(B)** GO enrichment analysis of DEGs. **(C)** KEGG analysis of DEGs.

correlations and 36 positive correlations (Figure 2A). Subsequently, 18 of the most characteristic inflammation-related genes were identified from 46 prognosis-related genes by Lasso regression (Figure 2B). Afterward, the prognostic genes were sorted, and the top 10 genes of relative importance were identified based on the random forest algorithm (Figure 2C). In order to identify more robust prognostic markers, the intersection of genes was taken from Lasso regression and random forest screening, and 8 genes were obtained in total (Figure 2D). Further, a 4-gene model related to inflammation was constructed through multivariate stepwise Cox regression (Figure 2E). Its risk score (RS) = -1.096\* expression of TPBG + 1.199\* expression of MYC + 0.457\* expression of FFAR2 + 0.405\* expression of CCL2.

### Prognostic importance of the four-gene risk model

A prognostic classifier was constructed in the discovery set based on this four-gene risk model. As per the median value of its RS, the discovery set was sorted into a low-risk group and a high-risk group. The K-M curve revealed that the individuals in the high-risk group had a worse OS time as compared to the low-risk patients' group (hazard ratio (HR) = 3.48, p < 0.001, Figure 3A). Remarkable high scores were shown by patients in

the high-risk group, and higher mortality was also seen in this group (Figure 3B). PCA analysis revealed that there was a notable distribution variation in the low-risk group and the high-risk group (Figure 3C). Then, ROC analysis was utilized to evaluate the prediction efficiency of the risk model. It was shown by the results that 1-year, 3-year, and 5-year survival AUC values predicted by the model were 0.784, 0.835, and 0.921, respectively, showing a strong predictive ability (Figure 3D). Univariate and multivariate Cox analysis revealed that our risk model worked as an independent risk factor for the poor prognosis of individuals with IPF (Figure 3E). The concordance index showed that our model is much better than the prediction efficiency of age and gender, which has been maintained above 0.7 (Figure 3F). To illustrate the superiority of the 4-gene model, we compared it with the other 4 genes that were common to LASSO and RF but not included in the model. The K-M curve revealed significant survival differences between high-risk and low-risk groups (HR=3.3, P<0.001) (Figure S1A). The ROC analysis indicated that the risk model also has certain prediction ability in predicting 1- and 3- years, with an AUC of 0.700 and 0.756 (Figure S1B). This shows that the model constructed based on the multivariate Cox stepwise regression method is significantly superior to the model selected by random variables. In addition, we also compared our model with LASSO-based model and RF-based model. ROC analysis showed that the 4-gene signature exhibited a higher power of risk prediction than

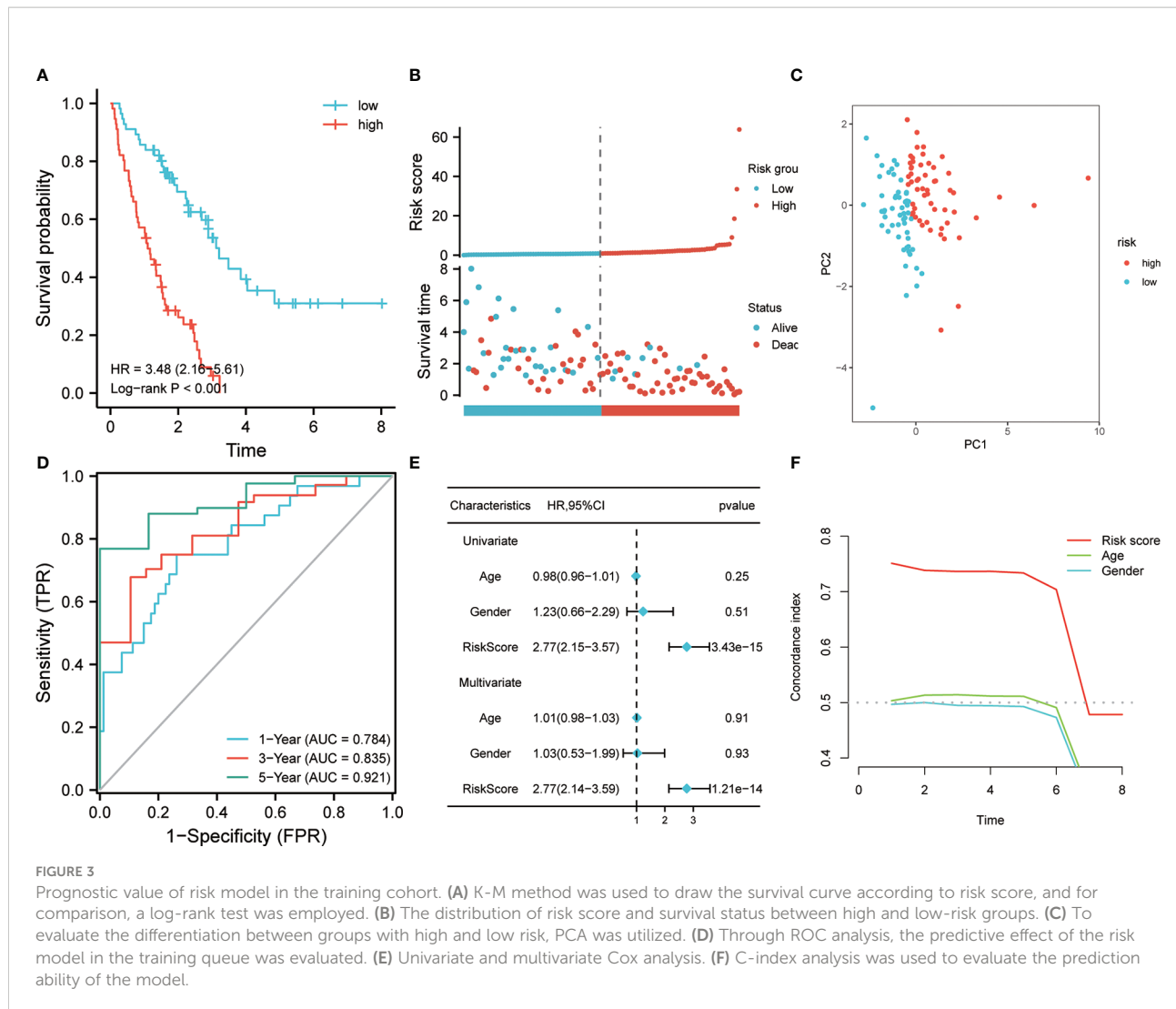


the 7-gene model based on LASSO regression (Figures S2A, B). Although, ROC analysis showed that the 10-gene signature (RF-based model) exhibited a higher power of risk prediction than the 4-gene model (Figures S2C, D). However, the number of genes in the 4-gene model is much less than the number of 10-gene model. In addition, we also performed multivariate Cox stepwise regression on these 10 genes, and finally the 4-gene model still had the lowest Akaike information criterion (AIC).

### Verification of risk model

The 64 patients from the GSE70866 dataset GPL17077 platform were used as the validation set. K-M analysis revealed that the RS showed a certain correlation with the OS time, and the low RS predicted a better prognosis (Figure 4A). The distribution of RSs shows that the group with high risk has higher scores and higher mortality (Figure 4B). PCA analysis

highlighted that the groups with high and low risk exhibited remarkably different clustering trends (Figure 4C). ROC analysis revealed that good prediction efficiency in the validation set was shown by the model, and its respective AUC values in 1, 3, and 4 years were 0.786, 0.713, and 0.793 (Figure 4D). Univariate and multivariate Cox analysis also highlighted that this risk model still worked as an independent risk factor for the poor prognosis of individuals with IPF in the validation set (Figure 4E). C-index also showed the strong prediction ability of the model (Figure 4F). Considering that most of the BALF were immune cells, the predictive value of this model was verified further in the RNA-seq of PBMC of 75 IPF patients from the GSE28042 cohort. It was revealed by the K-M curve that the group with high risk in the GSE28042 cohort still showed a poor OS time (Figure 5A). The distribution of RSs and the correlation between RSs and survival data showed that high-risk groups had higher RSs and higher mortality (Figure 5B). As per the risk model and

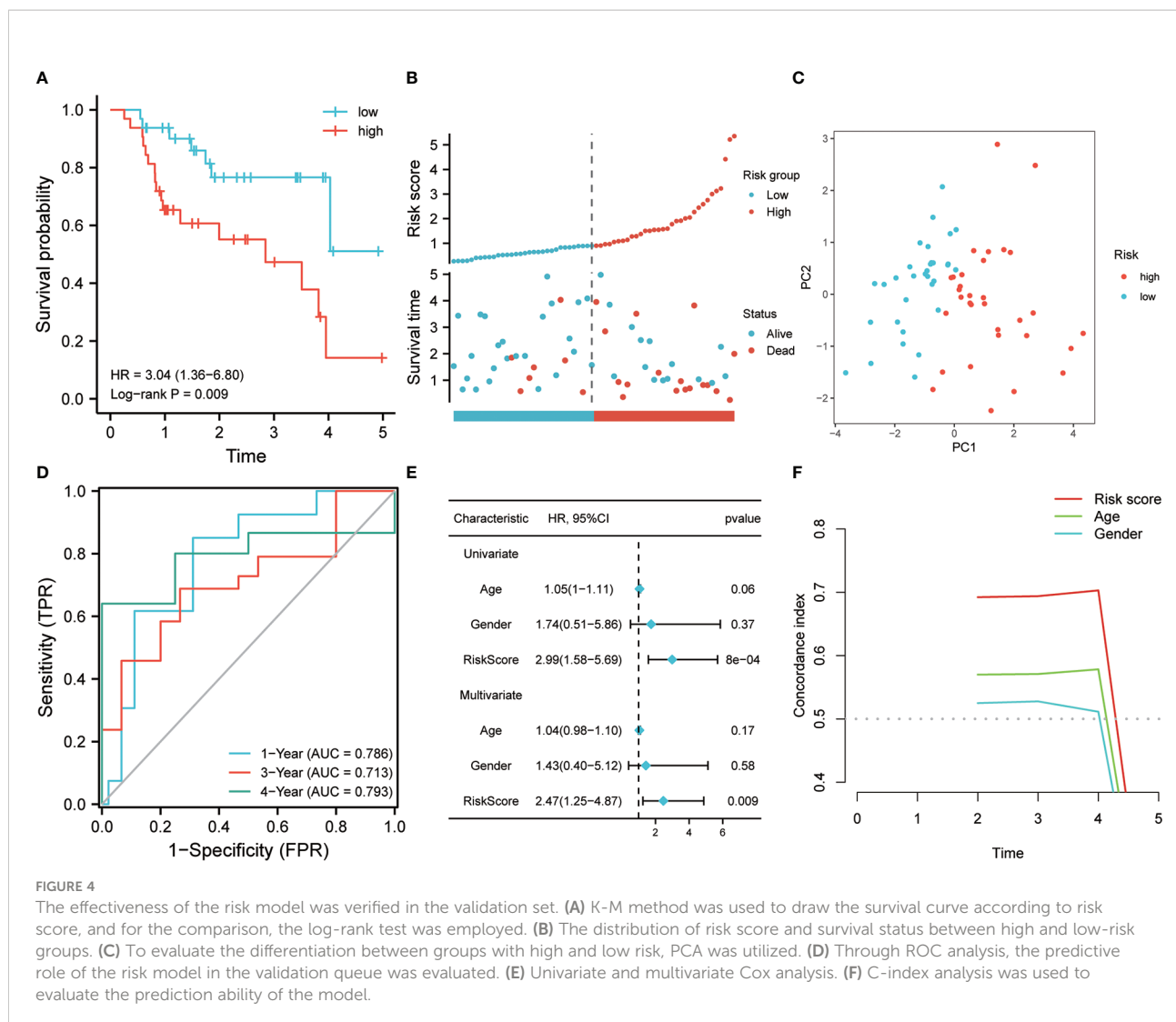


genome expression, PCA proved that the distribution in the group with low risk and group with high risk highlighted notable cluster variations (Figure 5C). It was also shown by ROC analysis that in the external validation set, the risk model showed good prediction efficiency, with respective AUC of 0.662, 0.732, and 0.723 in 1, 2, and 3 years (Figure 5D). Univariate and multivariate Cox regression analysis showed that for the poor prognosis of IPF in the external validation set, the risk model still worked as an independent risk factor (Figure 5E). Similarly, C-index analysis also indicated that the risk model performed well in predicting the OS time of 1, 2, and 3 years (Figure 5F). Further, we analyzed the prognostic significance of this risk model in different ages and gender. The results revealed that the high RS was still linked with poor OS time in people younger than 65 years old or older (Figures 6A, B). Similarly, a higher RS in both men and women predicts a poor prognosis (Figures 6C, D). These

results indicate that our risk model has strong predictive power in the prognosis evaluation of IPF patients.

## Comprehensive analysis of immune cell infiltration

As we know, BAL is rich in immune cells, so the ssGSEA algorithm was used for the evaluation of the variations in the distribution of immune cells in low- and high-risk groups. As expected, the results in the discovery set showed higher infiltration of macrophage, neutrophil, regulatory T cell, and NK cell infiltration in the high-risk group (Figure 7A). Further, the distribution of immune cells in the validation cohort from PBMC was also analyzed, and the results were also consistent with the discovery set (Figure 7B). The group with high risk had more macrophages, neutrophils, regulatory T, and NK cells



among others. The similar results were confirmed by Pearson correlation analysis (Figures S3A, B). In addition, CIBERSORT algorithm also showed that the risk score was positively correlated with Neutrophils, T cells gamma, T cells memory activated, and T cells CD8 (Figure S3C, D). These results suggest that there is higher infiltration of inflammatory cells in the group with high risk, and the stimulation of these inflammatory cells are possibly linked with the poor prognosis of IPF.

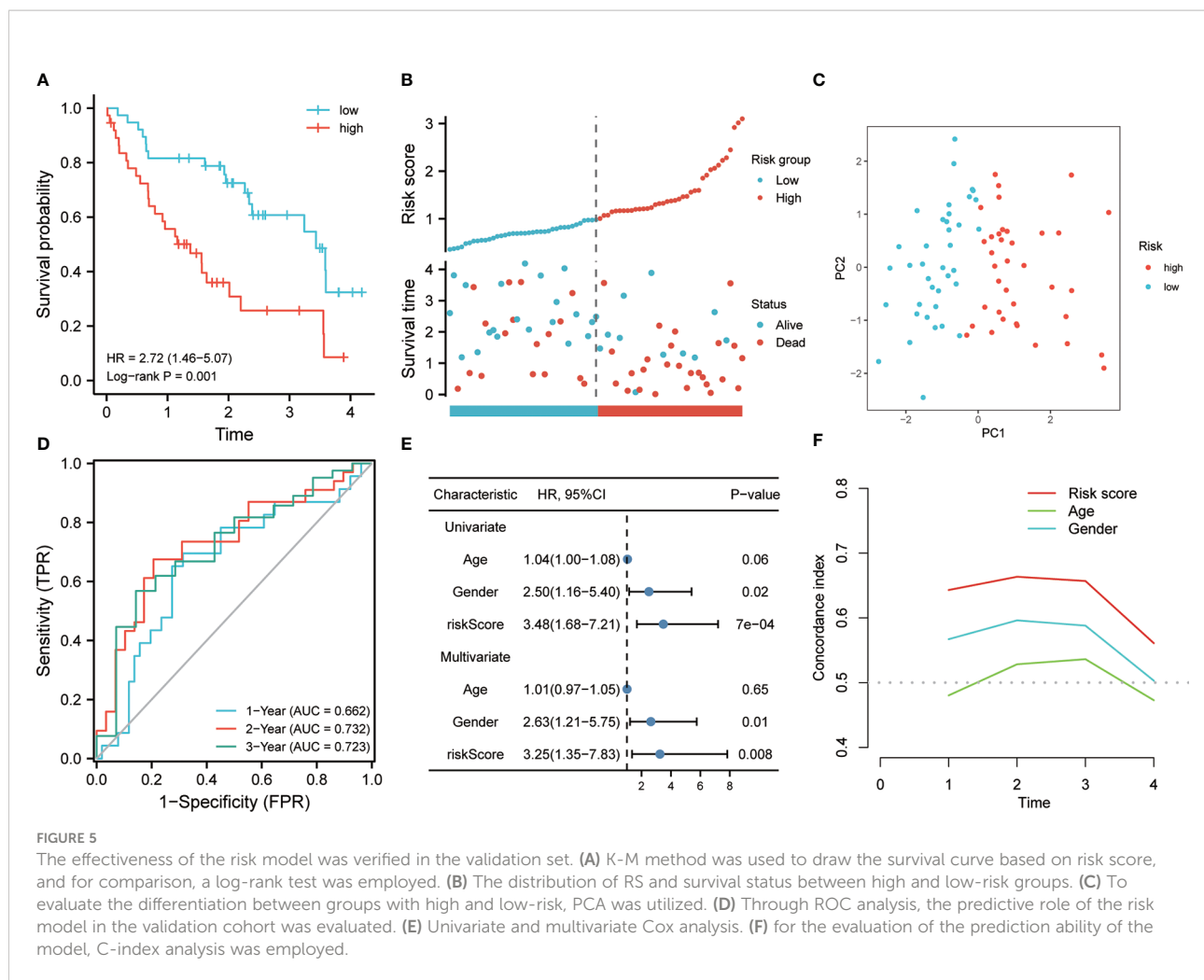
## Construction of nomogram

In order to promote the application of this risk model in clinical practice and facilitate the operability of clinical workers, a four-gene nomogram was constructed (Figure 8A). The nomogram can be used to assess the 1-, 3-, and 5-year OS rates of individuals with IPF. In addition, to find out the nomogram's effectiveness, a calibration chart was drawn, and

the results revealed that the nomogram had a very good prediction performance, which was almost close to the ideal model in the prediction of 1 and 3 years (Figure 8B).

## Discussion

Although the specific pathogenesis of IPF is still controversial, it has been reported by many studies that chronic inflammatory stimulation is crucially involved in the development and progression of IPF (22, 23). At present, researchers generally believe that chronic inflammation caused by lung epithelial cell injury is repeatedly stimulating, resulting in excessive wound repair and tissue cell remodeling, and fibrosis (24, 25). Fibroblasts can induce the secretion of interleukin-33, IL-13, IL-6, IL-1, PDGF, fibroblast growth factor, and TGF- $\beta$ , and promote their differentiation into myofibroblasts (5, 26). In addition, the disturbance of the homeostasis of various immune cells is also



considered to be an important reason for the formation of lung fibers (27). Macrophages are considered to be the main culprit of balance disorder in the process of wound healing (28, 29). Macrophages can also further promote the process of fibrosis by secreting related cytokines (26). However, the underlying mechanism of inflammation in the formation of pulmonary fibrosis still needs further exploration. In this study, we evaluated the expression of inflammation-linked genes in patients with IPF through the microarray expression chip of BALF and explored the link between their expression and the prognosis of patients. GO and KEGG enrichment analysis showed that these DEGs were also involved in a variety of viral infection related pathways, including viral protein interaction with cytokine and cytokine receptor, Human cytomegalovirus infection, and Epstein-Barr virus infection. This is also consistent with recent studies showing that the presence of viral infections is a significant risk factor for IPF pathogenesis (30, 31). In addition, a new four-gene risk model was designed based on multiple machine learning. The robustness of the model was verified in two independent validation queues.

Inflammation and immune microenvironment changes are the basis of IPF. Regulatory T cells are crucial in maintaining the stability of the immune microenvironment (32). Studies showed that regulatory T cells (Tregs) were significantly enriched in the lungs of IPF mice, and they were also found to be significantly activated in the peripheral blood of individuals with IPF, which was positively correlated with the progress of the disease (33). Tregs can also promote the progress of IPF by inducing the activation of Th2 and Th17 cells (34). Studies have shown that Th2 and Th17 cells can promote the disease process of pulmonary fibrosis, and their increase is associated with a positive correlation with the severity of the disease (35, 36). The researchers' knowledge regarding the role that macrophages have in the development of IPF is also constantly refreshed. Recently, results from single-cell RNA sequencing have identified many new macrophage subpopulations, and their different subpopulations are specifically expressed in different stages of IPF development (37). Myeloid-derived suppressor cells (MDSCs) are also an important part of regulating the immune microenvironment. Studies have shown that MDSCs are found



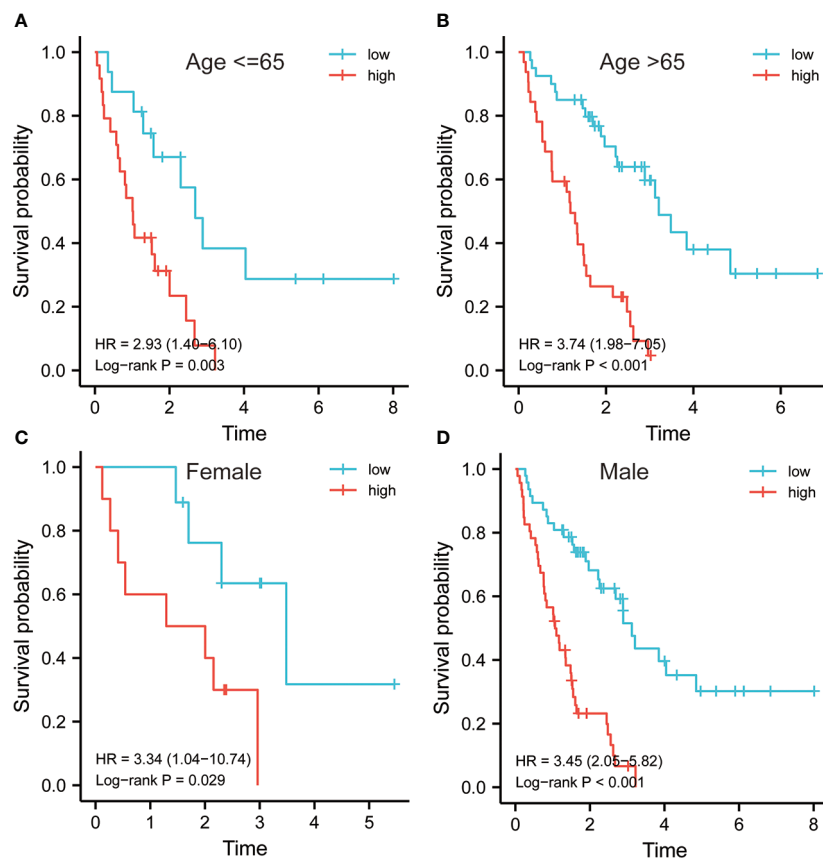


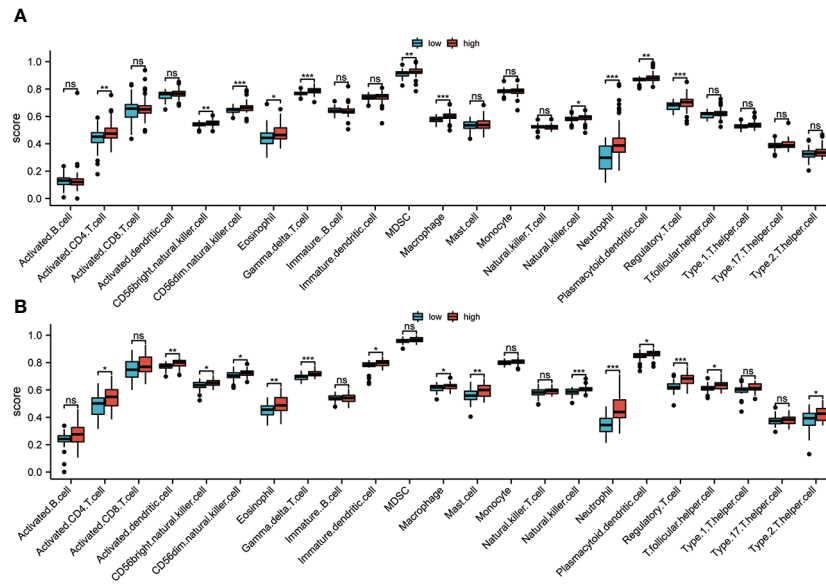
FIGURE 6

Survival analysis of risk model in different subgroups. Survival analysis of risk models in the subgroup of age  $\leq 65$  (A) and greater than 65 (B). Survival analysis of risk models in female (C) and male (D).

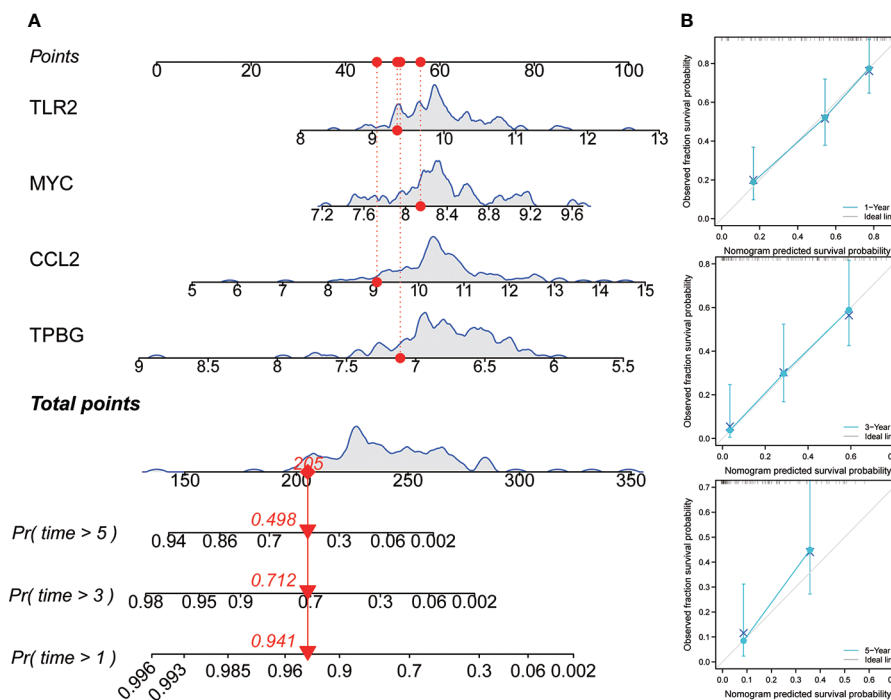
to increase significantly in the lungs, and peripheral blood of IPF patients and are positively correlated with poor lung function (38, 39). These findings are compatible with the findings of this investigation, according to several studies. The results of ssGSEA analysis in this study showed that in the group with high risk, there was enhanced infiltration of MDSC, macrophages, and regulatory T cells, which also predicted a worse OS time. In addition, we also found that in the high-risk group, there was higher infiltration of CD4, T cells, NK cells, dendritic cells, and neutrophils, which may further increase the inflammatory response and aggravate the disease process. A new perspective may be provided by our results for further understanding the characteristics of immune cell infiltration in IPF.

A number of studies have described the prognostic value of gene signature in IPF. Recently, Casanova NG et al. proposed a 21-gene prognostic model to predict the overall survival for IPF patients (40). Li X et al. developed a 9-gene prognostic signature based on hypoxia-immune-related genes (41). In this study, four inflammation-related genes, trophoblast glycoprotein (tpBG), Myc, free fatty acid receptor 2 (ffar2), and CCL2, were

identified as important prognostic markers. It is worth mentioning that the 4-gene signature has fewer genes, but exhibits greater predictive power. TPBG belongs to carcinoembryonic antigen and is mainly involved in cell adhesion (42). It is expressed highly in numerous tumor cells and is related to the poor prognosis of cancer (43, 44). However, there are few reports about its role in IPF. In this study, our results show that TPBG can be used as a good prognostic indicator of IPF. Reportedly, as a proto-oncogene, MYC has a crucial role in cell cycle progression, apoptosis, and cell transformation (45). Research has shown that MYC is up-regulated in pulmonary fibrosis cells of IPF mice and promote the growth and differentiation of pulmonary fibrosis cells by regulating the transcription of miR-9-5p (46). Yin X et al. also reported that MYC could promote the accumulation of HK2 in human and mouse lung fibroblasts, thereby increasing the proliferation and migration of fibroblasts (47). The protein encoded by FFAR2 is a member of the GP40 family that mainly play a role in the regulation of inflammatory response and plasma lipid level (48). Maslowski KM et al. confirmed that



**FIGURE 7**  
Immune infiltration analysis. **(A)** In the discovery set, the immune cell infiltration between groups with high and low risk was evaluated based on the ssGSEA algorithm. **(B)** In the validation set, the immune cell infiltration between high-risk and low-risk groups was analyzed based on the ssGSEA algorithm ns, non significance; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .



**FIGURE 8**  
Construction and evaluation of nomogram. **(A)** Nomogram was constructed based on the expression values of four genes to predict the OS rate at 1, 3, and 5 years. **(B)** The calibration curve was used to assess the nomogram.

short chain fatty acids promote the occurrence of the inflammatory response by binding to FFAR2 (49). The results of Sencio V et al. showed that FFAR2 was also related to the secondary infection of pneumococcus after influenza infection (50). However, we have little knowledge about the role of FFAR2 in IPF. A new direction may be provided by our results for understanding the occurrence and progress of IPF. There is increasing evidence that CCL2 is linked with the onset of pulmonary inflammatory diseases. Several studies have suggested that the concentration of CCL2 in BALF of individuals with IPF has increased significantly, and a significant correlation exists between its expression level and the increased risk of death (51). In addition, CCL2 can also combine with related transcription factors to regulate the proliferation of fibroblasts and promote the differentiation of fibroblasts. Our results further prove that CCL2 from BALF can be used as an effective predictor of the poor prognosis of IPF.

While this study has many strengths, it has some limitations as well. This study analyzed whole of IPF patients, and did not distinguish between rapidly progressive type and slowly progressive type. In addition, this remains to be verified by either further functional studies *in vitro* or *in vivo*.

## Conclusion

In conclusion, this report reveals the importance of genes linked with inflammation in the prognosis of individuals with IPF. A new four-gene risk model based on BALF expression profile was built to predict the progress and prognosis of IPF, which is conducive to further monitoring and management of IPF patients.

## Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

## Author contributions

JL, S-QZ, Y-QY, and FP all participated in research design and drafting the manuscript. JL, S-QZ, Y-QY, FP, H-JS, J-LX, LT, JW, and F-FJ took part in the data collection process of the

study and is responsible for the content of the manuscript. All authors read and approved the final manuscript.

## 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.1010345/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Prognostic value of risk model. (A) K-M method was used to draw the survival curve based on risk score, and for comparison, a log-rank test was employed. (B) Through ROC analysis, the predictive role of the risk model was evaluated.

### SUPPLEMENTARY FIGURE 2

Prognostic value of risk model. (A) K-M method was used to draw the survival curve based on risk score in LASSO-based model. (B) ROC analysis was used to evaluate the predictive power of LASSO-based model. (C) K-M method was used to draw the survival curve based on risk score in RF-based model. (D) ROC analysis was used to evaluate the predictive power of RF-based model.

### SUPPLEMENTARY FIGURE 3

Correlation analysis of risk score and immune cell. Pearson correlation analysis between risk score and immune cell infiltration abundance based on ssGSEA (A) and CIBERSORT (C) methods in discovery set. Pearson correlation analysis between risk score and immune cell infiltration abundance based on ssGSEA (B) and CIBERSORT (D) methods in validation set.

## References

- Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* (2017) 3:17074. doi: 10.1038/nrdp.2017.74
- Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet* (2017) 389(10082):1941–52. doi: 10.1016/S0140-6736(17)30866-8
- Olson AL, Gifford AH, Inase N, Fernandez Perez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. *Eur Respir Rev* (2018) 27(150):180077. doi: 10.1183/16000617.0077-2018
- Molyneaux PL, Maher TM. The role of infection in the pathogenesis of idiopathic pulmonary fibrosis. *Eur Respir Rev* (2013) 22(129):376–81. doi: 10.1183/09059180.00000713
- Phan THG, Paliogiannis P, Nasrallah GK, Giordo R, Eid AH, Fois AG, et al. Emerging cellular and molecular determinants of idiopathic pulmonary fibrosis. *Cell Mol Life Sci* (2021) 78(5):2031–57. doi: 10.1007/s00018-020-03693-7
- Raghu G, Chen SY, Hou Q, Yeh WS, Collard HR. Incidence and prevalence of idiopathic pulmonary fibrosis in US adults 18–64 years old. *Eur Respir J* (2016) 48(1):179–86. doi: 10.1183/13993003.01653-2015
- Navaratnam V, Fleming KM, West J, Smith CJ, Jenkins RG, Fogarty A, et al. The rising incidence of idiopathic pulmonary fibrosis in the U. *K Thorax* (2011) 66(6):462–7. doi: 10.1136/thx.2010.148031
- Lancaster LH, de Andrade JA, Zibrak JD, Padilla ML, Albera C, Nathan SD, et al. Pirfenidone safety and adverse event management in idiopathic pulmonary fibrosis. *Eur Respir Rev* (2017) 26(146):170057. doi: 10.1183/16000617.0057-2017
- Finnerty JP, Ponnuswamy A, Dutta P, Abdelaziz A, Kamil H. Efficacy of antifibrotic drugs, nintedanib and pirfenidone, in treatment of progressive pulmonary fibrosis in both idiopathic pulmonary fibrosis (IPF) and non-IPF: A systematic review and meta-analysis. *BMC Pulm Med* (2021) 21(1):411. doi: 10.1186/s12890-021-01783-1
- Sgalla G, Iovene B, Calvillo M, Ori M, Varone F, Richeldi L. Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir Res* (2018) 19(1):32. doi: 10.1186/s12931-018-0730-2
- Duck A, Pigram L, Errhalt P, Ahmed D, Chaudhuri N. IPF care: a support program for patients with idiopathic pulmonary fibrosis treated with pirfenidone in Europe. *Adv Ther* (2015) 32(2):87–107. doi: 10.1007/s12325-015-0183-7
- de Jong E, Bosco A. Unlocking immune-mediated disease mechanisms with transcriptomics. *Biochem Soc Trans* (2021) 49(2):705–14. doi: 10.1042/BST20200652
- Chambers DC, Carew AM, Lukowski SW, Powell JE. Transcriptomics and single-cell RNA-sequencing. *Respirology* (2019) 24(1):29–36. doi: 10.1111/resp.13412
- Li W, Lu J, Ma Z, Zhao J, Liu J. An integrated model based on a six-gene signature predicts overall survival in patients with hepatocellular carcinoma. *Front Genet* (2019) 10:1323. doi: 10.3389/fgene.2019.01323
- Wang H, Xu H, Cheng Q, Liang C. Identification of a novel stem cell subtype for clear cell renal cell carcinoma based on stem cell gene profiling. *Front Oncol* (2021) 11:758989. doi: 10.3389/fonc.2021.758989
- Nambiar S, Bong How S, Gummer J, Trengove R, Moodley Y. Metabolomics in chronic lung diseases. *Respirology* (2020) 25(2):139–48. doi: 10.1111/resp.13530
- Yamaguchi K, Iwamoto H, Mazur W, Miura S, Sakamoto S, Horimasa Y, et al. Reduced endogenous secretory RAGE in blood and bronchoalveolar lavage fluid is associated with poor prognosis in idiopathic pulmonary fibrosis. *Respir Res* (2020) 21(1):145. doi: 10.1186/s12931-020-01410-3
- Chanda D, Otoupalova E, Smith SR, Volckaert T, De Langhe SP, Thannickal VJ. Developmental pathways in the pathogenesis of lung fibrosis. *Mol Aspects Med* (2019) 65:56–69. doi: 10.1016/j.mam.2018.08.004
- Montero P, Milara J, Roger I, Cortijo J. Role of JAK/STAT in interstitial lung diseases; molecular and cellular mechanisms. *Int J Mol Sci* (2021) 22(12):6211. doi: 10.3390/ijms22126211
- Zhang Y, Bi L, Qiu Y, Wang Y, Ding J, Zhuang Y, et al. Elevated sL1-CAM levels in BALF and serum of IPF patients. *Respirology* (2016) 21(1):143–8. doi: 10.1111/resp.12659
- Sime PJ, Xing Z, Graham FL, Csaky KG, Gaudie J. Adenovector-mediated gene transfer of active transforming growth factor-beta1 induces prolonged severe fibrosis in rat lung. *J Clin Invest* (1997) 100(4):768–76. doi: 10.1172/JCI119590
- Heukels P, Moor CC, von der Thusen JH, Wijsenbeek MS, Kool M. Inflammation and immunity in IPF pathogenesis and treatment. *Respir Med* (2019) 147:79–91. doi: 10.1016/j.rmed.2018.12.015
- Racanelli AC, Kikkers SA, Choi AMK, Cloonan SM. Autophagy and inflammation in chronic respiratory disease. *Autophagy* (2018) 14(2):221–32. doi: 10.1080/15548627.2017.1389823
- Nie YJ, Wu SH, Xuan YH, Yan G. Role of IL-17 family cytokines in the progression of IPF from inflammation to fibrosis. *Mil Med Res* (2022) 9(1):21. doi: 10.1186/s40779-022-00382-3
- Nie Y, Sun L, Wu Y, Yang Y, Wang J, He H, et al. AKT2 regulates pulmonary inflammation and fibrosis via modulating macrophage activation. *J Immunol* (2017) 198(11):4470–80. doi: 10.4049/jimmunol.1601503
- Kral JB, Kuttke M, Schrottmaier WC, Birnecker B, Warszawska J, Wernig C, et al. Sustained PI3K activation exacerbates BLM-induced lung fibrosis via activation of pro-inflammatory and pro-fibrotic pathways. *Sci Rep* (2016) 6:23034. doi: 10.1038/srep23034
- van Geffen C, Deissler A, Quante M, Renz H, Hartl D, Kolahian S. Regulatory immune cells in idiopathic pulmonary fibrosis: Friends or foes? *Front Immunol* (2021) 12:663203. doi: 10.3389/fimmu.2021.663203
- Zhang L, Wang Y, Wu G, Xiong W, Gu W, Wang CY. Macrophages: friend or foe in idiopathic pulmonary fibrosis? *Respir Res* (2018) 19(1):170. doi: 10.1186/s12931-018-0864-2
- Gibbons MA, MacKinnon AC, Ramachandran P, Dhaliwal K, Duffin R, Phythian-Adams AT, et al. Ly6Chi monocytes direct alternatively activated profibrotic macrophage regulation of lung fibrosis. *Am J Respir Crit Care Med* (2011) 184(5):569–81. doi: 10.1164/rccm.201010-1719OC
- Mostafaei S, Sayad B, Azar MEF, Doroudian M, Hadifar S, Behrouzi A, et al. The role of viral and bacterial infections in the pathogenesis of IPF: a systematic review and meta-analysis. *Respir Res* (2021) 22(1):53. doi: 10.1186/s12931-021-01650-x
- Sheng G, Chen P, Wei Y, Yue H, Chu J, Zhao J, et al. Viral infection increases the risk of idiopathic pulmonary fibrosis: A meta-analysis. *Chest* (2020) 157(5):1175–87. doi: 10.1016/j.chest.2019.10.032
- Desai O, Winkler J, Minasyan M, Herzog EL. The role of immune and inflammatory cells in idiopathic pulmonary fibrosis. *Front Med (Lausanne)* (2018) 5:43. doi: 10.3389/fmed.2018.00043
- Lo Re S, Lecocq M, Uwambayinema F, Yakoub Y, Delos M, Demoulin JB, et al. Platelet-derived growth factor-producing CD4+ Foxp3+ regulatory T lymphocytes promote lung fibrosis. *Am J Respir Crit Care Med* (2011) 184(11):1270–81. doi: 10.1164/rccm.201103-0516OC
- Hou Z, Ye Q, Qiu M, Hao Y, Han J, Zeng H. Increased activated regulatory T cells proportion correlate with the severity of idiopathic pulmonary fibrosis. *Respir Res* (2017) 18(1):170. doi: 10.1186/s12931-017-0653-3
- Peng X, Moore MW, Peng H, Sun H, Gan Y, Homer RJ, et al. CD4+CD25+FoxP3+ regulatory tregs inhibit fibrocyte recruitment and fibrosis via suppression of FGF-9 production in the TGF-beta1 exposed murine lung. *Front Pharmacol* (2014) 5:80. doi: 10.3389/fphar.2014.00080
- Hirahara K, Aoki A, Morimoto Y, Kiuchi M, Okano M, Nakayama T. The immunopathology of lung fibrosis: amphiregulin-producing pathogenic memory T helper-2 cells control the airway fibrotic responses by inducing eosinophils to secrete osteopontin. *Semin Immunopathol* (2019) 41(3):339–48. doi: 10.1007/s00281-019-00735-6
- Neumark N, Cosme CJr., Rose KA, Kaminski N. The idiopathic pulmonary fibrosis cell atlas. *Am J Physiol Lung Cell Mol Physiol* (2020) 319(6):L887–L93. doi: 10.1152/ajplung.00451.2020
- Bryant AJ, Shenoy V, Fu C, Marek G, Lorentsen KJ, Herzog EL, et al. Myeloid-derived suppressor cells are necessary for development of pulmonary hypertension. *Am J Respir Cell Mol Biol* (2018) 58(2):170–80. doi: 10.1165/rcmb.2017-0214OC
- Fernandez IE, Greiffo FR, Frankenberger M, Bandres J, Heinzelmann K, Neurohr C, et al. Peripheral blood myeloid-derived suppressor cells reflect disease status in idiopathic pulmonary fibrosis. *Eur Respir J* (2016) 48(4):1171–83. doi: 10.1183/13993003.01826-2015
- Casanova NG, Zhou T, Gonzalez-Garay ML, Lussier YA, Sweiss N, Ma S-F, et al. MicroRNA and protein-coding gene expression analysis in idiopathic pulmonary fibrosis yields novel biomarker signatures associated to survival. *Transl Res* (2021) 228:1–12. doi: 10.1016/j.trsl.2020.07.009
- Li X, Cai H, Cai Y, Zhang Q, Ding Y, Zhuang Q. Investigation of a hypoxia-Immune-Related microenvironment gene signature and prediction model for idiopathic pulmonary fibrosis. *Front Immunol* (2021) 12:629854. doi: 10.3389/fimmu.2021.629854
- Spencer HL, Jover E, Cathery W, Avolio E, Rodriguez-Arabaolaza I, Thomas AC, et al. Role of TPBG (Trophoblast glycoprotein) antigen in human pericyte migratory and angiogenic activity. *Arterioscler Thromb Vasc Biol* (2019) 39(6):1113–24. doi: 10.1161/ATVBAHA.119.312665
- Xylinas E, Cha EK, Khani F, Kluth LA, Rieken M, Volkmer BG, et al. Association of oncofetal protein expression with clinical outcomes in patients with

urothelial carcinoma of the bladder. *J Urol* (2014) 191(3):830–41. doi: 10.1016/j.juro.2013.08.048

44. Su H, Yu S, Sun F, Lin D, Liu P, Zhao L. LINC00342 induces metastasis of lung adenocarcinoma by targeting miR-15b/TPBG. *Acta Biochim Pol* (2022) 69(2):291–7. doi: 10.18388/abp.2020\_5697

45. Chen H, Liu H, Qing G. Targeting oncogenic myc as a strategy for cancer treatment. *Signal Transduct Target Ther* (2018) 3:5. doi: 10.1038/s41392-018-0008-7

46. Qin H, Tang Y, Mao Y, Zhou X, Xu T, Liu W, et al. C-MYC induces idiopathic pulmonary fibrosis via modulation of miR-9-5p-mediated TBPL1. *Cell Signal* (2022) 93:110274. doi: 10.1016/j.cellsig.2022.110274

47. Yin X, Choudhury M, Kang JH, Schaeffbauer KJ, Jung MY, Andrianifahanana M, et al. Hexokinase 2 couples glycolysis with the profibrotic actions of TGF-beta. *Sci Signal* (2019) 12(612):eaax4067. doi: 10.1126/scisignal.aax4067

48. Kimura I, Ichimura A, Ohue-Kitano R, Igarashi M. Free fatty acid receptors in health and disease. *Physiol Rev* (2020) 100(1):171–210. doi: 10.1152/physrev.00041.2018

49. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, et al. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* (2009) 461(7268):1282–6. doi: 10.1038/nature08530

50. Sencio V, Barthelemy A, Tavares LP, Machado MG, Soulard D, Cuinat C, et al. Gut dysbiosis during influenza contributes to pulmonary pneumococcal superinfection through altered short-chain fatty acid production. *Cell Rep* (2020) 30(9):2934–47 e6. doi: 10.1016/j.celrep.2020.02.013

51. Gui X, Qiu X, Tian Y, Xie M, Li H, Gao Y, et al. Prognostic value of IFN-gamma, sCD163, CCL2 and CXCL10 involved in acute exacerbation of idiopathic pulmonary fibrosis. *Int Immunopharmacol* (2019) 70:208–15. doi: 10.1016/j.intimp.2019.02.039