



## OPEN ACCESS

## EDITED BY

Ashu Johri,  
Independent Researcher, New York,  
NY, United States

## REVIEWED BY

Xiaodong Han,  
Nanjing University, China  
R. Matthew Kottmann,  
University of Rochester, United States

## \*CORRESPONDENCE

Hao Zhang  
zhanghao@xzhmu.edu.cn

## SPECIALTY SECTION

This article was submitted to  
Cellular Endocrinology,  
a section of the journal  
Frontiers in Endocrinology

RECEIVED 23 July 2022

ACCEPTED 21 September 2022

PUBLISHED 04 October 2022

## CITATION

Lyu Y, Guo C and Zhang H (2022)  
Fatty acid metabolism-related genes in  
bronchoalveolar lavage fluid unveil  
prognostic and immune infiltration in  
idiopathic pulmonary fibrosis.  
*Front. Endocrinol.* 13:1001563.  
doi: 10.3389/fendo.2022.1001563

## COPYRIGHT

© 2022 Lyu, Guo and Zhang. 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.

# Fatty acid metabolism-related genes in bronchoalveolar lavage fluid unveil prognostic and immune infiltration in idiopathic pulmonary fibrosis

Yin Lyu<sup>1,2</sup>, Chen Guo<sup>1,2</sup> and Hao Zhang<sup>1,2\*</sup>

<sup>1</sup>Thoracic Surgery Laboratory, Xuzhou Medical University, Xuzhou, China, <sup>2</sup>Department of Thoracic Surgery, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

**Background:** Idiopathic pulmonary fibrosis (IPF) is a chronic and progressive condition with an unfavorable prognosis. A recent study has demonstrated that IPF patients exhibit characteristic alterations in the fatty acid metabolism in their lungs, suggesting an association with IPF pathogenesis. Therefore, in this study, we have explored whether the gene signature associated with fatty acid metabolism could be used as a reliable biological marker for predicting the survival of IPF patients.

**Methods:** Data on the fatty acid metabolism-related genes (FAMRGs) were extracted from databases like Kyoto Encyclopedia of Genes and Genomes (KEGG), Hallmark, and Reactome pathway. The GSE70866 dataset with information on IPF patients was retrieved from the Gene Expression Omnibus (GEO). Next, the consensus clustering method was used to identify novel molecular subgroups. Gene Set Enrichment Analysis (GSEA) was performed to understand the mechanisms involved. The Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) algorithm was used to evaluate the level of immune cell infiltration in the identified subgroups based on gene expression signatures of immune cells. Finally, the Least Absolute Shrinkage and Selection Operator (LASSO) regression and multivariate Cox regression analysis were performed to develop a prognostic risk model.

**Results:** The gene expression signature associated with fatty acid metabolism was used to create two subgroups with significantly different prognoses. GSEA reveals that immune-related pathways were significantly altered between the two subgroups, and the two subgroups had different metabolic characteristics. High infiltration of immune cells, mainly activated NK cells, monocytes, and activated mast cells, was observed in the subgroup with a poor prognosis. A risk model based on FAMRGs had an excellent ability to predict the prognosis of IPF. The nomogram constructed using the clinical features and the risk model could accurately predict the prognosis of IPF patients.

**Conclusion:** The fatty acid metabolism-related gene expression signature could be used as a potential biological marker for predicting clinical outcomes and the level of infiltration of immune cells. This could eventually enhance the accuracy of the treatment of IPF patients.

KEYWORDS

idiopathic pulmonary fibrosis, fatty acid metabolism, immune infiltration, biomarker, prognosis

## Introduction

Idiopathic pulmonary fibrosis (IPF), an interstitial lung disease (ILD), is a chronic, progressive, and age-related disease. IPF has high morbidity and a poor prognosis. However, the cause of IPF is unclear (1, 2). Alveolar epithelial cells' apoptosis, generation of pro-fibrotic factors, and stimulation of myofibroblasts and fibroblasts are involved in the pathogenesis of IPF (3–5). The Food and Drug Administration has approved two clinical drugs, nintedanib, and pirfenidone, for the treatment of IPF. However, these drugs have poor tolerability and cannot reverse the lung damage caused by IPF or reduce IPF-related death (6–8). IPF patients experience unfavorable clinical outcomes. The median survival rates are 50% and 20%, 3–5 years post IPF diagnosis, respectively, similar to lung cancer (9, 10). Consequently, IPF patients need personalized, targeted treatment, which requires developing a strategy for risk classification and the identification of prognostic genes.

Recently, substantial alterations in the metabolism of fatty acids (FAs) have been observed in IPF patients, which indicates that alterations in FA metabolism play an important role in the pathogenesis of IPF. FA metabolism pathways are complex, and disturbance in FA metabolism pathways in the lungs may result in the development of a pro-fibrotic phenotype in epithelial cells, macrophages, and myofibroblasts/fibroblasts. Sunaga et al. reported that treating cultured alveolar type (AT) 2 cells with palmitic acid induces cell apoptosis and transforming growth factor (TGF)- $\beta$ 1 expression in these cells (11). Mounting evidence has shown the involvement of the immune system in the progression of fibrosis and response to treatments. It has also gained extensive attention in interstitial lung disorders (12, 13). Immune cells like macrophages are predominantly distributed in the lung tissues and have an important role in the pathophysiology of pulmonary fibrosis. Macrophages can polarize into two phenotypes: M1 and M2. Once macrophages are activated, M2 macrophages release profibrotic cytokines, including TGF- $\beta$ 1, which stimulate the activity of fibroblasts and the deposition of extracellular matrix (14). There may be a strong association between the polarization of macrophages and

FA metabolism. The M2 phenotype of the macrophages depends on the transcription factor peroxisome proliferator-activated receptor (PPAR)- $\gamma$ , which plays an important role in FA metabolism as its natural ligand (15). In addition, FAs are responsible for enhancing the expression of FA receptor CD36, which induces the M2 phenotype by increasing FA uptake and self-reinforcing the profibrotic activation cycle (16). Dysfunctional epithelial cells and polarized macrophages produce numerous profibrotic cytokines, which induce the differentiation of fibroblasts into myofibroblasts. Myofibroblasts in IPF lungs produce an excessive extracellular matrix, disrupt the basement membrane, and are considered primary effector cells (17, 18). Reddy et al. have shown that nitrated FAs are PPAR- $\gamma$  receptor agonists and stimulate the dedifferentiation of myofibroblasts by inhibiting the impacts of TGF- $\beta$ 1 (19). Furthermore, FA and their derivatives modulate CCAAT enhancer-binding protein (C/EBP)  $\alpha$  and promote the dedifferentiation of myofibroblasts to lipofibroblasts (20). Therefore, targeting FA metabolism is an innovative approach for the treatment of IPF and is currently tested at the pre-clinical stage on pulmonary fibrosis models (21–23). However, the role of fatty acid metabolism-related genes (FAMRGs) in IPF pathogenesis is poorly understood, and the prognostic significance of these genes is yet to be elucidated.

Currently, lung biopsy is widely used for determining the molecular biomarkers for diagnosis and prognosis of IPF. However, lung biopsies are invasive, and hence their applications are limited. Prasse A et al. reported that the gene expression patterns of bronchoalveolar lavage (BAL) cells could predict mortality in IPF patients (24). Some studies showed that the levels of saturated long-chain FAs, including stearic acid, oleic acid, and palmitic acid, were low in bronchoalveolar lavage fluid (BALF) of IPF patients, in contrast, others showed that the levels of stearic and palmitic acid were high in BALF of IPF patients (25, 26). This may seem contradictory, but it gives us a clue that the changes in FA metabolism in the alveolar compartment of IPF patients could indicate disease progression.

In this study, we have analyzed FAMRGs in BAL cells to comprehensively examine the impacts of FA metabolism on the

survival of IPF patients. In addition, a risk score model based on FAMRGs was developed to examine the predictive ability of FAMRGs in IPF patients. These results may offer novel insights into understanding the molecular basis of IPF. It may also provide a novel approach to the targeted therapeutic for IPF and facilitate the development of personalized therapy for IPF patients.

## Materials and methods

### Datasets and samples

For this study, gene expression profiles and clinical data of 176 IPF patients were used. Gene expression of BAL cells from the GSE70866 dataset was retrieved from Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). The training cohort consisted of 112 IPF patients (from Siena, Italy, and Freiburg, Germany). The RNA microarray chips from IPF patients were performed using the Agilent-028004 SurePrint G3 Human GE 8x60K Microarray. The validation cohort consisted of 64 IPF patients (from Leuven, Belgium), and the RNA microarray chips from these patients were performed using Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray. The details of the patients included in this study are presented in **Table 1**. In addition, the raw data of IPF patients from the Freiburg and Siena groups were merged into the training cohort using the R package *inSilicoMerging* (27). The batch effect was removed using the Empirical Bayes method (28). The Uniform Manifold Approximation and Projection (UMAP) method was used to determine if the batch effect was eliminated (29). Data on 309 FAMRGs were obtained from the databases like Kyoto Encyclopedia of Genes and Genomes (KEGG), Hallmark, and Reactome.

### Identification of molecular subgroups

Univariate Cox regression analysis discovered 95 genes associated with the prognosis of IPF patients. Consensus clustering was performed based on the expression matrix of the 95 genes and was carried out using the R package *ConsensusClusterPlus* (30).

### Enrichment and immune analysis

Gene Set Enrichment Analysis (GSEA) was carried out using the same dataset to examine the differences across the clusters. Meanwhile, Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) (31) analysis was used to determine the proportions of 22 human immune cell subsets in the BALF of IPF patients based on the gene expression data.

### Construction and verification of the risk model

Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis was performed using R package *glmnet*, based on a 5-fold cross-validation in the training cohort to filter out FAMRGs associated with the survival of IPF patients. The lowest lambda value was determined as the optimal value. LASSO regression analysis identified genes related to the prognosis of IPF patients (survival time and status). These genes were used to generate a prognostic risk signature based on the regression coefficients. The following formula was used to determine the risk score of each patient within the training and validation cohorts:  $\text{Riskscore} = \sum \text{Exp (mRNA}_i) \times \text{Coefficient}$

TABLE 1 Patients' features in the training and validation cohorts.

Variables	Training Cohort	Validation Cohort	P-value
n	112	64	
Group, n (%)			< 0.001
Freiburg	62 (35.2%)	–	
Siena	50 (28.4%)	–	
Leuven	–	64 (36.4%)	
Sex, n (%)			0.726
Female	19 (10.8%)	13 (7.4%)	
Male	93 (52.8%)	51 (29%)	
Status, n (%)			< 0.001
Alive	36 (20.5%)	40 (22.7%)	
Dead	76 (43.2%)	24 (13.6%)	
Age, median (IQR)	69.5 (62, 76)	68.5 (63.75, 75)	0.920
Survival time (days), median (IQR)	569.5 (291, 961.25)	566.5 (346, 963.75)	0.625

IQR, Interquartile range.

(mRNAi). Patients were categorized into low- and high-risk groups based on the median value. The predictive ability of the model was assessed using the survival analysis and time-dependent receiver operating characteristic (ROC) curves.

## Statistical analysis

R package (version 4.0.2) was used to perform statistical analysis. The Kaplan-Meier survival analysis and the log-rank test were used for survival analyses. The predictive accuracy of the risk model was assessed by the time-dependent ROC using the R package survivalROC. The patients were classified based on age and gender for subgroup analysis. The student's t-test was used to compare the two groups, and the comparisons among multiple groups were carried out using one-way ANOVA. Statistical significance was defined as  $p < 0.05$ .

## Results

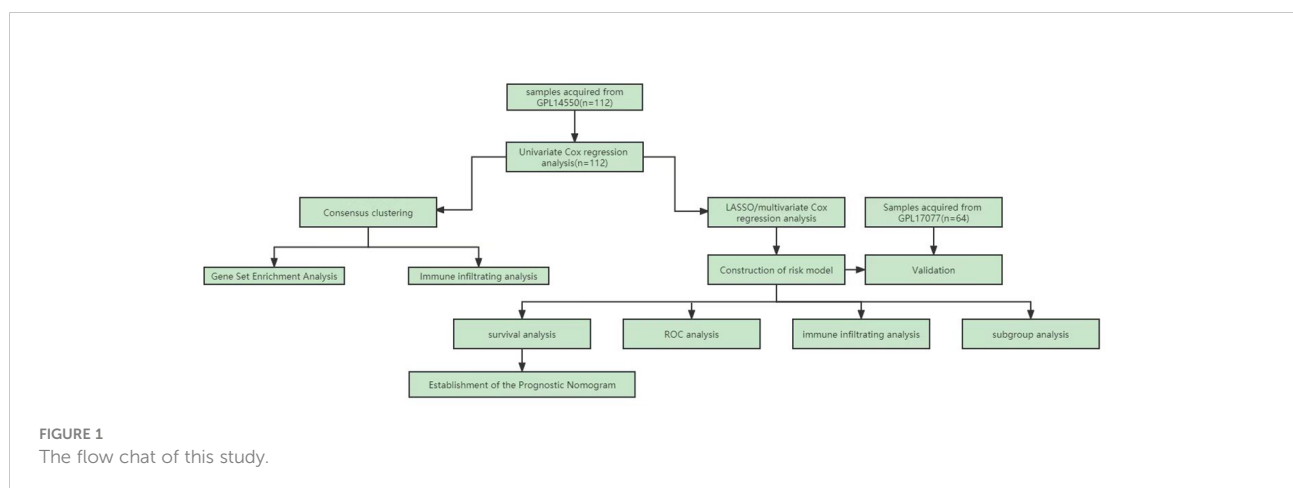
### Removal of batch effects in the training cohort

Figure 1 shows the flowchart of the workflow of the study. The empirical Bayes method was used to eliminate batch effects between the Freiburg and Siena groups in the training cohort. The boxplot shows that the sample distribution of each group differed significantly before the elimination of the batch effect, indicating the existence of the batch effect. Once the batch effect was removed, the data distribution of each group was consistent (Figure 2A). UMAP results show that the samples from each group clustered together before batch effect elimination, suggesting the presence of the batch effect. Once the batch effect was eliminated, the samples from each group clustered together and intertwined, suggesting that the batch effect was eliminated (Figure 2B).

### Molecular subtype identification prediction based on FAMRGs

The consensus clustering method was used to classify the IPF patients from the training cohort into distinct groups based on 95 prognostic genes obtained by univariate Cox analysis (Table S1).  $K = 2$  was identified as the optimum value for clustering stability (Figures 3A–D). Cluster 1 (C1) included 53 IPF patients, and cluster 2 (C2) had 59 IPF patients. The expression levels of FAMRG in the two clusters were visualized using the heatmap (Figure 3E), and the results revealed that C1 and C2 had significantly different gene expression patterns. Further, the patients in C2 exhibited a significantly higher ( $p < 0.0001$ ) overall survival (OS) compared to patients in C1 (Figure 3F). These results show that IPF patients can be classified into two distinct molecular subgroups based on FAMRGs, and each subgroup had a different OS. Hence, these results indicated that it was possible to create a prediction model based on FAMRGs.

GSEA was used to elucidate the possible mechanisms responsible for the difference between C1 and C2. As shown in Figure 4A, immune-related pathways, including the Nod-like receptor signaling pathway, chemokine signaling pathway, leukocyte transendothelial migrations, natural killer cell-mediated cytotoxicity, B cell receptor signaling pathway, Toll-like receptor signaling pathway, and metabolism-related pathways (i.e., arginine and proline metabolism, glycosaminoglycan biosynthesis heparan sulfate, glycerophospholipid metabolism, arachidonic acid metabolism, and histidine metabolism) were significantly ( $p < 0.05$ ) enriched by cluster 1, indicative of poor survival. Further, five metabolism-related pathways, including propanoate metabolism, butanoate metabolism, aminoacyl tRNA biosynthesis, riboflavin metabolism, and limonene and pinene degradation, were significantly ( $p < 0.05$ ) enriched by cluster 2 and were different compared to pathways enriched by cluster 1 (Figure 4B). Taken together, these findings show a correlation between the FAMRGs expression and dysregulation



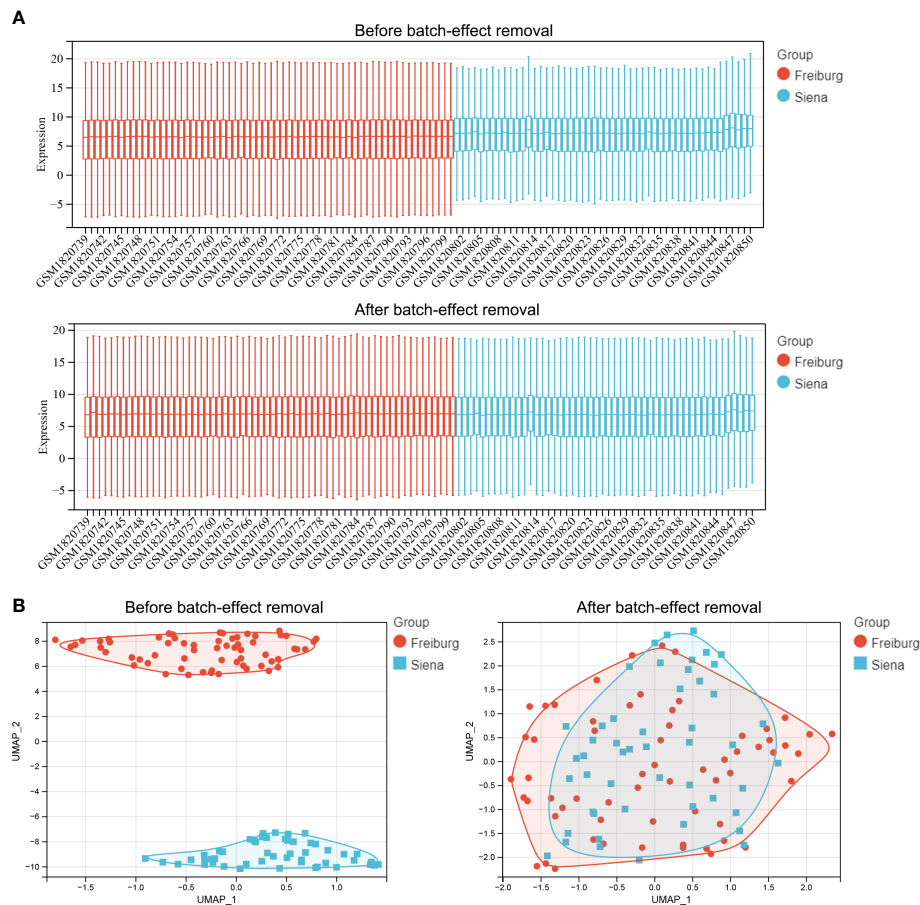


FIGURE 2

Comparison of expression data distribution and Uniform Manifold Approximation (UMAP) distribution before and after removal of the batch effect. (A) Expression data distribution. (B) UMAP distribution. The colors distinguish the Freiburg and Siena groups.

of the immune system and changes in nutrient metabolism status. This may be linked to the poor prognosis of IPF patients.

To establish the association between clusters and the infiltration of immune cells in the BALF of IPF patients, a CIBERSORT analysis was conducted to compare the relative levels of immune cells in C1 and C2 (Figure 4C). High infiltration of activated natural killer (NK) cells, monocytes, and activated mast cells were observed in the BALF of IPF patients in C1. The infiltration levels of resting dendritic cells, naive B cells, resting mast cells, M0 and M2 macrophages, and resting NK cells were high in the BALF of IPF patients in C2. These results demonstrate considerable differences in the immune status of the two clusters.

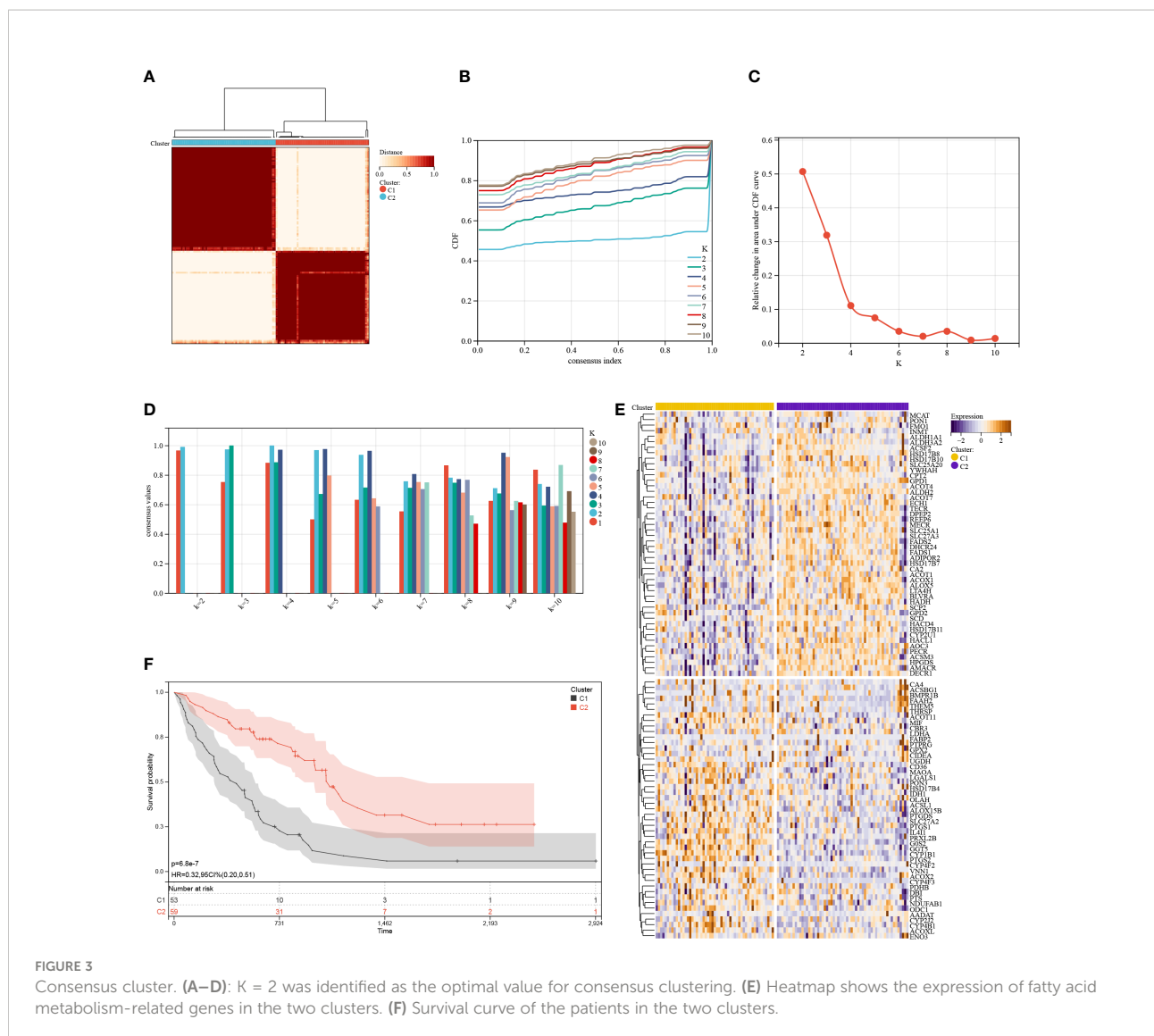
## Construction of FA metabolism-related prognostic risk model

Next, a risk signature model was created to determine the contribution of FAMRGs for the accurate prognosis of IPF

patients. The LASSO algorithm and the lambda value set at 0.23 were used to identify the most reliable prognostic FAMRGs (Figure 5A). Five genes (*GGT5*, *ACOX2*, *CYP4F3*, *HACD4*, and *ODC1*) were identified and used to generate the FAMRGs-associated prognostic risk model (Figure 5B). The Kaplan-Meier survival analysis shows that each gene could independently serve as a prognostic indicator for IPF patients (Figure S1). The following formula was used to determine the Riskscore for each patient in the training and validation cohorts:

$$\text{Riskscore} = 0.0949389697248446 \times \text{expression value of } GGT5 + 0.315559675709913 \times \text{expression value of } ACOX2 + 0.00450989430661515 \times \text{expression value of } CYP4F3 - 0.284593576349762 \times \text{expression value of } HACD4 + 0.190681441039345 \times \text{expression value of } ODC1.$$

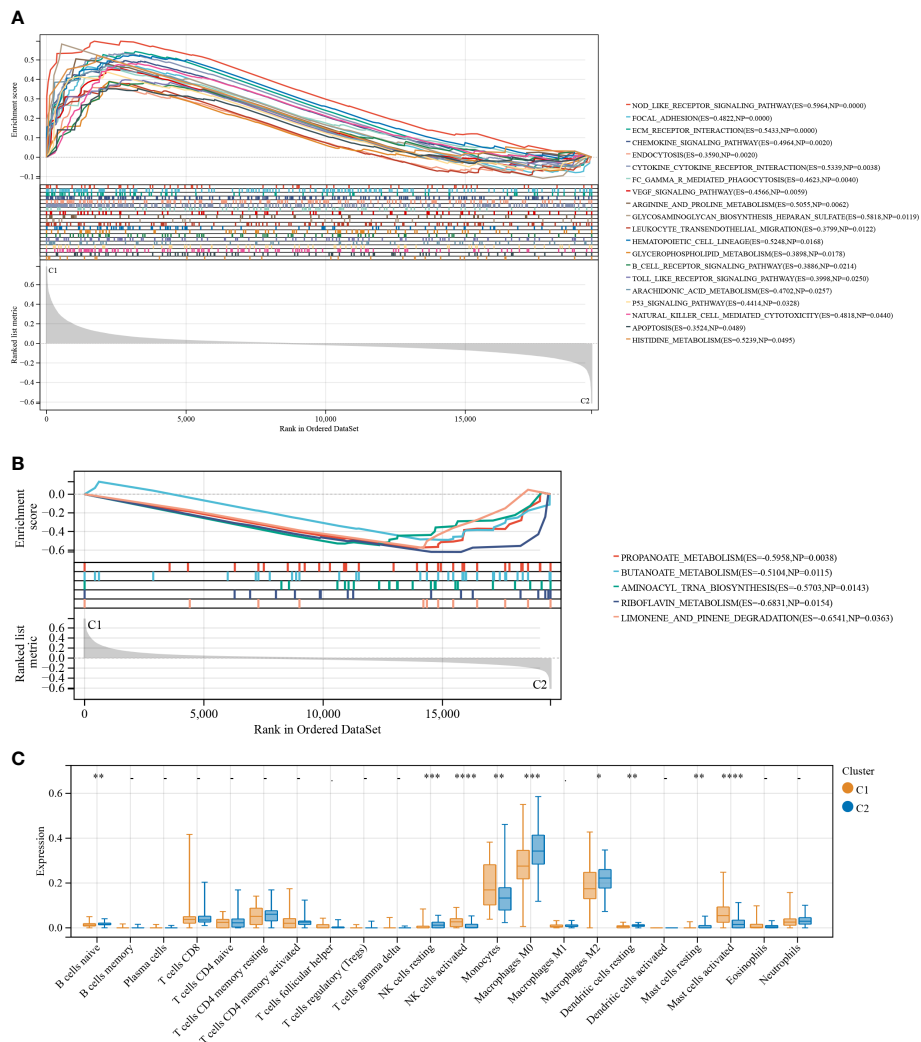
The patients were classified into low- and high-risk groups based on the median score. Kaplan-Meier survival analysis (Figures 5C, D) shows that IPF patients in the high-risk group experienced significantly more rapid progression compared to IPF patients in the low-risk group in both the training ( $p <$



0.0001) and the validation cohort ( $p < 0.01$ ). Therefore, the established prognostic risk model successfully grouped the patients with IPF into low- and high-risk groups. Moreover, high expression of the four candidate genes (*GGT5*, *ACOX2*, *CYP4F3*, and *ODC1*) and low expression of a candidate gene (*HACD4*) were observed in the high-risk group compared to the low-risk group (Figures 5E, F). The ROC curve was used to investigate the predictive ability of the Riskscore for patient prognosis and the area under the curve (AUC) values for 1-, 2-, 3-years were calculated. In the training cohort, the AUC values for determining the accuracy of Riskscore as a predictive marker for 1-, 2-, 3-years were 0.794, 0.849, and 0.879, respectively (Figure 5G). Further, in the validation cohort, the AUC values for 1-, 2-, and 3-years were 0.622, 0.679, and 0.723, respectively (Figure 5H). Based on these results, the established risk model had the potential to predict the prognosis of IPF patients.

Finally, the CIBERSORT algorithm was used to determine the differences in infiltration levels of different immune cells between the two groups and the association between the Riskscore and the immune cell infiltration in the BALF (Figure 6A). Patients with elevated Riskscore showed increased levels of activated mast cells, monocytes, and activated NK cells, and the levels of resting mast cells, naive B cells, resting dendritic cells, M2 macrophages, activated memory CD4, and CD8 T cells were relatively lower. Correlation analysis reveal an inversely correlation between Riskscore and the naive B cells (Figure 6B;  $r = -0.249$ ,  $p < 0.01$ ), CD8 T cells (Figure 6C;  $r = -0.218$ ,  $p < 0.05$ ), activated memory CD4 T cells (Figure 6D;  $r = -0.257$ ,  $p < 0.01$ ), M2 macrophages (Figure 6G;  $r = -0.317$ ,  $p < 0.001$ ), resting dendritic cells (Figure 6H;  $r = -0.432$ ,  $p < 0.001$ ), resting mast cells (Figure 6I;  $r = -0.272$ ,  $p = 0.004$ ). A positive correlation was observed between activated NK cells (Figure 6E;  $r = 0.327$ ,  $p <$





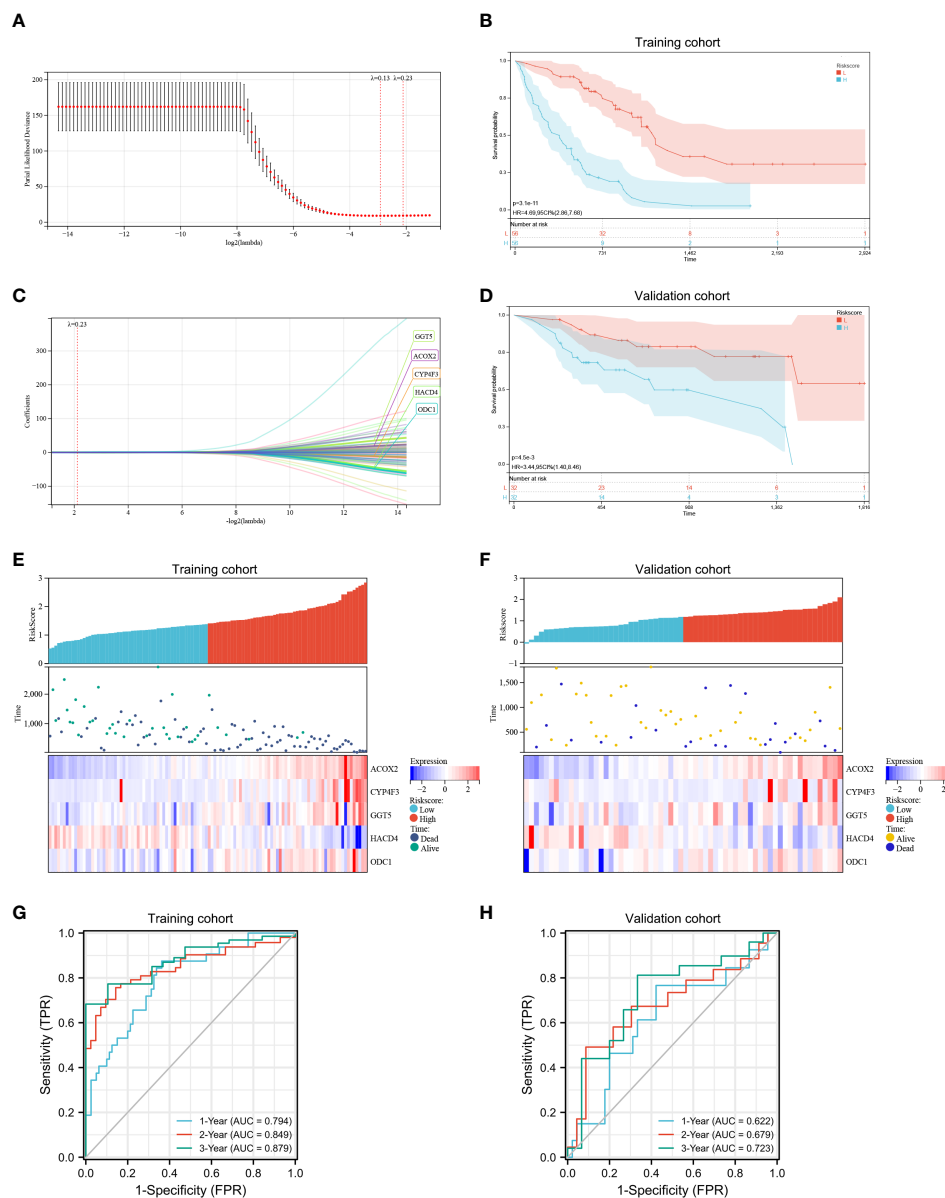
**FIGURE 4**  
 Enrichment analysis and the immune infiltration characteristics of the two identified subgroups with distinct prognoses. **(A)** GSEA for cluster 1 (C1) and cluster 2 (C2). Immune and metabolism-related pathways were significantly enriched by cluster 1. **(B)** and another five metabolism-related pathways were significantly enriched by cluster 2. **(C)** Boxplots shows differences in the infiltrating immune cells between clusters 1 and 2. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; and \*\*\*\* $p < 0.0001$ .

0.001), monocytes (Figure 6F;  $r = 0.252$ ,  $p < 0.01$ ), activated mast cells (Figure 6j;  $r = 0.464$ ,  $p < 0.001$ ) and Riskscore. These results show a close correlation between the constructed risk model and the level of immune cell infiltration in the BALF of IPF patients.

### Established risk model's independence

Furthermore, the correlation between the Riskscore and clinical characteristics was evaluated, and the independence of the risk model was verified by performing subgroup and regression analysis. There was no difference in the Riskscore

between IPF patients of different ages (Figure 7A) and sex (Figure 7B), suggesting no correlation between Riskscore and clinical parameters. Once the patients were classified based on age (Figures 7C, D) and sex (Figures 7E, F), the risk model demonstrated powerful predictive ability. The IPF patients with lower Riskscore had more favorable outcomes. Additionally, univariate and multivariate Cox regression analysis show that the developed risk model could independently serve as a predictive indicator for the prognosis of IPF patients (Table 2). These results suggest that the constructed risk model demonstrated independence in predicting the prognosis of IPF patients.



**FIGURE 5** Development and assessment of a prognostic fatty acid metabolism-related signature for patients with IPF. (A, B) LASSO regression analysis with optimal lambda. (C, D) Kaplan–Meier survival analysis of patients with high- and low- RiskScore in training (C) and validation cohorts (D), IPF patients in the high- RiskScore group had poor survival prognosis compared to patients in the low- RiskScore group. (E, F) Distribution of the RiskScore, survival status, and the expression of the five candidate genes in the high- and low- RiskScore groups in the training (E) and validation (F) cohorts. (G, H) In the training (G) and validation (H) cohorts, time-dependent receiver operating characteristic analysis showed that the RiskScore had better performance in predicting the survival of patients with IPF.

## Development of a predictive nomogram

A predictive nomogram was constructed to estimate the prognosis risk of IPF patients by calculating the chances of survival of patients 1, 2, and 3 years after diagnosis (Figure 8A). The accuracy of the nomogram was verified by testing it on training and validation cohorts. The C-index, the

calibration curve (Figures 8B–G), and decision curve analysis (Figure S2) demonstrated satisfactory accuracy of the diagnostic nomogram model. In the training cohort, the C-index for the nomogram was 0.7642 (95% CI: 0.7125–0.7962), and the C-index for the validation cohort was 0.6906 (95% CI: 0.5578–0.8234). The actual overall survival of the patients for 1, 2, and 3 years was very similar to the predicted overall survival of the patients



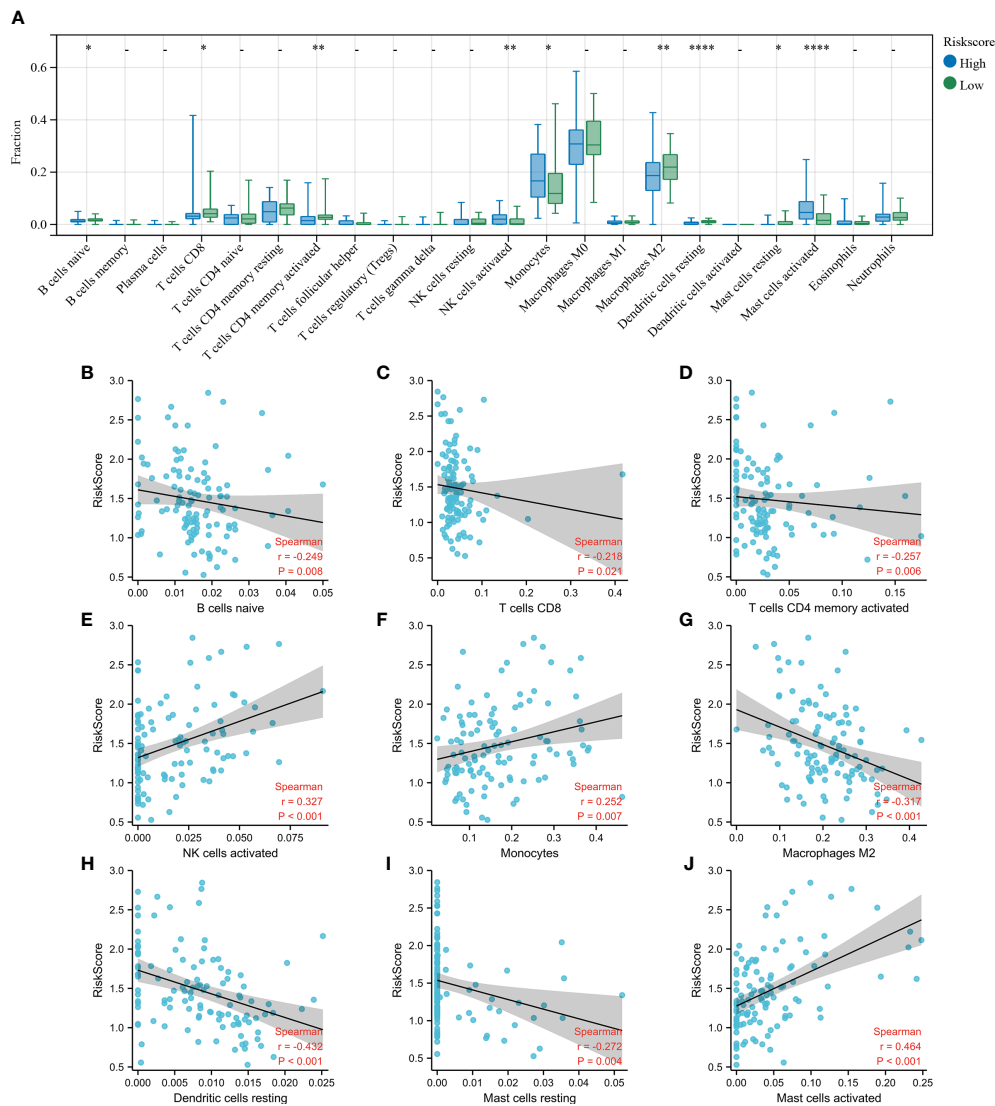


FIGURE 6

Estimation of immune cell infiltration in the bronchoalveolar lavage fluid (BALF). (A) Level of infiltration in the high- and low- RiskScore groups using the CIBERSORT algorithm. The plots show the differences in immune infiltration scores between the high- and low- RiskScore groups. (B–J) The Spearman’s rank correlation between RiskScore and the fraction of BALF immune cells is shown. Scatterplots show that the RiskScore negatively correlated with naive B cells (B), CD8 T cells (C), activated memory CD4 T cells (D), M2 macrophages (G), resting dendritic cells (H), and resting mast cells (I), while RiskScore positively correlated with activated NK cells (E), monocytes (F), and activated mast cells (J). \* $p < 0.05$ ; \*\* $p < 0.01$ ; and \*\*\*\* $p < 0.0001$ .

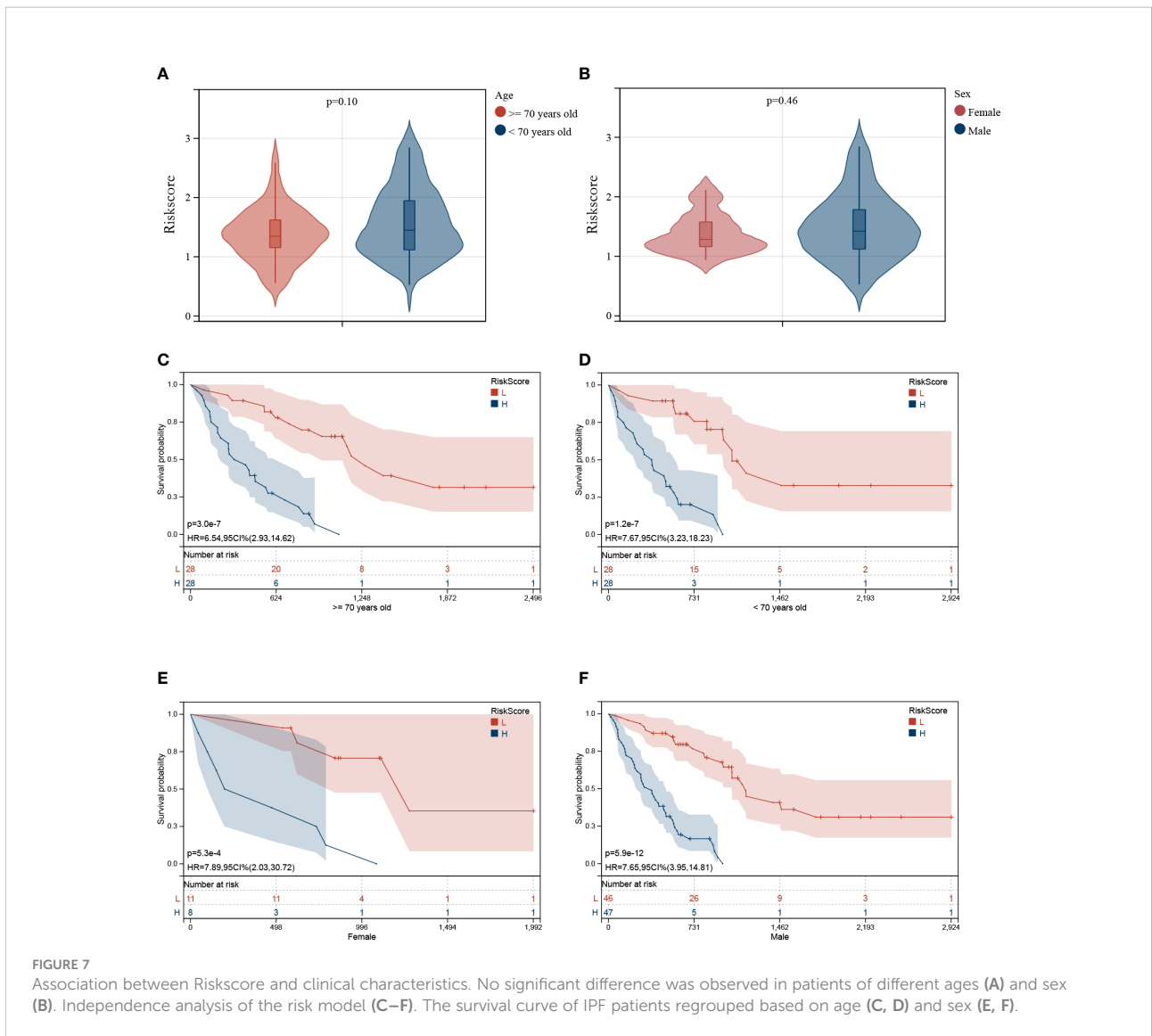
in the training cohort (Figures 8B–D). Consistent results were observed in the validation cohort (Figures 8E–G). Based on these results, it is tempting to conclude that the nomogram could accurately predict the prognosis of IPF patients.

## Discussion

The condition of IPF patients worsens over time and could be potentially fatal. Very few therapeutic options are available for

the treatment of IPF patients, and the therapeutic outcomes are poor and have an unfavorable prognosis. Despite the advancements in diagnosis and therapeutic strategies, the survival rate in IPF patients has not improved. The clinical progression of IPF is highly variable and unpredictable; hence, better risk assessment strategies and personalized, targeted therapy for IPF patients are necessary (32).

In this study, two subgroups of IPF patients were identified based on the expression profiles of genes associated with FA metabolism and had significantly different overall survival. The



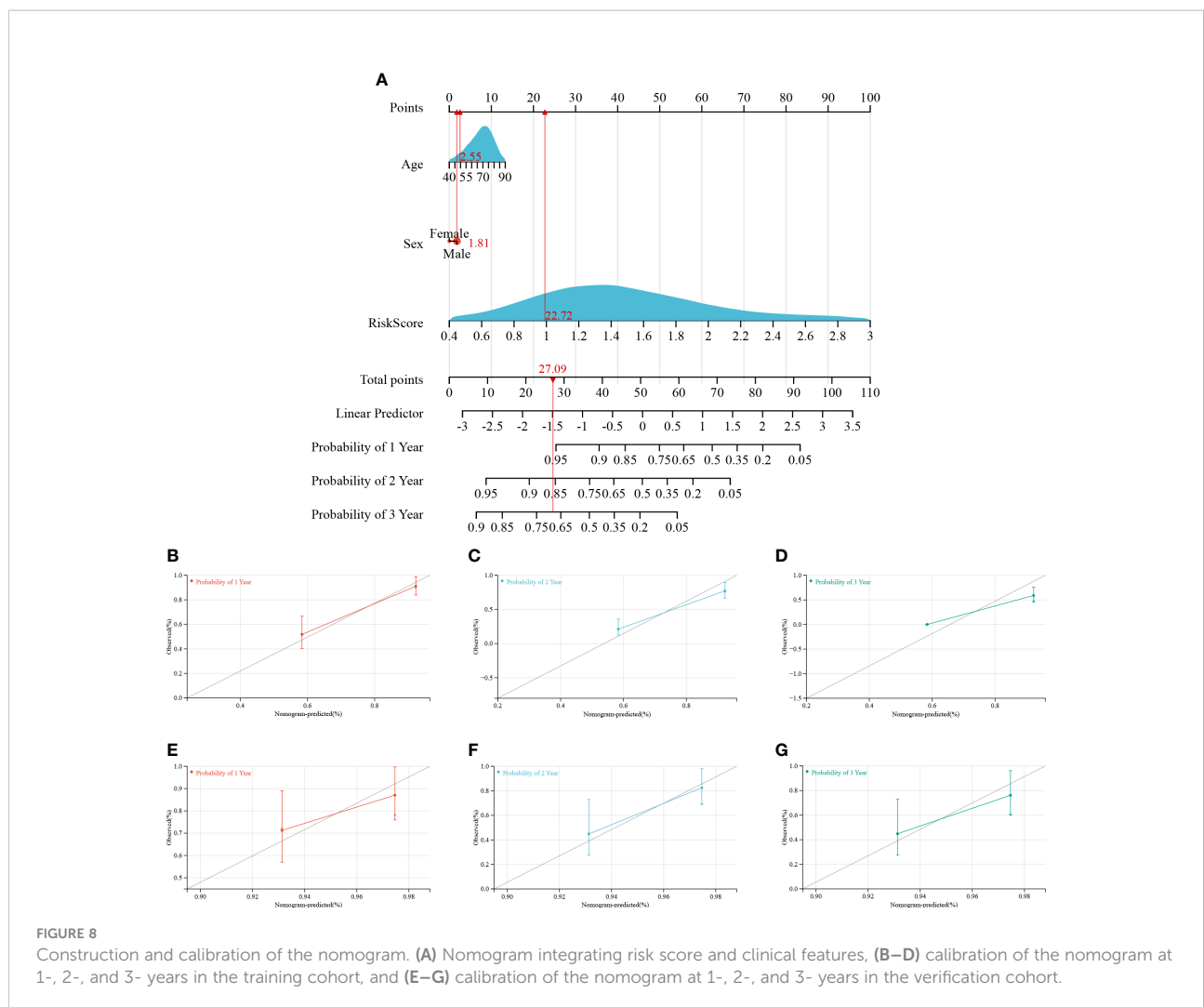
enrichment analysis revealed that the two subgroups enriched different immune and metabolism pathways. Further, the immune analysis revealed that two subgroups had different immune cell infiltration patterns. Additionally, a predictive

risk model based on FAMRGs was developed and could accurately predict the prognosis of IPF patients. Therefore, the results of our study shed light on developing IPF-specific therapies.

TABLE 2 Univariate and multivariate analysis of RiskScore and characteristics in the training cohort.

Variates	Total (n)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P-value	Hazard ratio (95% CI)	P-value
Age	112	0.986 (0.963-1.010)	0.255		
Sex	112				
Male	93	Reference			
Female	19	0.810 (0.436-1.504)	0.505		
RiskScore	112	10.354 (5.919-18.113)	<0.001	10.354 (5.919-18.113)	<0.001

CI, Confidence interval.



Consensus clustering was used to classify the IPF patients into different groups based on the gene expression matrix, and the results were reliable. Two subgroups were initially identified by consensus clustering based on the FAMRGs expression in IPF patients. The results revealed that the two subgroups had significantly different overall survival. These findings further confirmed that the FA metabolic subtypes in IPF patients influence the prognosis, and the prediction models constructed based on FAMRGs were reliable.

Furthermore, the enrichment analysis was conducted between the two subgroups to investigate the inherent biological processes. GSEA was used to elucidate the gene expression pattern in various subgroups (33). GSEA results revealed the enrichment of immune-related pathways, including natural killer cell-mediated cytotoxicity, B cell receptor signaling pathway, Toll-like receptor signaling pathway, leukocyte transendothelial migration, chemokine signaling pathway, and Nod-like receptor signaling pathway in cluster 1 with poor outcomes. Based on the results, it is likely

that dysregulated immune system could predict the role of FA metabolism in the onset and progression of IPF. Further, the association between FA metabolism and immune dysregulation could be regulated by an epigenetic process such as DNA methylation influenced by genetic variation (34). The results showed that cluster 1 and cluster 2 had different metabolic signatures. The pathways enriched in cluster 1 were predominantly linked to the metabolism of amino acids, including arginine, proline, and histidine metabolism, and lipid metabolism pathways, including the metabolism of glycerophospholipid and arachidonic acid. Short-chain FA metabolism pathways, including propanoate metabolism and butanoate metabolism, and other metabolism pathways, including aminoacyl tRNA biosynthesis, riboflavin metabolism, limonene, and pinene degradation, were mainly enriched in cluster 2. Since the classification was based on FA metabolism-related genes, the result showed that the crosstalk between FA metabolism and other nutrient metabolism could affect IPF pathogenesis and was worth further exploration.

As mentioned previously, immune systems play an integral role in the onset and progression of IPF and are also associated with FA metabolism. The CIBERSORT algorithm is a tool that analyses biological information and evaluates the expression levels of immune cells based on RNA-seq data, and obtains ratios of different immune cells from samples. It is extensively used to examine the infiltration of immune cells in various diseases, such as tumors, osteoarthritis, and lupus nephritis (35–37). Li et al. established a hypoxia-immune-based prediction model to determine the prognosis among IPF patients using CIBERSORT (38). Hence, CIBERSORT was used to determine the level of immune cell infiltration in the two clusters. The results revealed the presence of several immune cells in BALF of patients that could potentially be associated with IPF prognosis. The immune cells identified were mainly activated and resting mast cells, naive B cells, M0 and M2 macrophages, monocytes, activated and resting NK cells, and resting dendritic cells. Increased infiltration of activated NK cells, activated mast cells, and monocytes were observed in patients in cluster 1 with poor survival outcomes. Similarly, in the constructed risk model, an increase in infiltration of these three immune cell populations was observed in the high-risk group with poor prognosis and was positively correlated with Riskscore. Previous studies have shown that pulmonary fibrosis patients had higher levels of NK cell infiltration in BALF compared to patients with sarcoidosis (39). Scott et al. indicated that elevated levels of circulating monocyte count could be a cellular biomarker for poor outcomes in IPF patients (40). Kawanami et al. have shown that patients with fibrotic lung disease had significantly higher numbers of mast cells in their lungs. Mast cells are often localized around the thickened regions of the alveolar septa and are located near abnormal epithelial cells (41). Taken together, these results revealed that activated mast cells, monocytes, and activated NK cells in the BALF of patients with IPF may promote disease progression.

Based on our results, it is tempting to infer that dysregulation in FA metabolism results in the disturbance in the immune system and metabolism in BALF of IPF patients, thereby contributing to the poor prognosis of the IPF patients. As aforementioned, reprogramming FA metabolism was identified as a distinctive characteristic in IPF patients with poor prognoses. To further verify the influence of FA metabolic disorders on IPF and examine the prognostic value of FAMRGs in IPF patients, a prognostic risk model using FAMRGs was created and tested on a separate validation cohort. The results showed that five genes used to establish the risk model were remarkably associated with the onset and progression of IPF. Gamma-glutamyltransferase 5 (GGT5) cleaves glutathione peptides to maintain the glutathione balance in the human body (42). Previous research has shown that mice lacking gamma-glutamyl transpeptidase developed a less severe bleomycin-induced pulmonary fibrosis (43). However, the role of GGT5 in FA metabolism and the progression of IPF is

unknown. Branched-chain acyl-CoA oxidase (*ACOX2*), a peroxisomal enzyme, plays a role in the metabolism of bile acid intermediates and branched-chain FAs. Further, *ACOX2* deficiency has been linked to an increased risk of developing liver fibrosis (44), but its role in pulmonary fibrosis is unknown. Leukotriene-B4 omega-hydroxylase (*CYP4F3*) consists of enzymes *CYP4F3A* and *CYP4F3B*, responsible for the metabolism of leukotriene B4 and 5-hydroxyicosatetraenoic acid *via* omega oxidation reaction, which results in the inhibition and deterioration of inflammatory markers (45, 46). *CYP4F3* is associated with inflammatory diseases, like inflammatory bowel disease (47); however, the role of *CYP4F3* in IPF is unknown. 3-hydroxyacyl-CoA dehydratase 4 (*HACD4*) is involved in elongation, and the biosynthesis of very-long-chain FA; therefore, its role in IPF is worth exploring. Further, *ODC1* encodes for rate-limiting enzymes associated with the polyamine biosynthetic pathways and acts as a catalyst for converting ornithine to putrescine. This indicates the significance of *ODC1* in IPF. The survival analysis demonstrated that the developed risk model had an excellent ability to predict the survival of IPF patients in both cohorts. The prognosis of the IPF patients could be determined independently by each of the five genes. The independent and subgroup analysis demonstrated that the FAMRGs-based risk model could independently predict the prognosis of IPF patients, regardless of age and sex. Ultimately, a nomogram incorporating the Riskscore and the clinical characteristics was developed, calibrated, and tested. The results revealed that it had a powerful ability to predict patient survival. Taken together, these results provide additional evidence that FAMRGs have a prognostic and predictive role in IPF.

Due to the poor prognosis, high variability, and unpredictable nature of IPF progression, there is a need for efficient risk classification and treatment strategy for developing personalized, targeted therapy. Our results demonstrate highlights of FA metabolism in comparison to earlier studies. Our study was based on FA metabolism, which has gained considerable attention in IPF-related research. Using consensus clustering, two molecular clusters with different patient prognoses and immunological statuses were identified. Second, the biological processes based on consensus clustering results were identified, and the fundamental mechanisms associated with them were partially elucidated. Lastly, the role of FA metabolism on the infiltration of immune cells in BALF of IPF patients was elucidated. However, the role of five selected genes in IPF is unclear, and further studies are necessary to understand their role in IPF. Our results offer a good theoretical foundation for research on IPF.

However, our study also has several limitations. First, we could not establish the involvement of FAMRG in the progression of IPF due to insufficient data regarding the pulmonary function of IPF patients. Second, our results were based on bioinformatics analysis, and hence further

experimental validation is required. Thirdly, the data used for the study were retrieved from publicly available databases since our clinical practice had an insufficient number of IPF patients. Therefore, further studies evaluating the performance of the prognostic model and the mechanism of the five FAMRGs in the IPF pathogenesis using molecular biology tools and prospective cohort studies are required.

## Conclusions

In this study, consensus clustering was used to identify two distinct molecular clusters based on the FAMRGs in IPF patients. The enrichment and immune analyses show that dysregulation of FA metabolism contributes to disorders of the immune system and nutrition metabolism, which result in poor prognosis. Our study may shed light on developing novel targeted treatment approaches and offer a theoretical basis for personalized treatment strategies.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70866>.

## Author contributions

HZ and YL designed the study and revised the manuscript. YL and CG collected the data. YL analyzed and interpreted the data and drafted the manuscript. All authors have read and approved the final manuscript.

## References

1. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Resp Crit Care* (2011) 183(6):788–824. doi: 10.1164/rccm.2009-040GL
2. Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *New Engl J Med* (2018) 378(19):1811–23. doi: 10.1056/NEJMra1705751
3. Shin KH, Shin SW. Idiopathic pulmonary fibrosis. *New Engl J Med* (2018) 379(8):795–6. doi: 10.1056/NEJMra1705751
4. Mora AL, Rojas M, Pardo A, Selman M. Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. *Nat Rev Drug Discovery* (2017) 16(11):755–72. doi: 10.1038/nrd.2017.170
5. Wang ZN, Tang XX. New perspectives on the aberrant alveolar repair of idiopathic pulmonary fibrosis. *Front Cell Dev Biol* (2020) 8:580026. doi: 10.3389/fcell.2020.580026
6. Spagnolo P, Kropski JA, Jones MG, Lee JS, Rossi G, Karampitsakos T, et al. Idiopathic pulmonary fibrosis: Disease mechanisms and drug development. *Pharmacol Therapeut* (2021) 222:107798. doi: 10.1016/j.pharmthera.2020.107798
7. Kreuter M, Bonella F, Wijnsbeek M, Maher TM, Spagnolo P. Pharmacological treatment of idiopathic pulmonary fibrosis: Current

## Funding

This study was supported by grants from the Social Development Projects of Key R&D Programs in Jiangsu Province (BE2019643), the National Natural Science Foundation of Jiangsu Province (BK20171178), General Program of Jiangsu Commission of Health (H2017083), and the Project of Invigorating Health Care through Science, Technology, and Education, Jiangsu Provincial Medical Youth Talent (QNRC2016778).

## 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/fendo.2022.1001563/full#supplementary-material>

approaches, unsolved issues, and future perspectives. *BioMed Res Int* (2015) 2015:329481. doi: 10.1155/2015/329481

8. Galli JA, Pandya A, Vega-Olivo M, Dass C, Zhao H, Criner GJ. Pirfenidone and nintedanib for pulmonary fibrosis in clinical practice: Tolerability and adverse drug reactions. *Respirol* (2017) 22(6):1171–8. doi: 10.1111/resp.13024

9. Olson AL, Swigris JJ, Lezotte DC, Norris JM, Wilson CG, Brown KK. Mortality from pulmonary fibrosis increased in the united states from 1992 to 2003. *Am J Resp Crit Care* (2007) 176(3):277–84. doi: 10.1164/rccm.200701-044OC

10. Vancheri C, Failla M, Crimi N, Raghu G. Idiopathic pulmonary fibrosis: A disease with similarities and links to cancer biology. *Eur Respir J* (2010) 35(3):496–504. doi: 10.1183/09031936.00077309

11. Sunaga H, Matsui H, Ueno M, Maeno T, Iso T, Samsunarno MR, et al. Deranged fatty acid composition causes pulmonary fibrosis in Elovf6-deficient mice. *Nat Commun* (2013) 4:2563. doi: 10.1038/ncomms3563

12. 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

13. Zheng J, Dong H, Zhang T, Ning J, Xu Y, Cai C. Development and validation of a novel gene signature for predicting the prognosis of idiopathic

pulmonary fibrosis based on three epithelial-mesenchymal transition and immune-related genes. *Front Genet* (2022) 13:865052. doi: 10.3389/fgene.2022.865052

14. 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

15. Namgaladze D, Brüne B. Macrophage fatty acid oxidation and its roles in macrophage polarization and fatty acid-induced inflammation. *Biochim Biophys Acta* (2016) 1861(11):1796–807. doi: 10.1016/j.bbali.2016.09.002

16. Parks BW, Black LL, Zimmerman KA, Metz AE, Steele C, Murphy-Ullrich JE, et al. CD36, but not G2A, modulates efferocytosis, inflammation, and fibrosis following bleomycin-induced lung injury. *J Lipid Res* (2013) 54(4):1114–23. doi: 10.1194/jlr.M035352

17. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet* (2017) 389(10082):1941–52. doi: 10.1016/S0140-6736(17)30866-8

18. Lin Y, Xu Z. Fibroblast senescence in idiopathic pulmonary fibrosis. *Front Cell Dev Biol* (2020) 8:593283. doi: 10.3389/fcell.2020.593283

19. Reddy AT, Lakshmi SP, Zhang Y, Reddy RC. Nitrated fatty acids reverse pulmonary fibrosis by dedifferentiating myofibroblasts and promoting collagen uptake by alveolar macrophages. *FASEB J* (2014) 28(12):5299–310. doi: 10.1096/fj.14-256263

20. Liu W, Meridew JA, Aravamudan A, Ligresti G, Tschumperlin DJ, Tan Q. Targeted regulation of fibroblast state by CRISPR-mediated CEBPA expression. *Respir Res* (2019) 20(1):281. doi: 10.1186/s12931-019-1253-1

21. Romero F, Hong X, Shah D, Kallen CB, Rosas I, Guo Z, et al. Lipid synthesis is required to resolve endoplasmic reticulum stress and limit fibrotic responses in the lung. *Am J Resp Cell Mol* (2018) 59(2):225–36. doi: 10.1165/rcmb.2017-0340OC

22. Kheirollahi V, Wasnick RM, Biasin V, Vazquez-Armendariz AI, Chu X, Moiseenko A, et al. Metformin induces lipogenic differentiation in myofibroblasts to reverse lung fibrosis. *Nat Commun* (2019) 10(1):2987. doi: 10.1038/s41467-019-10839-0

23. Zhao H, Chan-Li Y, Collins SL, Zhang Y, Hollowell RW, Mitzner W, et al. Pulmonary delivery of docosahexaenoic acid mitigates bleomycin-induced pulmonary fibrosis. *BMC Pulm Med* (2014) 14:64. doi: 10.1186/1471-2466-14-64

24. 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 Resp Crit Care* (2019) 199(5):622–30. doi: 10.1164/rccm.201712-2551OC

25. Kim HS, Yoo HJ, Lee KM, Song HE, Kim SJ, Lee JO, et al. Stearic acid attenuates profibrotic signaling in idiopathic pulmonary fibrosis. *Respirol* (2021) 26(3):255–63. doi: 10.1111/resp.13949

26. Chu SG, Villalba JA, Liang X, Xiong K, Tsoyi K, Ith B, et al. Palmitic acid-rich high-fat diet exacerbates experimental pulmonary fibrosis by modulating endoplasmic reticulum stress. *Am J Resp Cell Mol* (2019) 61(6):737–46. doi: 10.1165/rcmb.2018-0324OC

27. Taminau J, Meganck S, Lazar C, Steenhoff D, Coletta A, Molter C, et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC Bioinf* (2012) 13:335. doi: 10.1186/1471-2105-13-335

28. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* (2007) 8(1):118–27. doi: 10.1093/biostatistics/kxj037

29. Trozzi F, Wang X, Tao P. UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: A comparison study. *J Phys Chem B* (2021) 125(19):5022–34. doi: 10.1021/acs.jpcc.1c02081

30. Wilkerson MD, Hayes DN. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26(12):1572–3. doi: 10.1093/bioinformatics/btq170

31. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337

32. Martinez FJ, Safran S, Weycker D, Starko KM, Bradford WZ, King TE, et al. The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med* (2005) 142(Pt 1):963–7. doi: 10.7326/0003-4819-142-12\_Part\_1-200506210-00005

33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *P Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

34. Hawe JS, Wilson R, Schmid KT, Zhou L, Lakshmanan LN, Lehne BC, et al. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. *Nat Genet* (2022) 54(1):18–29. doi: 10.1038/s41588-021-00969-x

35. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor-infiltrating immune cells with CIBERSORT. *Methods Mol Biol* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1\_12

36. Deng YJ, Ren EH, Yuan WH, Zhang GZ, Wu ZL, Xie QQ, GRB10 and E2F3 as diagnostic markers of osteoarthritis and their correlation with immune infiltration. *Diagnostics (Basel)* (2020) 10(3):171. doi: 10.3390/diagnostics10030171

37. Cao Y, Tang W, Tang W. Immune cell infiltration characteristics and related core genes in lupus nephritis: Results from the bioinformatics analysis. *BMC Immunol* (2019) 20(1):37. doi: 10.1186/s12865-019-0316-x

38. 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

39. Bergantini L, Cameli P, d'Alessandro M, Vagaggini C, Refini RM, Landi C, et al. NK and NKT-like cells in granulomatous and fibrotic lung diseases. *Clin Exp Med* (2019) 19(4):487–94. doi: 10.1007/s10238-019-00578-3

40. 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 Resp Med* (2019) 7(6):497–508. doi: 10.1016/S2213-2600(18)30508-3

41. Kawanami O, Ferrans VJ, Fulmer JD, Crystal RG. Ultrastructure of pulmonary mast cells in patients with fibrotic lung disorders. *Lab invest* (1979) 40(6):717–34.

42. Heisterkamp N, Groffen J, Warburton D, Sneddon TP. The human gamma-glutamyltransferase gene family. *Hum Genet* (2008) 123(4):321–32. doi: 10.1007/s00439-008-0487-7

43. Pardo A, Ruiz V, Arreola JL, Ramirez R, Cisneros-Lira J, Gaxiola M, et al. Bleomycin-induced pulmonary fibrosis is attenuated in gamma-glutamyl transpeptidase-deficient mice. *Am J Resp Crit Care* (2003) 167(6):925–32. doi: 10.1164/rccm.200209-1007OC

44. Vilarinho S, Sari S, Mazzacava F, Bilgüvar K, Esendagli-Yilmaz G, Jain D, et al. ACOX2 deficiency: A disorder of bile acid synthesis with transaminase elevation, liver fibrosis, ataxia, and cognitive impairment. *P Natl Acad Sci USA* (2016) 113(40):11289–93. doi: 10.1073/pnas.1613228113

45. Corcos L, Lucas D, Le Jossic-Corcos C, Dréano Y, Simon B, Plée-Gautier E, et al. Human cytochrome P450 4F3: Structure, functions, and prospects. *Drug Metabol Drug Interact* (2012) 27(2):63–71. doi: 10.1515/dmdi-2011-0037

46. Powell WS, Rokach J. Biosynthesis, biological effects, and receptors of hydroxyeicosatetraenoic acids (HETEs) and oxo-eicosatetraenoic acids (oxo-ETEs) derived from arachidonic acid. *Biochim Biophys Acta* (2015) 1851(4):340–55. doi: 10.1016/j.bbali.2014.10.008

47. Scafoli E, Liverani E, Belluzzi A. The imbalance between n-6/n-3 polyunsaturated fatty acids and inflammatory bowel disease: A comprehensive review and future therapeutic perspectives. *Int J Mol Sci* (2017) 18(12):2619. doi: 10.3390/ijms18122619