



OPEN ACCESS

EDITED BY

Guochang Hu,
University of Illinois at Chicago,
United States

REVIEWED BY

David Dolivo,
Northwestern University, United States
Xuxu Gou Baylor College of Medicine,
United States

*CORRESPONDENCE

Nanshan Zhong
nanshan@vip.163.com
Jianxing He
drjianxing.he@gmail.com
Xiao Xiao Tang
tangxiaoxiao@gird.cn

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

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

RECEIVED 18 July 2022

ACCEPTED 31 August 2022

PUBLISHED 20 September 2022

CITATION

Huang G, Xu X, Ju C, Zhong N,
He J and Tang XX (2022) Identification
and validation of autophagy-related
gene expression for predicting
prognosis in patients with idiopathic
pulmonary fibrosis.
Front. Immunol. 13:997138.
doi: 10.3389/fimmu.2022.997138

COPYRIGHT

© 2022 Huang, Xu, Ju, Zhong, He and
Tang. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Identification and validation of autophagy-related gene expression for predicting prognosis in patients with idiopathic pulmonary fibrosis

Guichuan Huang^{1†}, Xin Xu^{1†}, Chunrong Ju^{1†},
Nanshan Zhong^{1,2*}, Jianxing He^{1*} and Xiao Xiao Tang^{1,2*}

¹State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, ²Guangzhou Laboratory, Guangzhou, China

Background: Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, and fatal fibrotic pulmonary disease with unknown etiology. Owing to lack of reliable prognostic biomarkers and effective treatment measures, patients with IPF usually exhibit poor prognosis. The aim of this study is to establish a risk score prognostic model for predicting the prognosis of patients with IPF based on autophagy-related genes.

Methods: The GSE70866 dataset was obtained from the gene expression omnibus (GEO) database. The autophagy-related genes were collected from the Molecular Signatures Database (MSigDB). Gene enrichment analysis for differentially expressed genes (DEGs) was performed to explore the function of DEGs. Univariate, least absolute shrinkage and selection operator (LASSO), as well as multivariate Cox regression analyses were conducted to identify a multi-gene prognostic model. Receiver operating characteristic (ROC) curve was applied to assess the prediction accuracy of the model. The expression of genes screened from the prognostic model was validated in clinical samples and human lung fibroblasts by qPCR and western blot assays.

Results: Among the 514 autophagy-related genes, a total of 165 genes were identified as DEGs. These DEGs were enriched in autophagy-related processes and pathways. Based on the univariate, LASSO, and multivariate Cox regression analyses, two genes (MET and SH3BP4) were included for establishing the risk score prognostic model. According to the median value of the risk score, patients with IPF were stratified into high-risk and low-risk groups. Patients in high-risk group had shorter overall survival (OS) than low-risk group in both training and test cohorts. Multivariate regression analysis indicated that prognostic model can act as an independent prognostic indicator for IPF. ROC curve analysis confirmed the reliable predictive value of prognostic model. In the validation experiments, upregulated MET expression and

downregulated SH3BP4 expression were observed in IPF lung tissues and TGF- β 1-activated human lung fibroblasts, which is consistent with results from microarray data analysis.

Conclusion: These findings indicated that the risk score prognostic model based on two autophagy-related genes can effectively predict the prognosis of patients with IPF.

KEYWORDS

idiopathic pulmonary fibrosis, prognostic model, autophagy, MET, SH3BP4

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive and fatal interstitial lung disease with unknown etiology (1). It is characterized by repetitive epithelial cell injury, fibroblast activation and overwhelming extracellular matrix (ECM) deposition which ultimately cause progressive loss of lung function and even death owing to respiratory failure (1). In the USA, the annual occurrence rate of IPF was 6.8-8.8 per 100,000 population with narrow case definitions and 16.3-16.7 per 100,000 population with broad case definitions (2). The annual occurrence rate in Europe was 0.22-7.4 per 100,000 population (2). The medial survival after diagnosis is only 2-3 years and the 5-year survival rate is no more than 40% (3, 4). Due to the complex etiology and unclear pathogenesis, there is still a lack of effective drugs. Pirfenidone and nintedanib, the two FDA-approved drugs, can't stop disease progression or reduce mortality (5). Therefore, it is important to identify the pathogenesis of IPF, explore novel treatment strategies and develop prognosis model.

Autophagy is a multi-step dynamic process that regulated by autophagy-related genes. In this process, autophagosomes are generated by phagocytosis of unwanted organelles and cytoplasmic proteins in a double membraned-surrounded vesicle (6). Then, autophagosomes are fused with lysosomes to degrade the contents of vesicles (6). Dysregulation of autophagy is involved in various lung diseases, including pulmonary hypertension, asthma, chronic obstructive pulmonary disease, and pulmonary fibrosis (7). For instance, a study has shown that leucine-rich repeat kinase 2 (LRRK2) is conducive to alleviate pulmonary fibrosis *via* preventing alveolar type II epithelial dysfunction and regulating the innate immune responses (8). Kim et al. reported that interleukin-37 (IL-37) attenuates IPF by blocking the transforming growth factor- β 1 (TGF- β 1) pathway and enhancing autophagy in IPF fibroblast (9). Wan et al. found that the downregulation of thymocyte differentiation antigen-1 (Thy-1) and upregulation of integrin β 3 can lead to pulmonary

fibrosis *via* activating PI3K/AKT/mTOR pathway and inhibiting lung fibroblast autophagy (10). Nevertheless, the role of autophagy-related genes in the prognosis of IPF remains largely unclarified and awaits further study.

In the present study, the autophagy-associated differentially expressed genes (DEGs) between control samples and IPF samples were analyzed in GSE70866 dataset. Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses were performed for DEGs. Then, based on the univariate, least absolute shrinkage and selection operator (LASSO) as well as multivariate regression analyses, two autophagy genes were included to establish a risk score prognostic model in the training set. Finally, this risk score model was proved to be an independent and reliable prognostic factor in patients with IPF.

Materials and methods

Acquisition of dataset and autophagy-related genes

The microarray data and clinical information in GSE70866 dataset (GPL14550 platform) were downloaded from gene expression omnibus (GEO) database [20 normal bronchoalveolar lavage fluid (BALF) samples and 112 IPF BALF samples]. The diagnosis of IPF in the dataset was confirmed by a multidisciplinary board at each institution according to the American Thoracic Society/European Respiratory Society criteria. To obtain BALF, pre-warmed and sterile saline was instilled by 20ml aliquots with immediate aspiration by gentle suction after each aliquot. Additional information regarding the collected BALF samples can be seen in this article (11). The raw microarray data were pre-processed for quality control with the use of "limma" package, including background adjustment and normalization. The general information of 112 patients with IPF was presented in

Table 1. A total of 12 autophagy-related gene sets were downloaded from the Molecular Signatures Database (MSigDB) (version 7.5.1) (Supplementary Table 1). After deleting the overlapping genes, 504 autophagy-associated genes were included for analysis (Supplementary Table 2).

Identification of autophagy-associated DEGs

The autophagy-associated DEGs between normal samples and IPF samples were investigated using “limma” R package. A gene with $p < 0.05$ was considered as DEG. Then, based on the GO and KEGG analyses, the biological functions and mechanism pathways for these autophagy-associated DEGs were explored.

Construction of risk score prognostic model

A total of 112 IPF samples was randomly divided into the training set ($n=56$) and the test set ($n=56$) with the use of “caret” R package. First, the prognosis-related genes from autophagy-associated DEGs were identified utilizing a univariate Cox regression analysis in the training set. Then, in order to avoid overfitting, we adopted the LASSO regression analysis to obtain the crucial autophagy-associated DEGs. Finally, a multivariate Cox regression analysis was conducted to select the autophagy-associated genes for establishing a risk score prognostic model. The formula of risk score model was presented as follows: Risk score = [(expression value of gene 1 $\times \beta_1$) + (expression value of gene 2 $\times \beta_2$) + ... + (expression value of gene $n \times \beta_n$)], where β is the corresponding gene’s regression coefficient. The risk score of each sample was calculated according to the formula. Samples

were stratified into the high-risk group and low-risk group on the basis of the median value of risk score. Kaplan-Meier analysis and log-rank test were performed to compare the survival differences between the two risk groups using “survival” R package. Receiver operating characteristic (ROC) curve was conducted to assess the model’s prediction accuracy using “survivalROC” R package. Cox regression analysis, including univariate and multivariate, was performed to evaluate whether the risk score model is an independent factor in IPF.

Pre and post risk score prognostic model comparison for principal component analysis

First, based on all autophagy-associated genes, PCA was performed to explore the sample distribution between two risk groups in the training set. Then, based on the two genes from risk score model, PCA was performed again. Finally, the “ggplot2” R package was employed to visualize the results.

The relationship between risk scores and clinical parameters

The relationship between risk scores and clinical parameters was explored, including age, gender, gender-age-physiology (GAP) index. GAP index is a staging system for patients with IPF and can be calculated using gender (G), age (A), and two lung physiology variables (P), including forced vital capacity (FVC) and diffusing capacity for carbon monoxide (DLCO) (12). GAP index may be used as a simple and quick approach for assessing risk in patients with IPF (12). IPF samples were divided into different groups according to the clinical parameters and the difference of risk scores among these groups was compared.

TABLE 1 Baseline characteristic of patients with IPF in training and test set.

Characteristic	Total (n=112)	Training set (n=56)	Test set (n=56)	P value
Age (Mean \pm SD)	67.97 \pm 10.06	69.54 \pm 10.36	66.41 \pm 9.61	0.812
Age, n (%)				
<70	56 (50.00)	25 (44.64)	31 (53.6)	
\geq 70	56 (50.00)	31 (55.36)	25 (44.64)	0.257
Sex, n (%)				
Male	93 (83.04)	44 (78.57)	49 (87.50)	
Female	19 (16.96)	12 (21.43)	7 (12.50)	0.208
GAP index (Mean \pm SD)	4.54 \pm 1.73	4.64 \pm 1.88	4.43 \pm 1.58	0.535
Status, n (%)				
Alive	36 (32.14)	19 (33.93)	17 (30.36)	
Dead	76 (67.86)	37 (66.07)	39 (69.64)	0.686

SD, Standard deviation; GAP, Gender-age-physiology.

Immune cell infiltration and immune-related function analyses in the two risk groups

The 22 kinds of immune cells in two risk groups were determined using CIBERSORT. CIBERSORT is a deconvolution method that assesses the immune cell composition of tissue from their gene expression profile (13). It applies linear support vector regression (SVR) (a machine learning approach) to deconvolute a mixture of gene expression. It has been shown that the results are correlated well with flow cytometric analysis. Therefore, CIBERSORT is also referred as “digital cytometry” (14). In addition, the 13 kinds of immune-related function were explored in the two risk groups.

Gene set variation analysis

GSVA is a nonparametric and unsupervised analysis method to condense information from gene expression profiles into a pathway summary (15). To investigate the biological process between the two risk groups, GSVA was performed with the use of “GSVA” R package. $P < 0.05$ was considered as statistically significant. The gene set of “c2.cp.kegg.v7.5.1.symbols”, downloaded from the MSigDB, was used as a reference.

Validation of the risk score model in the test set

According to the median value of risk score from the training set, the patients with IPF in test set were split into the high-risk and low-risk groups. The OS in two groups were compared using Kaplan-Meier analysis and log-rank test. ROC curve was conducted to evaluate the prediction accuracy of the model in the test set.

Preliminary experimental validation of the genes from risk score model

Lung tissue samples were obtained from six patients with IPF and six healthy controls at the First Affiliated Hospital of Guangzhou Medical University. This study was approved by the ethics committee of the First Affiliated Hospital of Guangzhou Medical University and was carried out in accordance with the Declaration of Helsinki.

Human lung fibroblasts were purchased from the ATCC. Cells were cultured in DMEM medium supplemented with 10% fetal bovine serum, 100U/ml penicillin, and 100mg/ml streptomycin (Gibco) at 37°C in a 5% carbon dioxide atmosphere. The fibroblasts were stimulated with 10ng/ml TGF-

$\beta 1$ for 48h to induce them differentiate into myofibroblasts. Then, the total RNA and protein were collected for further analysis.

qPCR

To detect the mRNA levels, total RNA from cells was obtained using NucleoZOL reagent (Macherey-nagel GmbH & Co. Kg, Germany). RNA concentration was determined with a NanoDrop 2000 micro-spectrophotometer (Thermo Fisher Scientific, USA). Then, total RNA was reverse-transcribed into complementary DNA (cDNA) using Hifair[®] III 1st Strand cDNA Synthesis SuperMix (Yeasen Biotechnology, China). Subsequently, the cDNA was amplified by SYBR Green Master Mix (Yeasen Biotechnology, China). The relative expression levels of mRNA were normalized to the levels of GAPDH and calculated by the $2^{-\Delta\Delta CT}$ method.

The primers sequences were as follows:

MET, forward, 5'-AGCGTCAACAGAGGGACCT-3', reverse, 5'-GCAGTGAACCTCCGACTGTATG-3'; SH3BP4, forward, 5'-ACCCTGATTGACCTGAGCGA-3', reverse, 5'-GGGGTTGTCTACGAGCAAGG-3'.

Western blot

Tissues from explanted IPF lungs or healthy donor lungs were collected and stored in liquid nitrogen before use. For protein extraction, tissues or cells were homogenized in ice-cold RIPA lysis buffer supplemented with phenylmethylsulfonyl fluoride (Biosharp, China) and phosphatase inhibitor cocktail (Sigma, USA). After centrifuging at 14,000 xg for 30 min at 4 °C, the supernatant was collected as total protein and the protein concentration was determined using BCA protein assay kit (Thermo Fisher Scientific, USA). Equal amounts of protein (20 μ g) were separated by 10% SDS-PAGE and transferred to PVDF membranes. After blocking with 5% non-fat milk at room temperature for 1h, the membranes were soaked in primary antibodies solutions at 4°C overnight. On the next day, membranes were washed with TBST for three times, then incubated with secondary antibodies at room temperature for 1h. Finally, the protein bands were visualized *via* an electrochemiluminescence reagent (Thermo Fisher Scientific, USA). The images were analyzed by Image J software. The following primary antibodies were utilized: anti-MET (1:1000, 25869-1-AP, Proteintech), anti-SH3BP4 (1:200, sc-393730, Santa Cruz).

Statistical analysis

The statistical analysis was implemented by R software (version 4.1.3). Gene expression in two groups was compared

by Wilcoxon test. Univariate Cox, LASSO and multivariate Cox regression analyses were performed to identify the prognosis-associated genes. The Kaplan-Meier analysis combined with a log-rank test was used to explore the differences in OS between two groups. ROC analysis was conducted to assess accuracy of the risk score prognostic model.

Raw data from qPCR and western blot analysis were presented as “mean ± standard error of mean (SEM)” and were further compared by Student’s t test. GraphPad software (version 8.0) was used to visualize the statistical results. $p < 0.05$ was considered as statistically significant.

Results

Identification of autophagy-related DEGs and function enrichment analysis in IPF

The overall flow chart for this study was presented in **Figure 1**. A total of 504 autophagy-related genes were collected from the MSigDB database (**Supplementary Table 2**). We explored the expression of autophagy-related genes between the 20 normal samples and 112 IPF samples. Among the 504 autophagy-related genes, 165 genes with $p < 0.05$ were considered as DEGs (**Supplementary Table 3**). Of the 165 DEGs, 113 genes were downregulated and 52 genes were upregulated in IPF.

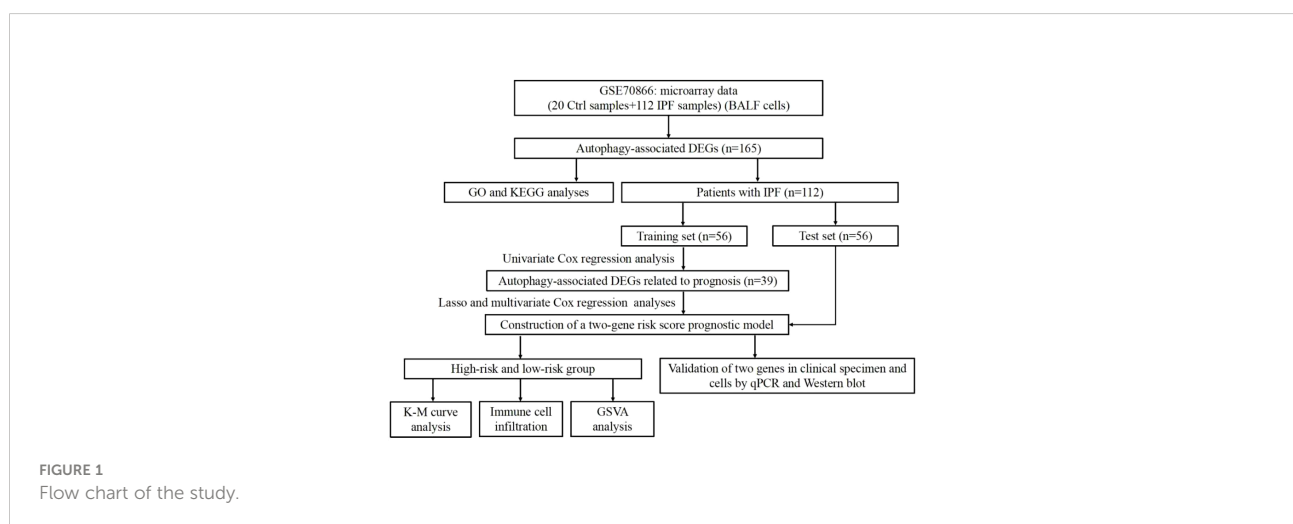
To further understand the function of these DEGs in IPF, we conducted GO and KEGG enrichment analyses. Three categories, including biological process (BP), cellular component (CC) and molecular functions (MF), were presented to describe the GO analysis. With respect to BP, DEGs were mainly enriched in autophagy-related biological activities (**Figure 2A**). Regarding CC, the top three items were autophagosome, vacuolar membrane, and autophagosome membrane, which were associated with autophagy (**Figure 2A**). In terms of MF, DEGs were closely related to ubiquitin protein

ligase which can degrade the proteins (**Figure 2A**). For the KEGG pathways, DEGs were also enriched in autophagy-related pathways, such as mTOR, JAK-STAT, and autophagy signaling pathways (**Figure 2B**).

Establishment of an autophagy-related gene risk score model in the training set

IPF samples ($n=112$) were randomly divided into the training set ($n=56$) and test set ($n=56$) (**Table 1**). First, 165 autophagy-related DEGs were included for a univariate Cox regression analysis. The results showed that 39 autophagy-related genes were associated with prognosis of patients with IPF ($p < 0.05$) (**Supplementary Table 4**). LASSO and multivariate Cox regression analyses were performed to select the key genes from the above 39 genes for construction of a risk score prognostic model. Finally, two genes were identified to establish the risk score prognostic model using the following formula: Risk score = $MET \times 0.545 + SH3BP4 \times (-0.461)$. In IPF, MET is a risk factor with $HR > 1$, whereas SH3BP4 is a protective factor for $HR < 1$ (**Figure 3** and **Table 2**). In addition, upregulated MET expression and downregulated SH3BP4 expression were found in IPF group as compared with the control group (**Figure 4**).

Based on the formula of risk score, patients were stratified into high-risk and low-risk groups with a median value of risk score as a cut-off point. First, we explored the distribution of patients using PCA analysis. As shown in **Figure 5**, the genes from risk score model could distinguish IPF from different risk groups. In order to assess the performance of risk score model in predicting the prognosis of patients with IPF, Kaplan-Meier curves were conducted. The results demonstrated that patients in high-risk group had a shorter OS as compared to the low-risk group ($p < 0.001$) (**Figure 6D**). The distribution of risk score in different risk groups was displayed in **Figure 6A**. Survival status of each patient was shown in **Figure 6B**. The heatmap presented



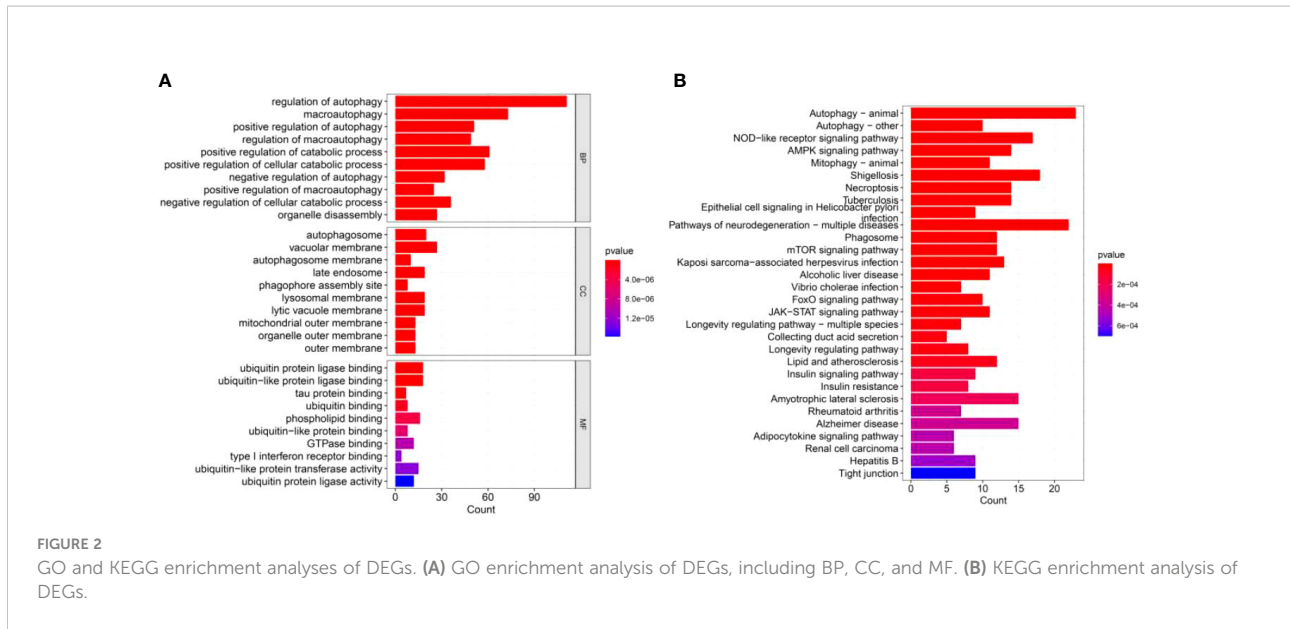


FIGURE 2
GO and KEGG enrichment analyses of DEGs. **(A)** GO enrichment analysis of DEGs, including BP, CC, and MF. **(B)** KEGG enrichment analysis of DEGs.

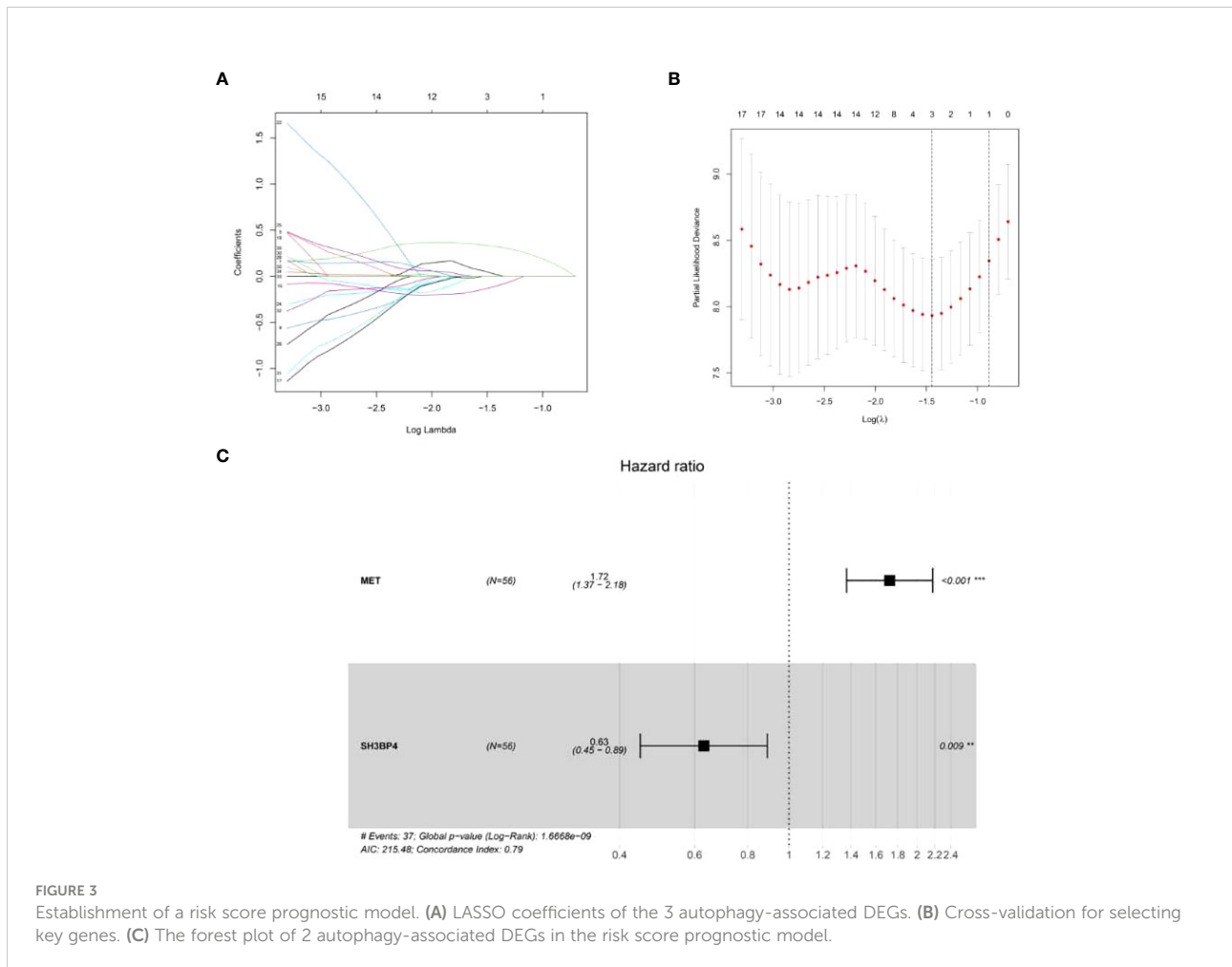


FIGURE 3
Establishment of a risk score prognostic model. **(A)** LASSO coefficients of the 3 autophagy-associated DEGs. **(B)** Cross-validation for selecting key genes. **(C)** The forest plot of 2 autophagy-associated DEGs in the risk score prognostic model.

TABLE 2 Details of the two genes from the risk score prognostic model.

Gene	Location	Coefficient	HR	P value
MET	chr7:116,672,196-116,798,386	0.545	1.724	<0.001
SH3BP4	chr2:234,952,017-235,055,714	-0.461	0.631	0.009

HR, hazard ratio.

the expression profiles of the two genes in the high-risk and low-risk groups (Figure 6C).

The risk score serves as an independent prognostic indicator

In order to determine if the risk score prognostic model is an independent prognostic factor for IPF, univariate and

multivariate regression analyses were performed. We integrated the risk score and clinical parameters (including age, sex, and GAP index) for analysis. The univariate regression analysis showed that GAP index and risk score were correlated to prognosis (Figure 7A). In the multivariate regression analysis, risk score as well as GAP index was proved to be an independent prognostic indicator (Figure 7B). These findings demonstrated that risk score prognostic model is reliable in forecasting the survival of patients with IPF.

We further performed ROC analysis to assess the risk score. The area under the ROC curve (AUC) for risk score at one, three, and five years was 0.889, 0.816, and 0.725, respectively (Figure 8A). In addition, we found that the AUC value for risk score at one year (0.889) was higher than age (0.497), sex (0.550) and GAP index (0.710) (Figure 8B). Taken together, this risk score model had a good prediction accuracy.

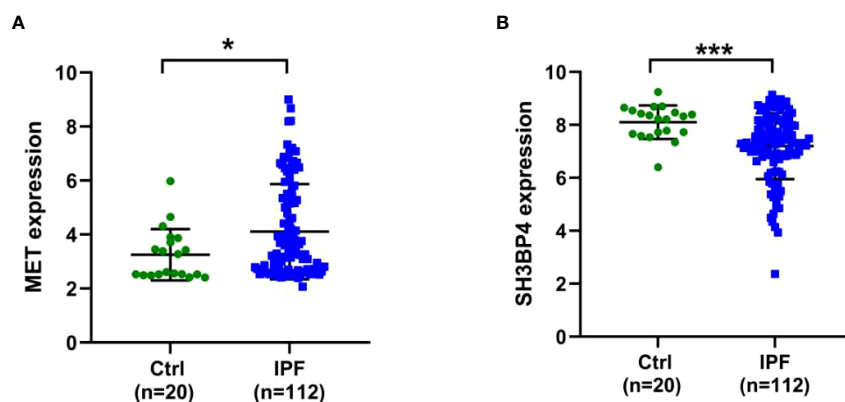


FIGURE 4 Expression levels of two genes in Ctrl and IPF groups. (A) MET. (B) SH3BP4. * $p < 0.05$; *** $p < 0.001$.

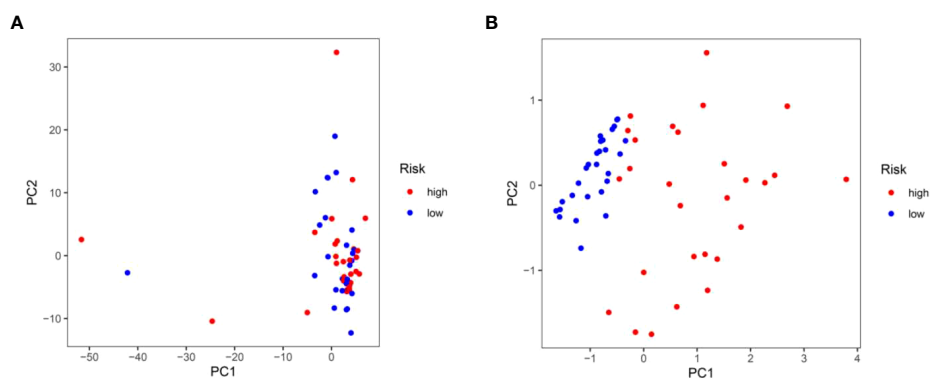
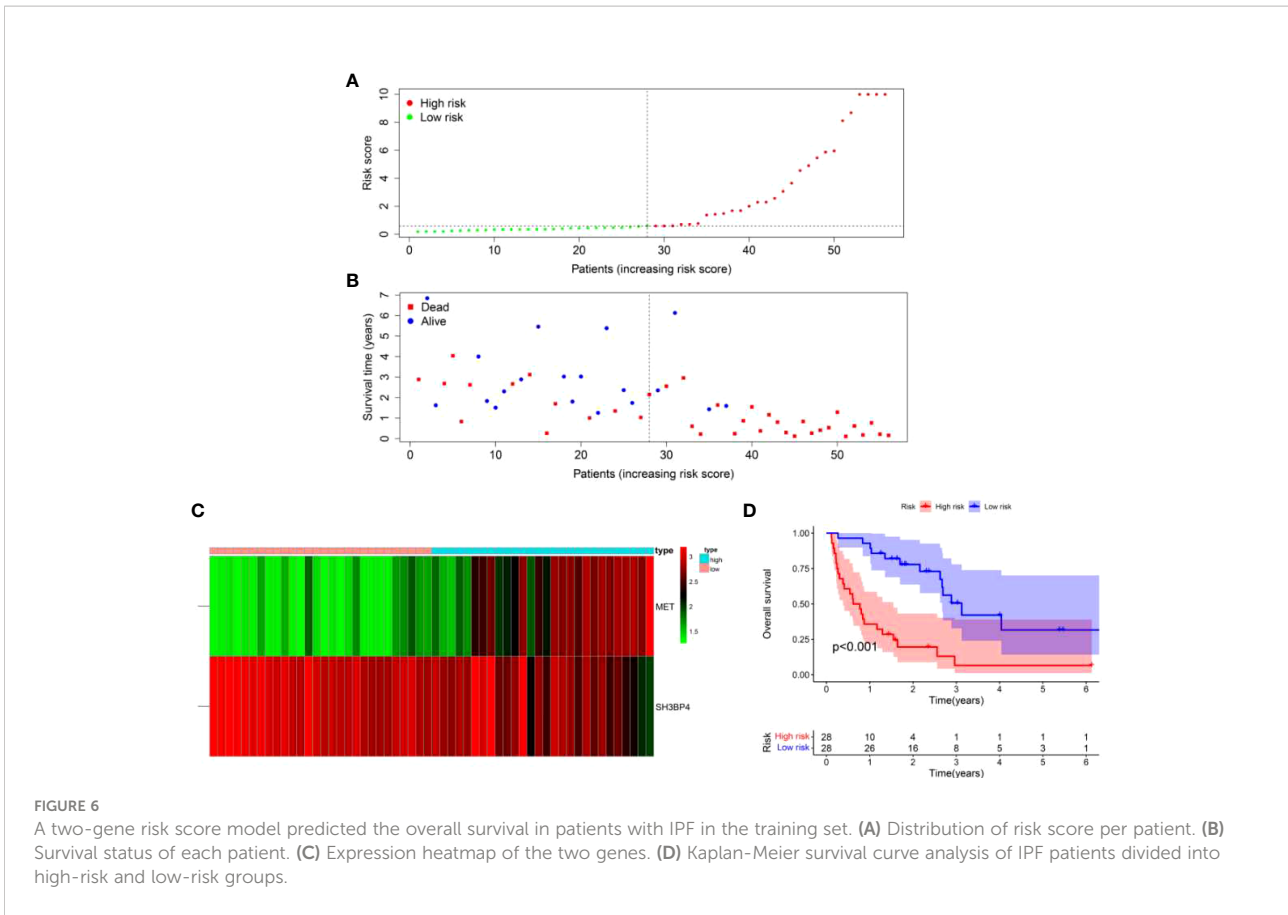


FIGURE 5 Principal component analysis. (A) Principal component analysis based upon all autophagy-associated genes in the training test. (B) Principal component analysis based upon two autophagy-associated genes from risk score prognostic model.

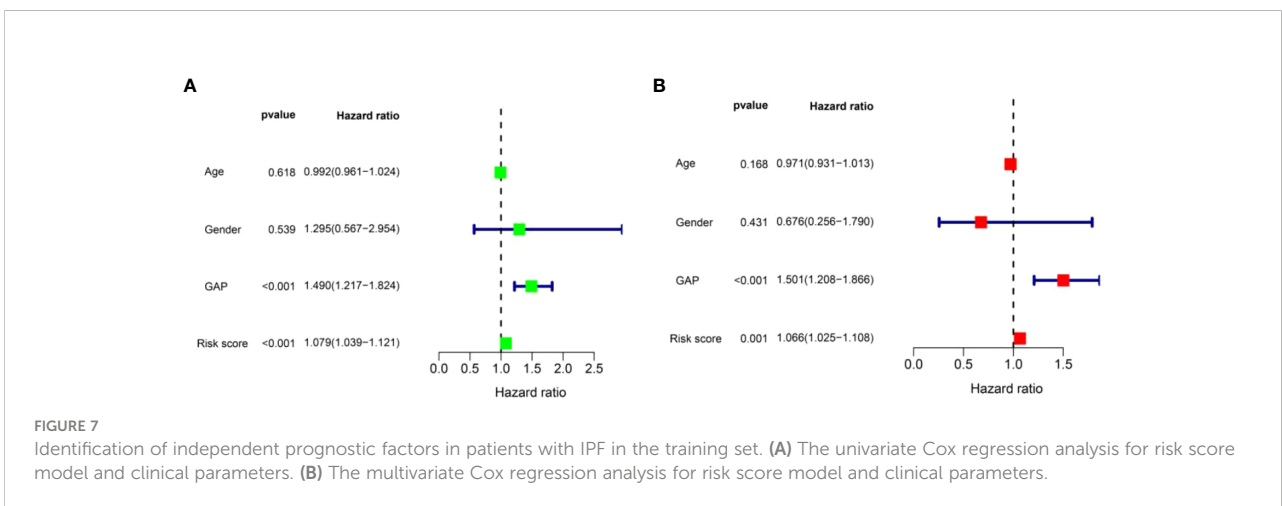


The relationship between risk scores and clinical features

To explore the association between risk scores and clinical features, we analyzed the distribution of risk scores in age, sex and GAP index. There was no statistical difference in risk score associations with age and sex, while higher risk scores were related to high GAP index (Figure 9).

Comparison analysis of immune cells or immune functions between high-risk and low-risk groups

Studies have shown that BALF contains different kinds of blood cells which might affect the progression of pulmonary fibrosis (16). Therefore, we explored the infiltration level of immune cells in high-risk and low-risk groups. We found more



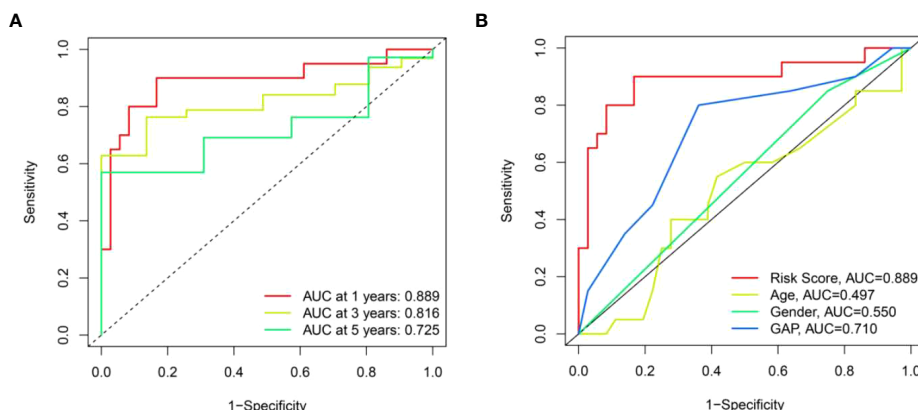


FIGURE 8
The prognostic value of risk score model in the training set. **(A)** ROC curves of risk score model at 1-, 3-, and 5-year overall survival. **(B)** ROC curves of clinical parameters and risk score model at 1-year overall survival.

monocytes and macrophages in the high-risk group than in the low-risk group (Figure 10A). On the contrary, low-risk group has more resting CD4⁺ T cells as compared with high-risk group (Figure 10A). In addition, high-risk group exhibits high scores of antigen-presenting cell (APC) co-stimulation, cytokine-cytokine receptor (CCR) interaction, parainflammation and type II IFN response (Figure 10B), while, human leukocyte antigen (HLA) was increased in low-risk group (Figure 10B).

GSVA

In order to explore the differences in pathway activity between the high-risk and low-risk groups in the training set, GSVA was performed. A total of 23 pathways were found to be statistically significant, such as pathways related to cancer, p53 signaling pathway, cytokine-cytokine receptor interaction, ECM receptor interaction, and Toll-like receptor signaling pathway

(Figure 11). These pathways were closely linked to the development of IPF (17–21).

Verification of the risk score model in the test set

To further validate the universality of the risk score model from the training set, the formula was applied in the test set. The risk score of each patient in the test set was calculated according to the formula from the training set. Subsequently, patients in the test set were divided into high-risk and low-risk groups based on the median value of risk score from the training set. The patients' risk curve and distribution of survival status in the test set were analyzed. We discovered that risk curve, survival status, and heat map were similar to those in the training set (Figures 12A–C). Likewise, we found that patients in high-risk group had poorer OS than low-risk group in the test set

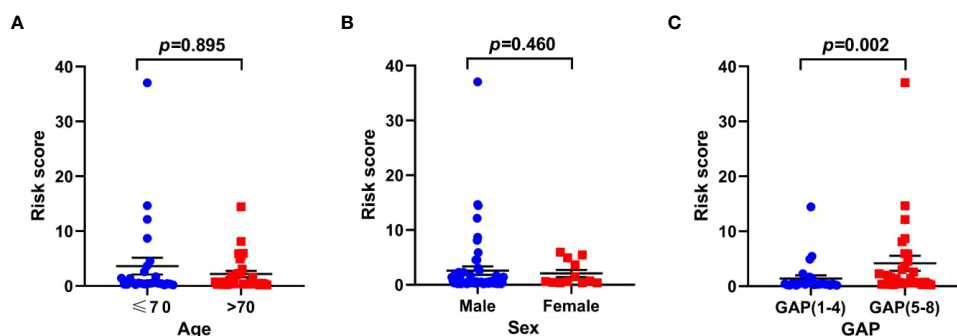


FIGURE 9
The relationship between risk scores and clinical parameters in the training set. **(A)** age; **(B)** sex; **(C)** GAP index.

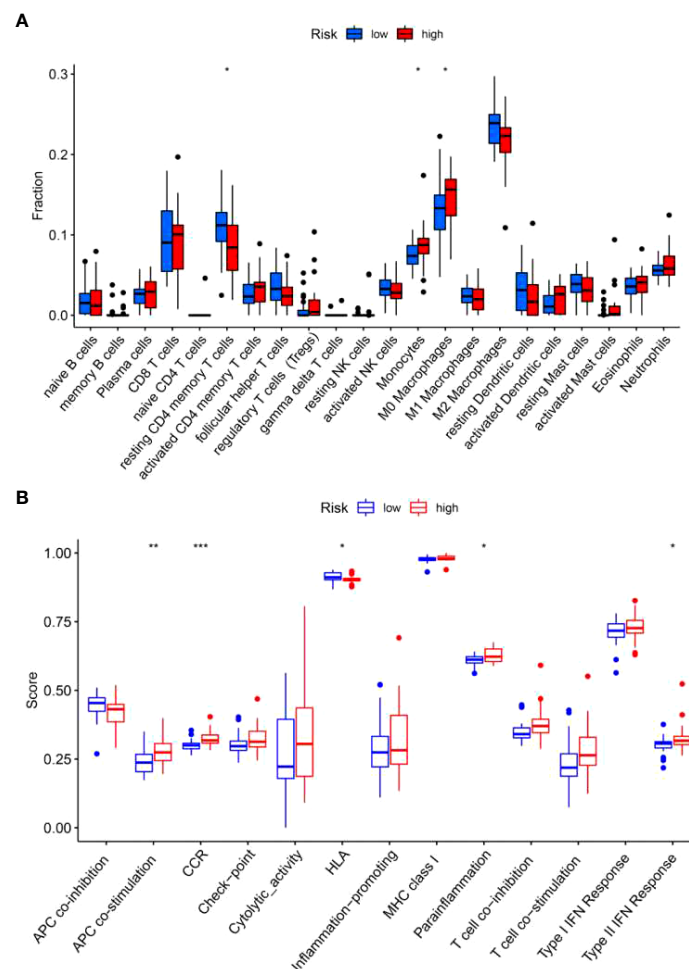


FIGURE 10

Immune cells and immune-related functions of the two risk groups in the training set. The proportion of 22 types of immune cells (A) and 13 immune-related functions (B) were analyzed in the high-risk and low-risk group. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

(Figure 12D). In addition, the AUC value was 0.706 in one year, 0.818 in three years and 0.819 in five years, respectively (Figure 13A). Furthermore, the AUC value of risk score at one year was better than age, sex, and GAP index (Figure 13B). These findings proved the universality and robustness of the risk score prognostic model.

Validation of model gene expression in clinical specimens and fibroblasts

To validate the gene expression, we performed qPCR and western blot analysis in clinical specimens and TGF- β 1-activated human lung fibroblasts. As shown in Figures 14A–C, the protein expression of MET was increased in IPF lung tissues as compared with normal lung tissues. On the contrary, the

protein expression of SH3BP4 in IPF lung tissues was decreased as compared to the normal lung tissues, although no statistical significance was observed. This could be due to the relative small sample size or different types of the samples as the microarray data were obtained from BALF cells, while our samples were lung homogenates. The differential expression of autophagy-related genes in the BALF may be coming from different cell types in the BALF between control subjects and IPF patients or from different gene expression profiles within the same types of cells, or combination of these factors, which needs further study.

Activated lung fibroblasts are the principal effector cells of progressive fibrotic process in IPF (22). TGF- β 1, a well-known pro-fibrotic factor, was used to activate fibroblasts. Similar as in IPF lung tissues, we found an upregulated MET expression and downregulated SH3BP4 expression in the TGF- β 1-activated fibroblasts (Figures 15A–E). These results are consistent with the microarray data analysis.

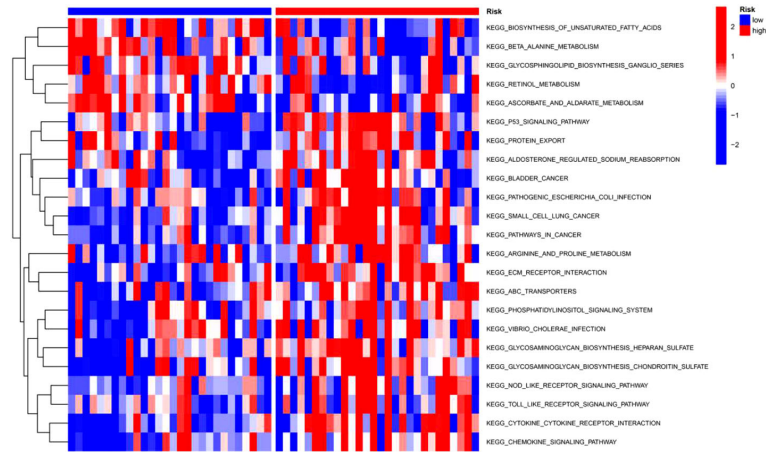


FIGURE 11
GSEA enrichment analysis between the two risk groups in the training set.

Discussion

IPF is a progressive lung disease with a poor prognosis. Molecular signatures of gene expression from lung tissue are associated with prognosis of IPF (23). Nevertheless, genomic

signatures are not widely applied in clinics as lung biopsy is an invasive and unpleasant operation. Bronchoalveolar lavage (BAL) is a medical procedure that sterile saline solution is injected into lung and then collected by a bronchoscope (24). BAL fluid exhibits biochemical changes due to lung diseases and

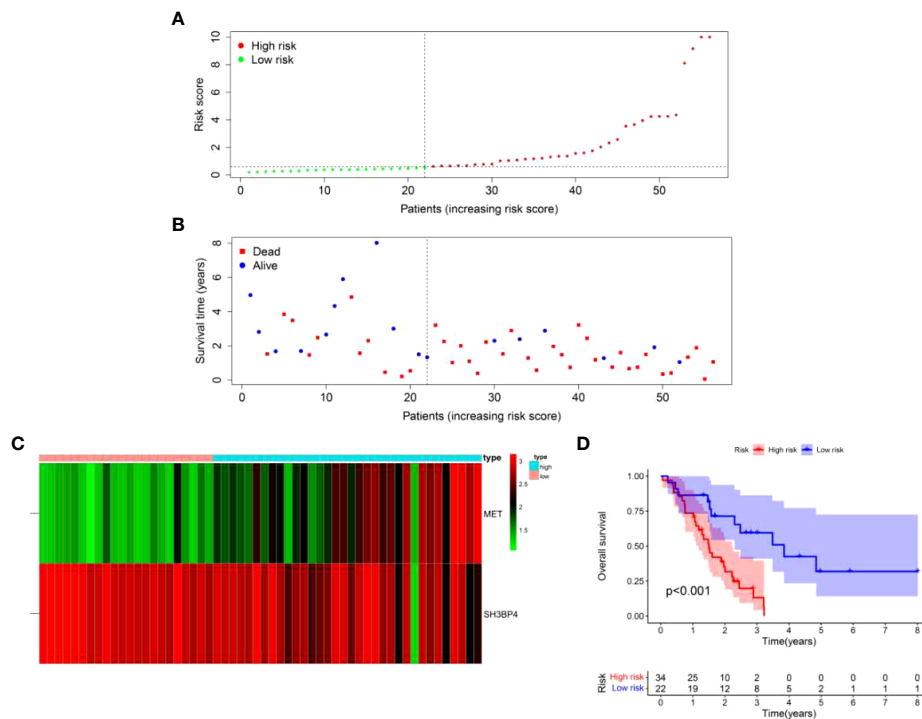


FIGURE 12
Validation of risk score model in the test set. (A) Distribution of risk score per patient. (B) Survival status of each patient. (C) An expression heatmap of the two genes. (D) Kaplan-Meier survival curve analysis of IPF patients divided into the high-risk and low-risk groups.

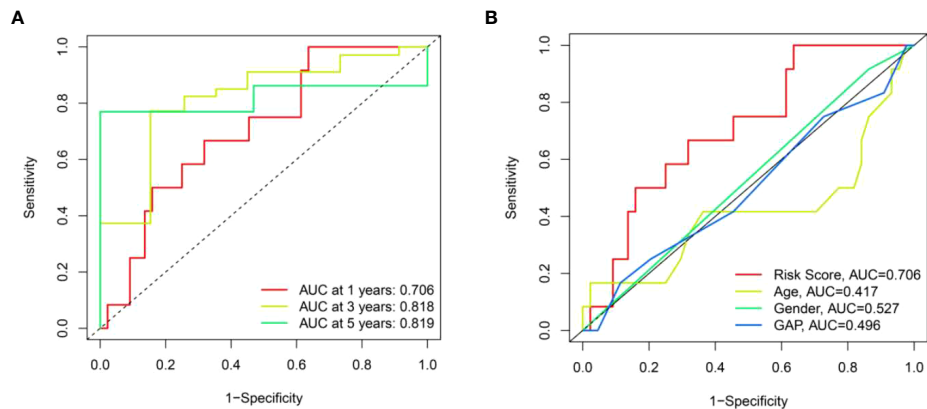


FIGURE 13
The prognostic value of risk score model in the test set. **(A)** ROC curves of risk score model at 1-, 3-, and 5-year overall survival. **(B)** ROC curves of clinical parameters and risk score model at 1-year overall survival.

external factors (24). The analysis of cells in BAL fluid may be conducive to diagnose and treat lung diseases, even assess the prognosis of patients with lung disease (25). Therefore, establishment of a multi-gene prognostic model based on the BALF cells is necessary to predict the prognosis of patients with IPF.

In this study, a total of 165 autophagy-related DEGs was determined between IPF BAL samples and normal BAL samples in the GSE70866 data set. Among the 165 DEGs, 39 DEGs were

further identified to be related to prognosis of patients with IPF in the training set. Subsequently, we constructed a two-gene risk score prognostic model based on the LASSO and multivariate Cox regression analyses and further validated it in the test set. Moreover, this risk score model could serve as an independent indicator for patients with IPF. Additionally, the ROC curve indicated that this risk score model had a reliable and effective prediction accuracy. Finally, we found an increased MET expression and decreased SH3BP4 expression in both IPF lung

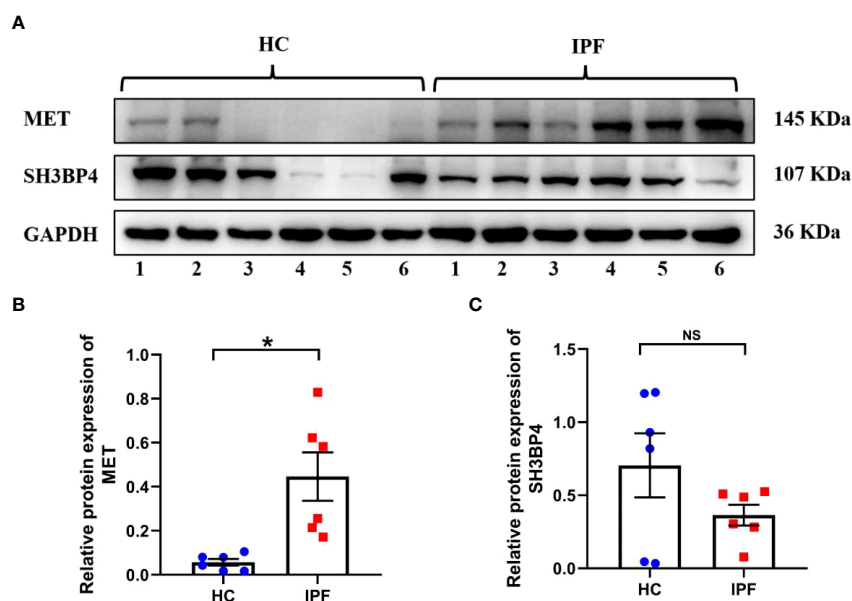


FIGURE 14
The expression of two model genes in HC and IPF lung tissues. **(A–C)** The protein expression of MET and SH3BP4 in healthy control and IPF lung tissues was detected by western blot assay. HC, healthy control. * $p < 0.05$. NS, not significant. $n = 6$.

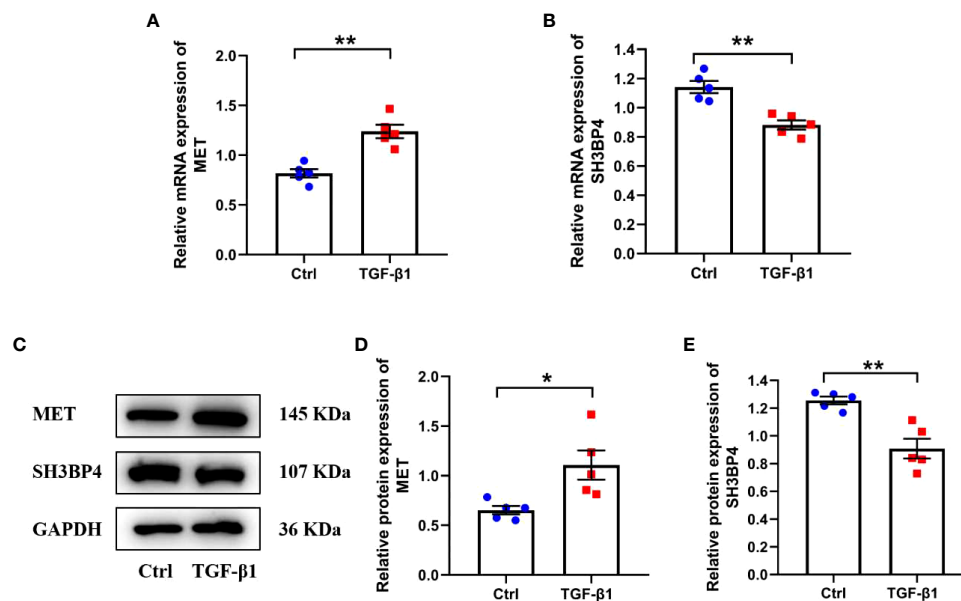


FIGURE 15

The expression of two model genes in human lung fibroblasts. Human lung fibroblasts were treated with 10ng/ml TGF- β 1 for 48h. (A, B) The mRNA expression of MET and SH3BP4 was detected by qPCR. (C–E) The protein expression of MET and SH3BP4 was detected by western blot assay. * $p < 0.05$; ** $p < 0.01$. $n = 5$.

tissues and TGF- β 1-activated human lung fibroblasts, consistent with the microarray results. These findings may aid clinicians in identifying high-risk patients and designing individualized treatment strategy for them.

The prognostic model in the present study was composed of two autophagy-related genes (MET and SH3BP4). Receptor tyrosine kinase MET, also known as c-MET, is a receptor of hepatocyte growth factor (HGF) (26). It has been reported that HGF/c-MET signaling pathway participates in multiple cellular processes, including cell survival, proliferation, motility, invasion and metastasis (27). In addition, MET is tightly linked to the process of autophagy (28–30). A number of studies have indicated that MET is closely involved in fibrotic diseases. Marquardt et al. reported that lack of c-MET can promote carbon tetrachloride-induced liver fibrosis in mice (31). Another study has shown an increased MET expression in lung fibroblasts from patients with pulmonary fibrosis as compared with lung fibroblasts from normal people (32). Moreover, MET has been implicated in driving profibrotic phenotypes and leading to pulmonary fibrosis (33, 34). Activation of lung fibroblast plays a major role in the pathogenesis of IPF (22). TGF- β 1 has been considered as the main growth factor involved in the differentiation of lung fibroblasts into myofibroblasts (35). In agreement with previous studies, we found that MET expression was upregulated in both IPF lung tissues and TGF- β 1-activated human lung fibroblasts, indicating that MET may promote the progression of IPF. SRC homology 3 domain-binding protein 4 (SH3BP4), also known as

transferrin receptor trafficking protein (TTP), was first discovered in human corneal fibroblasts (36). SH3BP4 affects autophagy process by negatively regulating Rag GTPase- mTOR complex 1 (mTORC1) signaling pathway (37). Besides, SH3BP4 negatively regulates Wnt signaling *via* modulating β -catenin's subcellular localization, thus suppressing tumor development (38). Kim et al. reported that SH3BP4 is a direct target gene of miR-125b and is negatively regulated by miR-125b (39). In another study, upregulated miR-125b was found in both human cardiac fibrosis and TGF- β -treated human cardiac fibroblasts (40). Consistent with the microarray data, we observed decreased expression of SH3BP4 in IPF lung tissues and TGF- β 1-activated human lung fibroblasts. We speculated that SH3BP4 is a negative regulator in the occurrence and progression of IPF and the underlying mechanism needs to be further elucidated.

Increasing evidences indicate that immune cells are linked to the development of IPF (1, 19, 41, 42). Kreuter et al. reported that increased monocyte count was related to elevated risks of IPF progression, hospitalization and mortality for patients with IPF (43). Another study also indicated that high absolute monocyte count is an IPF specific marker of mortality and poor outcomes (44). Macrophages play important roles in IPF. Single-cell transcriptomic analysis identified a distinct population of profibrotic alveolar macrophages exclusively in patients with pulmonary fibrosis (45). Also, accumulation of CD163⁺ and CD204⁺ macrophages in lung leads to worse clinical course in IPF patients (46). Consistent with these studies, we found a higher

amount of monocytes and macrophages in high-risk group than low-risk group, indicating that the elevated monocytes and macrophages may be related to the progression of IPF. We also found a reduced amount of resting memory CD4⁺ T cells in BALF of the high-risk group as compared with the low-risk group, indicating that resting memory CD4⁺ T cells may exhibit a protective role in IPF. Besides, significant differences in APC co-stimulation, CCR, HLA, parainflammation and type II IFN response were identified between the high-risk and low-risk groups. The exact role of these immune cells in the pathogenesis of IPF awaits further study.

There are limitations in the present study. First of all, the construction and validation of prognostic model were based on the retrospective data from GEO database, and the sample size in cohort was relatively small. Thus, a prospective study of large sample size is necessary to identify its clinical application. Second, clinical information was not complete in the data set, such as patients' therapy approaches, laboratory test, lung function data and so on, therefore, the significance of prognostic model was restricted. Last but not least, the association between risk score and immune activity needs to be further explored in basic experiments.

Conclusion

In summary, our study identified a novel risk score prognostic model of two autophagy-related genes, providing a new approach to predict the prognosis 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.

Ethics statement

This study was approved by the ethics committee of the First Affiliated Hospital of Guangzhou Medical University and was carried out in accordance with the Declaration of Helsinki. All subjects have provided written informed consent. IPF patients included in this study were according to the diagnostic criteria of 2018 ATS/ERS/JRS/ALAT Clinical Practice Guideline (47).

Author contributions

XT, NZ and GH conceived the study. XT designed and supervised the study, revised and edited the manuscript. GH

performed data analysis and drafted the manuscript. XX, CJ and JH contributed to clinical specimen collection for validation. All authors reviewed and approved the final version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (XT), the National High-Level Talents Program (XT), Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program (2017BT01S155), Open Project of State Key Laboratory of Respiratory Disease (SKLRD-OP-202109), Guangzhou Institute of Respiratory Health Open Project (XT) and Postdoctoral Startup Fund of Guangzhou City (GH).

Acknowledgments

We would like to express our gratitude to the GEO databases for providing us with this excellent public data.

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.2022.997138/full#supplementary-material>

References

- Heukels P, Moor CC, Von Der Thüsen 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
- Nalysnyk L, Cid-Ruzafa J, Rotella P, Esser D. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *Eur Respir Rev* (2012) 21 (126):355–61. doi: 10.1183/09059180.00002512
- King TE, Albera C, Bradford WZ, Costabel U, du Bois RM, Leff JA, et al. All-cause mortality rate in patients with idiopathic pulmonary fibrosis. implications for the design and execution of clinical trials. *Am J Respir Crit Care Med* (2014) 189 (7):825–31. doi: 10.1164/rccm.201311-1951OC
- Cai M, Zhu M, Ban C, Su J, Ye Q, Liu Y, et al. Clinical features and outcomes of 210 patients with idiopathic pulmonary fibrosis. *Chin Med J (Engl)* (2014) 127:1868–73. doi: 10.3760/cma.jissn.0366-6999.20132528
- Shenderov K, Collins SL, Powell JD, Horton MR. Immune dysregulation as a driver of idiopathic pulmonary fibrosis. *J Clin Invest* (2021) 131(2):e143226. doi: 10.1172/JCI143226
- Klionsky DJ. Autophagy: from phenomenology to molecular understanding in less than a decade. *Nat Rev Mol Cell Biol* (2007) 8(11):931–7. doi: 10.1038/nrm2245
- Racanello 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
- Tian Y, Lv J, Su Z, Wu T, Li X, Hu X, et al. LRRK2 plays essential roles in maintaining lung homeostasis and preventing the development of pulmonary fibrosis. *Proc Natl Acad Sci U.S.A.* (2021) 118(35):e2106685118. doi: 10.1073/pnas.2106685118
- Kim MS, Baek AR, Lee JH, Jang AS, Kim DJ, Chin SS. IL-37 attenuates lung fibrosis by inducing autophagy and regulating TGF- β 1 production in mice. *J Immunol* (2019) 203(8):2265–75. doi: 10.4049/jimmunol.1801515
- Wan H, Xie T, Xu Q, Hu X, Xing S, Yang H, et al. Thy-1 depletion and integrin β 3 upregulation-mediated PI3K-Akt-mTOR pathway activation inhibits lung fibroblast autophagy in lipopolysaccharide-induced pulmonary fibrosis. *Lab Invest* (2019) 99(11):1636–49. doi: 10.1038/s41374-019-0281-2
- Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. BAL cell gene expression is indicative of outcome and airway basal cell involvement in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2019) 199(5):622–30. doi: 10.1164/rccm.201712-2551OC
- Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* (2012) 156(10):684–91. doi: 10.7326/0003-4819-156-10-201205150-00004
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Alizadeh, robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* (2019) 37(7):773–82. doi: 10.1038/s41587-019-0114-2
- Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
- Meyer KC, Raghu G, Baughman RP, Brown KK, Costabel U, du Bois RM, et al. An official American thoracic society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am J Respir Crit Care Med* (2012) 185(9):1004–14. doi: 10.1164/rccm.201202-0320ST
- Jiang C, Liu G, Luckhardt T, Antony V, Zhou Y, Carter AB, et al. Serpine 1 induces alveolar type II cell senescence through activating p53-p21-Rb pathway in fibrotic lung disease. *Aging Cell* (2017) 16(5):1114–24. doi: 10.1111/acel.12643
- Shetty SK, Tiwari N, Marudamuthu AS, Puthusseri B, Bhandary YP, Fu J, et al. p53 and miR-34a feedback promotes lung epithelial injury and pulmonary fibrosis. *Am J Pathol* (2017) 187(5):1016–34. doi: 10.1016/j.ajpath.2016.12.020
- She YX, Yu QY, Tang XX. Role of interleukins in the pathogenesis of pulmonary fibrosis. *Cell Death Discov* (2021) 7(1):52. doi: 10.1038/s41420-021-00437-9
- 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
- Karamitsakos T, Woolard T, Bourros D, Tzouveleki A. Toll-like receptors in the pathogenesis of pulmonary fibrosis. *Eur J Pharmacol* (2017) 808:35–43. doi: 10.1016/j.ejphar.2016.06.045
- King TE, Pardo A, Selman M. Idiopathic pulmonary fibrosis. *Lancet* (2011) 378(9807):1949–61. doi: 10.1016/S0140-6736(11)60052-4
- Boon K, Bailey NW, Yang J, Steel MP, Groshong S, Kervitsky D, et al. Molecular phenotypes distinguish patients with relatively stable from progressive idiopathic pulmonary fibrosis (IPF). *PLoS One* (2009) 4(4):e5134. doi: 10.1371/journal.pone.0005134
- Landi C, Bargagli E, Carleo A, Bianchi L, Gagliardi A, Prasse A, et al. A system biology study of BALF from patients affected by idiopathic pulmonary fibrosis (IPF) and healthy controls. *Proteomics Clin Appl* (2014) 8(11-12):932–50. doi: 10.1002/prca.201400001
- Korthagen NM, van Moorsel CHM, Barlo NP, Ruven HJT, Kruit A, Heron M, et al. Serum and BALF YKL-40 levels are predictors of survival in idiopathic pulmonary fibrosis. *Respir Med* (2011) 105(1):106–13. doi: 10.1016/j.rmed.2010.09.012
- Gallo S, Sala V, Gatti S, Crepaldi T. Cellular and molecular mechanisms of HGF/Met in the cardiovascular system. *Clin Sci (Lond)* (2015) 129(12):1173–93. doi: 10.1042/CS20150502
- Noriega-Guerra H, Freitas VM. Extracellular matrix influencing HGF/c-MET signaling pathway: Impact on cancer progression. *Int J Mol Sci* (2018) 19 (11):3300. doi: 10.3390/ijms19113300
- Bell ES, Coelho PP, Park M. LC3C mediates selective autophagy of the MET RTK, inhibiting cancer cell invasion. *Autophagy* (2020) 16(5):959–61. doi: 10.1080/15548627.2020.1728099
- Liu Y, Liu JH, Chai K, Tashiro SI, Onodera S, Ikejima T. Inhibition of c-met promoted apoptosis, autophagy and loss of the mitochondrial transmembrane potential in oridonin-induced A549 lung cancer cells. *J Pharm Pharmacol* (2013) 65(11):1622–42. doi: 10.1111/jphp.12140
- Bell ES, Coelho PP, Ratcliffe CDH, Rajadurai CV, Peschard P, Vaillancourt R, et al. LC3C-mediated autophagy selectively regulates the MET RTK and HGF-stimulated migration and invasion. *Cell Rep* (2019) 29(12):4053–4068.e6. doi: 10.1016/j.celrep.2019.11.063
- Marquardt JU, Seo D, Gómez-Quiroz LE, Uchida K, Gillen MC, Kitade M, et al. Loss of c-met accelerates development of liver fibrosis in response to CCl(4) exposure through deregulation of multiple molecular pathways. *Biochim Biophys Acta* (2012) 1822(6):942–51. doi: 10.1016/j.bbdis.2012.02.012
- Ghatak S, Bogatkevich GS, Atelishvili I, Akter T, Feghali-Bostwick C, Hoffman S, et al. Overexpression of c-met and CD44v6 receptors contributes to autocrine TGF- β 1 signaling in interstitial lung disease. *J Biol Chem* (2014) 289 (11):7856–72. doi: 10.1074/jbc.M113.505065
- Wilson CL, Hung CF. Another weapon in the battle against idiopathic pulmonary fibrosis? *Am J Respir Cell Mol Biol* (2019) 60(4):386–7. doi: 10.1165/rccm.2018-0387ED
- Stella GM, Gentile A, Balderacchi A, Meloni F, Milan M, Benvenuti S. Ockham's razor for the MET-driven invasive growth linking idiopathic pulmonary fibrosis and cancer. *J Transl Med* (2016) 14(1):256. doi: 10.1186/s12967-016-1008-4
- Upagupta C, Shimbori C, Alsilmi R, Kolb M. Matrix abnormalities in pulmonary fibrosis. *Eur Respir Rev* (2018) 27(148):180033. doi: 10.1183/16000617.0033-2018
- Dunlevy JR, Berryhill BL, Vergnes JP, SundarRaj N, Hassell JR. Cloning, chromosomal localization, and characterization of cDNA from a novel gene, SH3BP4, expressed by human corneal fibroblasts. *Genomics* (1999) 62(3):519–24. doi: 10.1006/geno.1999.5994
- Kim YM, Stone M, Hwang TH, Kim YG, Dunlevy JR, Griffin TJ, et al. SH3BP4 is a negative regulator of amino acid-rag GTPase-mTORC1 signaling. *Mol Cell* (2012) 46(6):833–46. doi: 10.1016/j.molcel.2012.04.007
- Antas P, Novellasdemunt L, Kucharska A, Massie I, Carvalho J, Oukrif D, et al. SH3BP4 regulates intestinal stem cells and tumorigenesis by modulating β -catenin nuclear localization. *Cell Rep* (2019) 26(9):2266–2273.e4. doi: 10.1016/j.celrep.2019.01.110
- Kim KH, Lee TR, Cho EG. SH3BP4, a novel pigmentation gene, is inversely regulated by miR-125b and MITF. *Exp Mol Med* (2017) 49(8):e367. doi: 10.1038/emm.2017.115
- Nagpal V, Rai R, Place AT, Murphy SB, Verma SK, Ghosh AK, et al. MiR-125b is critical for fibroblast-to-myofibroblast transition and cardiac fibrosis. *Circulation* (2016) 133(3):291–301. doi: 10.1161/CIRCULATIONAHA.115.018174
- Deng KM, Yang XS, Luo Q, She YX, Yu QY, Tang XX. Deleterious role of Th9 cells in pulmonary fibrosis. *Cells* (2021) 10(11):3209. doi: 10.3390/cells10113209
- Serezani APM, Pascoalino BD, Bazzano JMR, Vowell KN, Tanjore H, Taylor CJ, et al. Multiplatform single-cell analysis identifies immune cell types enhanced in pulmonary fibrosis. *Am J Respir Cell Mol Biol* (2022) 67(1):50–60. doi: 10.1165/rccm.2021-0418OC

43. Kreuter M, Lee JS, Tzouveleakis A, Oldham JM, Molyneaux PL, Weycker D, et al. Maher, monocyte count as a prognostic biomarker in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* (2021) 204(1):74–81. doi: 10.1164/rccm.202003-0669OC
44. Scott MKD, Quinn K, Li Q, Carroll R, Warsinske H, Vallania F, et al. Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: a retrospective, multicentre cohort study. *Lancet Respir Med* (2019) 7(6):497–508. doi: 10.1016/S2213-2600(18)30508-3
45. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* (2019) 199(12):1517–36. doi: 10.1164/rccm.201712-2410OC
46. Nouno T, Okamoto M, Ohnishi K, Kaieda S, Tominaga M, Zaizen Y, et al. Elevation of pulmonary CD163 and CD204 macrophages is associated with the clinical course of idiopathic pulmonary fibrosis patients. *J Thorac Dis* (2019) 11(9):4005–17. doi: 10.21037/jtd.2019.09.03
47. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of idiopathic pulmonary fibrosis. an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* (2018) 198(5):e44–68. doi: 10.1164/rccm.201807-1255ST