# Fatty Acid Metabolism-related Genes in Bronchoalveolar Lavage Fluid Unveil Prognostic and Immune Infiltration in Idiopathic Pulmonary Fibrosis

**Yin Lyu**
Xuzhou Medical University
**Chen Guo**
Xuzhou Medical University
**Hao Zhang** ( zhanghao@xzhmu.edu.cn )
Xuzhou Medical University

**Research Article**

**Keywords:** Idiopathic pulmonary fibrosis, fatty acid metabolism, immune infiltration, biomarker, prognosis

**Posted Date:** June 22nd, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1764279/v1

# Abstract

**Background:** Idiopathic pulmonary fibrosis, often abbreviated as IPF, is a condition that is both chronic and progressive, and has an unfavorable prognosis. Recent research has demonstrated that individuals with IPF exhibit characteristic alterations in the fatty acid metabolism in their lungs, suggesting an association with the disease pathogenesis. Here, we explored whether the gene signature associated with fatty acid metabolism might be used as a reliable biological marker for predicting the IPF patients' survival.

**Methods:** Data of fatty acid metabolism-related genes (FAMRGs) were extracted from the Kyoto Encyclopedia of Genes and Genomes (KEGG), Hallmark, and Reactome databases. The GSE70866 data set was retrieved for information on individuals diagnosed with IPF. Next, a consensus clustering approach was used to discover novel molecular subgroups. Then, Gene Set Enrichment Analysis (GSEA) was conducted to evaluate the mechanisms involved. CIBERSORT was utilized to ascertain the infiltration levels of immune cells in the identified subgroups premised on specific gene expression signatures of immune cells. Finally, the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm and multivariate Cox regression analysis were utilized in the development of a prognostic risk model.

**Results:** The patterns of gene expression that are linked to fatty acid metabolism were used to construct two subgroups that had remarkably different prognoses. GSEA indicated that immune-related pathways, were significantly altered between the two subgroups, and different subgroups had different metabolic characteristics. Poor prognosis was associated with high immune cell infiltration, particularly activated NK cells, monocytes, and activated mast cells. A risk model that was developed premised on FAMRGs showed excellent significance for use in predicting IPF. The prognosis for individuals diagnosed with IPF might be correctly predicted using a nomogram that incorporated clinical features and the risk model.

**Conclusion:** The proposed fatty acid metabolism-related gene signature is a potential biological marker for predicting the clinical outcomes and infiltration status of immune cells, and may ultimately increase the accuracy of treating patients with IPF.


# Background

Idiopathic pulmonary fibrosis (IPF) is a kind of interstitial lung disease (ILD) that is chronic, progressive, and age-related. The cause of this illness is unclear, and it leads to considerable morbidity and a poor prognosis (1, 2). Apoptosis of the alveolar epithelial cells, generation of pro-fibrotic factors, and stimulation of myofibroblasts and fibroblasts are the mechanistic characteristics (3–5). Although the Food and Drug Administration has given its approval to two clinical drugs, nintedanib, and pirfenidone, none of these drugs can reverse the lung damage caused by IPF, nor can reduce the death rate linked to it, and both of these drugs have tolerability problems (6–8). IPF patients experience unfavorable clinical outcomes, and the median survival rates at 3 and 5 years after diagnosis are 50% and 20%, respectively, like that of lung cancer patients (9, 10). Consequently, patients diagnosed with IPF need individualized

targeted treatment, which necessitates the development of a risk classification strategy and the identification of prognostic genes.

In recent times, substantial alterations in the metabolism of fatty acids (FAs) have been observed in patients with IPF, indicating that these alterations perform an integral function in the pathogenesis of this illness. The pathways of FA metabolism are complex, and disturbance of these pathways in the lungs might result in the formation of a pro-fibrotic phenotype in macrophages, epithelial cells, and myofibroblasts/ fibroblasts. Sunaga et al. reported that treatment with palmitic acid induces apoptosis and the expression of transforming growth factor (TGF)-β1 in cultured alveolar type (AT) 2 cells (11). With the accumulating knowledge of the immune system's involvement in the progression of fibrosis and the body's response to treatments, we are gaining an extensive interest in the context of interstitial lung disorders (12, 13). Macrophages, which are the immune cells that are predominantly distributed in the lung tissues, perform a noteworthy function in the pathophysiology of pulmonary fibrosis. Macrophages can polarize into two phenotypes-M1 and M2. After they have been activated, M2 macrophages release profibrotic factors including TGF-β1, which stimulate the activity of fibroblasts and the deposition of extracellular matrix (ECM) (14). There may be a strong link between the polarization of macrophages and the metabolism of FA. The M2 phenotype depends on the transcription factor peroxisome proliferator-activated receptor (PPAR)-γ, which has extensive FAs functioning as its natural ligands (15). In addition, FAs are responsible for enhancing the FA receptor CD36 expression, which induces the M2 phenotype by simultaneously increasing FA uptake and self-reinforcing the profibrotic activation cycle (16). Epithelial cells that are dysfunctional, as well as polarized macrophages, produce numerous profibrotic cytokines, which induce fibroblasts to differentiate into myofibroblasts. Myofibroblasts in IPF lungs, which can produce excessive extracellular matrix and cause basement membrane disruption, are considered the main effector cells (17, 18). Reddy et al. found that nitrated FAs are agonists for the PPAR-γ receptor, and they stimulate the dedifferentiation of myofibroblasts by inhibiting the impacts of TGF-β1 (19). Furthermore, It has been demonstrated that the CCAAT enhancer-binding protein (C/EBP) α, which is modulated by fatty acids as well as their derivatives, is responsible for promoting the dedifferentiation of myofibroblasts into lipofibroblasts (20). As a consequence, targeting the FA metabolism has been considered an innovative approach for treatment, and treatments that target the metabolism of FA have been tried in pre-clinical models of pulmonary fibrosis (21–23). However, the role of FAMRGs in the process of IPF is poorly understood and the prognostic significance of these genes is yet to be elucidated.

At the moment, lung biopsy is recognized to be the most effective way of determining the molecular biomarkers that may be used in the diagnostic and prognostic assessment of IPF. On the other hand, because of the intrusive nature of lung biopsy, its application is restricted. Prasse A et al. reported that bronchoalveolar lavage (BAL) cell gene expression signature could predict mortality in IPF (24). Other studies have shown that the expression levels of saturated long-chain FAs, including stearic acid, oleic acid, and palmitic acid, were low in IPF bronchoalveolar lavage fluid (BALF), while the levels of stearic and palmitic acid were high in BALF from IPF patients (25, 26). Therefore, we propose that the changes in FA metabolism in the alveolar compartment of patients with IPF are indicative of disease progression.

Here, we analyzed FAMRGs in BAL cells to comprehensively examine the impacts of FA metabolism on IPF patients' survival. In addition, we developed a risk score model that was premised on FAMRGs to examine the predictive power of FAMRGs in IPF. Our research may offer novel insights for investigating the molecular basis behind IPF, provide innovative perspectives on the targeted therapeutic techniques of IPF, and facilitate the development of individualized treatments for individuals who suffer from IPF.

# Methods

# Datasets and Samples

In all, there were 176 IPF patients whose gene expression profiles and clinical data were incorporated into this research. From the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/), we acquired data on the gene expression of BAL cells (GSE70866). Among them, The training cohort consisted of 112 individuals with IPF (from Siena and Freiburg), and RNA microarray chips from those patients were collected utilizing the Agilent-028004 SurePrint G3 Human GE 8x60K Microarray, whereas the validation cohort consisted of 64 IPF patients (from Leuven) who had their RNA microarray chips analyzed with an Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray. The details of the patients who were included in this study are presented in Table 1. In addition, the raw data of IPF patients in the Freibug and Siena groups were merged into the training cohort using the R package inSilicoMerging (27), followed by removing the batch effect using the empirical Bayes method (28). To determine whether or not the batch effect had been eliminated, the Uniform Manifold Approximation and Projection (UMAP) method was utilized (29). Datasets of 309 FAMRGs were obtained from the KEGG (hsa00071), Hallmark, and Reactome (R-HSA-8978868) databases.

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. | | | |

# Identification of Molecular Subgroups

The univariate Cox regression analysis discovered 95 genes that were linked to the prognosis of patients with IPF. With the help of the R program ConsensusClusterPlus, consensus clustering was carried out premised on the expression matrix of the 95 genes (30).

# Enrichment and Immune Analyses

GSEA was carried out using the same dataset to examine the variations (differences) that existed across the clusters. Meanwhile, Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) (31) analysis was utilized to ascertain the proportions of 22 human immune cell subsets in the BALF of IPF patients premised on the gene expression data.

# Construction and Verification of a Risk Model

By employing the R package glmnet, the Least Absolute Shrinkage and Selection Operator (LASSO) regression was performed based on a 5-fold cross-validation in the training cohort to filter out FAMRGs linked to IPF patients' survival. The value that was determined to be optimal was the one with the lowest lambda value. The genes that were found to be linked to IPF patients' prognosis (survival time and survival status) after being identified by LASSO regression were utilized to generate a prognostic risk

signature on the basis of the regression coefficients. The following formula was applied to determine each patient's risk score within the training and validation cohorts: Riskscore = Σ Exp (mRNAi) × Coefficient (mRNAi). Patients were categorized into low- and high-risk groups depending on the median value. To assess the model's prognostic predictive ability, survival analysis as well as time-dependent receiver operating characteristic (ROC) curves were utilized.

# Statistical Analysis

R version 4.0.2 was utilized to execute all analyses of statistical data. The Kaplan-Meier analyses, as well as the long-rank tests, were utilized to perform the survival analyses, and the risk model's predictive accuracy was assessed using the time-dependent ROC that is available via the R package survivalROC. After reclassifying the patients depending on their ages and genders, we carried out a subgroup analysis. The statistical comparison between two groups was performed utilizing the Student's t-test, whereas comparisons involving multiple groups were performed by performing one-way ANOVA. A P-value < 0.05 was judged to be the criterion of statistical significance.

## Results

# Removal of Batch Effects in the Training Cohort

The empirical Bayes method was used to eliminate batch effects between the Freiburg and Siena groups in the training cohort. According to the boxplot, the sample distribution of each group differed greatly 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 tended to be consistent (Fig. 1A). The results of the UMAP demonstrated that the samples from each group clustered together prior to batch effect elimination, which provided evidence for the presence of the batch effect. After the batch effect was eliminated, the samples from each group clustered together and became intertwined, which is an indication that the batch effect has been eliminated (Fig. 1B).

# Molecular Subtype Identification Predicated on FAMRGs

The consensus clustering method was used to classify the IPF patients who were included in the training cohort into distinct groups according to the 95 prognostic genes that were derived from the univariate Cox analysis (**Table S1**). K = 2 was determined to provide the optimum clustering stability (Fig. 2A-D). 53 patients were clustered to form cluster 1 (C1), and 59 patients were clustered to form cluster 2 (C2). The heatmap allowed for the visualization of the FAMRGs' expression levels for the two different clusters (Fig. 2E), and it was discovered that clusters 1 and 2 had significantly different patterns of expression. In addition to this, the patients who were classified into C2 exhibited a higher overall survival (OS) rate in contrast with those classified into C1 ($p < 0.0001$; Fig. 2F). The above results illustrated that the FAMRGs can group IPF patients into two distinct molecular clusters, each of which has a distinct OS rate; thus, it is feasible to create prediction models founded on FAMRGs.

GSEA was used to clarify the possible modulatory mechanisms responsible for the difference in outcomes between clusters 1 and 2. As shown in Fig. 3A, immune-related pathways, such as the Nod-like receptor signaling pathway, the chemokine signaling pathway, leukocyte transendothelial migrations, natural killer cell-mediated cytotoxicity, the B cell receptor signaling pathway, the 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 predominantly enriched in cluster 1, indicative of poor survival ($p < 0.05$), whereas another five metabolism-related pathways, which are different from those in cluster 1, including propanoate metabolism, butanoate metabolism, aminoacyl tRNA biosynthesis, riboflavin metabolism, and limonene and pinene degradation, were significantly enriched in cluster 2 ($p < 0.05$) (Fig. 3B). Taken together, these findings illustrate the correlation between the FAMRGs expression and dysregulation of the immune system as well as changes in nutrient metabolism status, which may be linked to unfavorable prognosis among IPF patients.

To establish the association between clusters and the infiltration status of immune cells in the BALF of IPF patients, the CIBERSORT was utilized to compare the relative levels of immune cells in clusters 1 and 2 (Fig. 3C). Activated natural killer (NK) cells, monocytes, and activated mast cells showed considerably elevated infiltration levels in the BALF of patients in cluster 1, whereas the infiltration levels of resting dendritic cells, naive B cells, resting mast cells, M0 and M2 macrophages, and resting NK cells, were increased in patients in cluster 2. These findings demonstrate substantial variations in the immune status of the two molecular clusters.

# Construction of FA Metabolism-Related Prognostic Risk Model

We next created a risk signature model to determine the extent to which FAMRGs contribute to the prognostic accuracy of IPF. Using the LASSO technique and setting the lambda parameter to its optimum value of 0.23, we searched the FAMRGs to identify the most reliable prognostic genes (Fig. 4A). Five genes (GGT5, ACOX2, CYP4F3, HACD4, and ODC1) were discovered and used to generate the FAMRGs-associated prognostic signature, which is displayed in Fig. 4B. According to the findings of a Kaplan-Meier study, each of these genes independently serves as a prognostic indicator in patients with IPF (**Figure S1**). The following formula was used to determine each patient's Riskscore within the training and validation cohorts: Riskscore = 0.0949389697248446 × expression value of GGT5 + 0.315559675709913 × expression value of ACOX2 + 0.00450989430661515 × expression value of CYP4F3 − 0.284593576349762 × expression value of HACD4 + 0.190681441039345 × expression value of ODC1, and the median score was used as the basis for categorizing patients into low- and high- risk groups. Kaplan-Meier analysis (Fig. 4C, D) demonstrated that patients diagnosed with IPF classified into the high-risk group experienced a considerably more rapid progression in contrast with those within the low-risk group in the training cohort ($p < 0.0001$) as well as the validation cohort ($p < 0.01$). The established risk model successfully grouped the patients with IPF into low- and high-risk groups. Moreover, the high-risk

group was shown to express the four candidate genes (GGT5, ACOX2, CYP4F3, and ODC1) higher and one candidate gene (HACD4) lower in contrast with those in the low-risk group (Fig. 4E, F). ROC curve analysis was used to investigate the predictive ability of the Riskscore for prognosis, and we evaluated the area under the curve (AUC) values over one, two, and three years. In the training cohort, the AUC values over one, two, and three years for determining the degree to which the Riskscore was accurate as a predictive marker were 0.794, 0.849, and 0.879, correspondingly (Fig. 4G); in the validation cohort, the AUC values for 1-, 2-, and 3- year were 0.622, 0.679, and 0.723, correspondingly (Fig. 4H). Based on these findings, it appears that the risk model that was created has great potential in predicting the prognosis of patients who have IPF.

Finally, the CIBERSORT algorithm was utilized to determine the variations in immune cell expression levels between patients having IPF to assess the association between the Riskscore and the infiltration status of immune cells in the BALF of these patients (Fig. 5A). Patients who had elevated scores exhibited substantially elevated levels of activated mast cells, monocytes, and activated NK cells, whereas the levels of resting mast cells, naive B cells, resting dendritic cells, M2 macrophages, activated memory CD4 T cells, and CD8 T cells were relatively lower. Analyses of correlations suggested that naive B cells (Fig. 5B; $r = -0.249$, $p = 0.008$), CD8 T cells (Fig. 5C; $r = -0.218$, $p = 0.021$), activated memory CD4 T cells (Fig. 5D; $r = -0.257$, $p = 0.006$), M2 macrophages (Fig. 5G; $r = -0.317$, $p < 0.001$), resting dendritic cells (Fig. 5H; $r = -0.432$, $p < 0.001$), and resting mast cells (Fig. 5I; $r = -0.272$, $p = 0.004$) were inversely linked to the Riskscore, whereas activated NK cells (Fig. 5E; $r = 0.327$, $p < 0.001$), monocytes (Fig. 5F; $r = 0.252$, $p = 0.007$), and activated mast cells (Fig. 5J; $r = 0.464$, $p < 0.001$) were positively correlated. These findings demonstrated that the risk model that was constructed had a close link to the infiltration status of immune cells that were observed in the BALF of individuals diagnosed with IPF.

# Established risk model's independence

Next, we examined the link between the Riskscore and clinical characteristics and verified the independence of the risk model by performing subgroup and regression analyses. There were no variations between patients of various ages (Fig. 6A) and sex (Fig. 6B) regarding Riskscore, demonstrating that no correlation existed between Riskscore and clinical parameters. After the patients had been regrouped depending on age (Fig. 6C, D) and sex (Fig. 6E, F), the risk model continued to demonstrate powerful prediction capacity, and patients exhibiting a lower Riskscore had a more favorable prognosis. Additionally, univariate and multivariate Cox regression analyses illustrated that the developed risk model independently served as a predictive indicator of the prognosis among individuals diagnosed with IPF (Table 2). These findings revealed that the risk model that was created had great independence in the prediction of IPF patients' prognoses.

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.

# Development of a Predictive Nomogram

A prognostic nomogram was designed to estimate the prognosis risk of individuals diagnosed with IPF by measuring their chance of survival over 1, 2, and 3 years after diagnosis (Fig. 7A). After that, we used the training and validation cohorts to verify whether the nomogram was accurate. The C-index, the calibration curve (Figs. 7B-G), as well as the decision curve analysis **(Figure S2)** all revealed a satisfactory level of accuracy of the diagnostic nomogram model. Within the training cohort, the C-index for the nomogram attained 0.7642 (95%CI: 0.7125–0.7962) and 0.6906 (95%CI: 0.5578–0.8234) for the validation cohort. The actual OS rate over 1, 2, and 3 years was very close to the predicted OS rate in the training cohort (Fig. 7B-D), and similar findings were shown in the validation cohort (Fig. 7E-G). Based on these findings, it was established that the integrated nomogram could reliably anticipate the patients' prognoses.

# Discussion

IPF is a condition that worsens with time, is fatal, has few therapy choices, inadequate therapeutic outcomes, and an unfavorable prognosis. Notwithstanding the booming development of multiple diagnoses and treatment strategies, the survival rate of patients with IPF has not been improved. IPF is characterized by a clinical course that is highly variable and prone to unpredictability; hence, improved risk assessment methodologies and tailored targeted therapy strategies are necessary (32).

In this research, we found two molecular clusters that are considerably distinct from one another premised on the expression of genes associated with FA metabolism. Enrichment analyses revealed different immune and metabolism actions between the two clusters. Further immune analyses identified several immune cells in BALF potentially associated with prognosis in IPF. Additionally, we developed a predictive risk model that was on the basis of FAMRGs and was able to correctly predict the prognosis of

patients who were diagnosed with IPF. Our findings could help in creating therapies that specifically target IPF.

The use of consensus clustering as a method for categorizing samples into various groups premised on the gene expression matrix was shown to be reliable. We initially identified two molecular clusters using consensus clustering based on the FAMRGs expression in patients diagnosed with IPF and discovered that these two molecular clusters had remarkably different OS rates. These findings provided further evidence that the FA metabolic subtypes of a patient's IPF influenced their prognoses and that the consequent construction of prediction models premised on FAMRGs was reliable.

Next, enrichment studies between the two clusters were carried out so that the inherent biological processes could be investigated. GSEA is a classic approach for incorporating gene expression data. Using this method, the expression pattern of gene sets in various groups may be elucidated directly. (33). First, GSEA findings illustrated an enrichment of immune-related pathways, including natural killer cell-mediated cytotoxicity, the B cell receptor signaling pathway, the Toll-like receptor signaling pathway, leukocyte transendothelial migration, the chemokine signaling pathway, and the Nod-like receptor signaling pathway, in cluster 1, which indicated poor OS. Based on these findings, it seems that an immune system that is dysregulated may be a potential mediator of the function that FA metabolism plays in the onset and progression of IPF. Mechanistically, the association between FA metabolism and immune dysregulation may be through epigenetic regulation, such as DNA methylation influenced by genetic variation (34). Notably, the results showed that clusters 1 and 2 had specific metabolism signatures. The pathways enriched in cluster 1 were predominantly linked to the metabolism of amino acids, including arginine, proline, and histidine metabolism, and other 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. Considering that the classification was based on FA metabolism relevant genes, the result suggested that there was crosstalk between FA metabolism and other nutrient metabolisms that affected the process of IPF and were worth further exploration.

As mentioned previously, immunology performs an integral function in the onset and progression of IPF, which is also associated with FA metabolism. CIBERSORT is a biological information analysis tool that evaluates the expression levels of immune cells premised on RNA-seq and obtains different immune cell ratios from samples. This algorithm is extensively utilized to examine the infiltration of immune cells in various human immune-related diseases, such as tumors, osteoarthritis, and lupus nephritis (35–37). Li et al. established a hypoxia immune-related prediction model for determining the prognosis among IPF patients using CIBERSORT (38); therefore, we implemented CIBERSORT to ascertain the infiltration levels of immune cells between the two clusters. We identified several BALF immune cells that had the potential to be associated with IPF prognosis, mainly involving activated mast cells, naive B cells, resting mast cells, M0 and M2 macrophages, monocytes, activated NK cells, resting dendritic cells, and resting NK cells. Activated NK cells, activated mast cells, and monocytes showed increased infiltration in cluster 1

patients with worse survival. Similarly, in the constructed risk model, these three immune cell populations also had an elevated infiltration level in the high-risk group with poor prognosis and were positively correlated with Riskscore. Earlier research findings have shown that individuals with pulmonary fibrosis exhibit a greater infiltration level of NK cells in BALF in contrast with those who have sarcoidosis (39). Scott et al. indicated that elevated circulating monocyte count could be a cellular biomarker for poor outcomes in IPF (40). According to the findings of Kawanami et al., individuals with fibrotic lung disease had significantly higher numbers of mast cells in their lungs. These mast cells are often localized around the thickened regions of the alveolar septa and are located near aberrant 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.

Synthesizing the findings of this study, it is reasonable to infer that FA metabolism-related dysregulation, resulted in the disorder of immune and metabolism status in BALF, hence contributing to unsatisfactory IPF prognosis. As aforementioned, the reprogramming of the FA metabolism was identified as a distinctive characteristic of a grim prognosis in individuals with IPF. To additionally verify the influence of FA metabolic disorders in IPF and examine the prognostic significance of FAMRGs in patients diagnosed with IPF, we created a prognostic risk model using FAMRGs, and then we confirmed it using a separate validation cohort. It has been shown that the five genes that were employed for the establishment of the risk model in this research are remarkably related to the onset and progression of IPF. GGT5 encodes gamma-glutamyl transferase 5, which cleaves glutathione peptides to maintain the glutathione balance in the human body (42). Prior studies have reported that mice lacking gamma-glutamyl transpeptidase had less severe cases of bleomycin-induced pulmonary fibrosis (43). However, the role of GGT5 in FA metabolism and the progression of IPF is unknown. The branched-chain acyl-CoA oxidase that is encoded by the ACOX2 gene is a peroxisomal enzyme that is thought to play a role in the metabolism of bile acid intermediates as well as branched-chain FAs. A deficiency in ACOX2 has been shown to be linked to an increased risk of developing liver fibrosis (44), but its role in pulmonary fibrosis is unknown. CYP4F3, also referred to as leukotriene-B4 omega-hydroxylase, consists of enzymes CYP4F3A and CYP4F3B that are responsible for the metabolism of leukotriene B4 and most probable 5-hydroxyeicosatetraenoic acid through an omega oxidation reaction. This results in the inhibition and deterioration of well-recognized inflammatory markers (45, 46). Thus, CYP4F3 is considered to be associated with inflammation-related diseases, such as inflammatory bowel disease (IBD) (47), but the role of CYP4F3 in IPF is unknown. HACD4 encodes 3-hydroxyacyl-CoA dehydratase 4, which is involved in FA elongation and the biosynthetic process of very-long-chain FA; therefore, its role in IPF is worth exploring. Although ODC1 is known to encode the rate-limiting enzyme of the polyamine biosynthetic pathways, which acts as a catalyst of ornithine to putrescine, its significance in IPF is only partially recognized. Survival analysis illustrated that the developed risk model showed outstanding prediction capacity for the survival of patients diagnosed with IPF in both cohorts. The patients' prognoses could be determined independently by each of the five genes, and independence and subgroup analyses demonstrated that the FAMRGs-based risk model can independently anticipate the prognosis of IPF, regardless of age and sex. Ultimately, a nomogram that incorporates both the risk score and the clinical

characteristics was developed, calibrated, and tested; the results revealed that it had a powerful predictive power regarding survival. Taken together, these findings provided more evidence that FAMRGs have a prognostic predictive role in IPF.

Owing to the poor prognosis, variability, and unpredictability of the clinical course of IPF, there is a need for a method of efficient risk classification as well as a treatment plan that emphasizes personalized targeting. Our research demonstrates highlights in comparison to earlier studies. First, our research was on the basis of FA metabolism, a topic that has been receiving more and more interest in IPF-related research. Using consensus clustering, we discovered two molecular clusters with distinct prognoses and immunological statuses. Second, we examined the biological processes based on the clustering findings, and we partially elucidated the fundamental mechanisms. Thirdly, we elucidated the function that the metabolism of FA plays in influencing the infiltration level of immune cells in BALF as well as the prognosis. Moreover, the role of each of the five selected genes is unclear in IPF, and further exploration is necessary. Our study offers a good theoretical foundation for future IPF research.

Nevertheless, our study also has several limitations. First, we could not establish the involvement of FAMRG in the progression of IPF because there was insufficient data regarding the pulmonary function of the individuals who had IPF. Second, our findings were derived from a bioinformatics study, which was not followed up with further experimentations to confirm its validity. Thirdly, the data that were utilized in this research were obtained by downloading them from a free-access database since our clinical practice had an insufficient number of patients diagnosed with IPF. Consequently, to additionally evaluate the performance of the predictive model and the mechanisms of the five FAMRGs in the pathogenic process behind IPF, a prospective cohort study as well as molecular biology tests ought to be devised and conducted.

# Conclusions

In this research, consensus clustering was utilized to identify two distinct molecular clusters premised on the FAMRGs in IPF. The results of the enrichment and immune analyses showed that dysregulation of FA metabolism contributes to disorders of the immune system and nutrition metabolism, resulting in an unfavorable prognosis. Our research may provide unique insights into the creation of novel targeted treatments, offer a theoretical basis for personalized treatment plans, and make it easier to classify people who have IPF according to their risk levels.

# Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Availability of data and materials

All the data corresponding to the IPF series used in this study are available in GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70866), which is public functional genomics data repository.

### Competing interests

The authors declare that they have no competing interests.


### 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).


### Author Contributions

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


### Acknowledgements

We gratefully acknowledge Gene Expression Omnibus for data availability.


## References

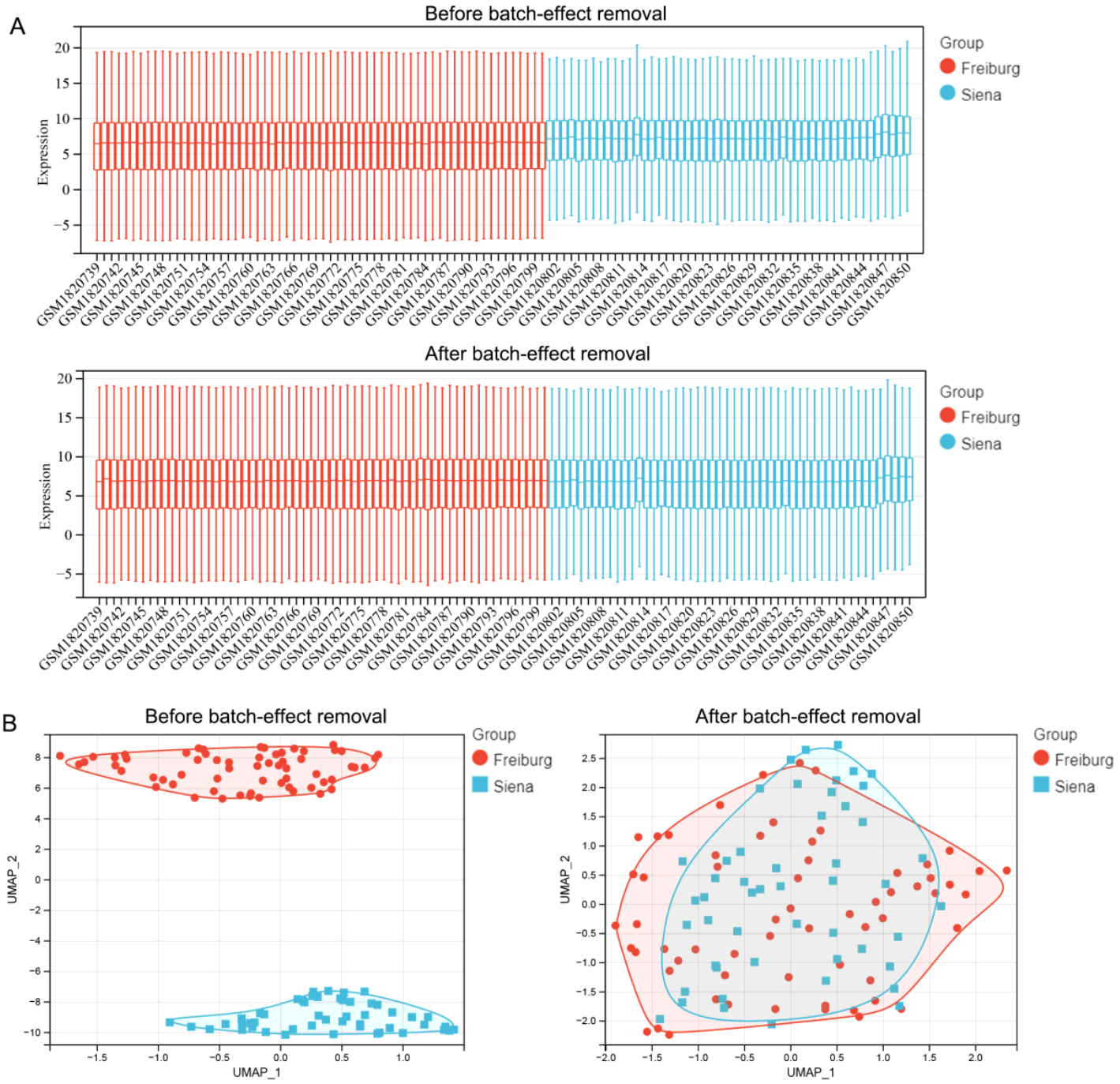

1. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am j resp crit care. 2011;183(6):788–824.
2. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. New engl j med. 2018;378(19):1811–23.
3. Shin KH, Shin SW. Idiopathic Pulmonary Fibrosis. New engl j med. 2018;379(8):795–6.
4. Mora AL, Rojas M, Pardo A, Selman M. Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. Nat rev drug discov. 2017;16(11):755–72.
5. Wang ZN, Tang XX. New Perspectives on the Aberrant Alveolar Repair of Idiopathic Pulmonary Fibrosis. Front Cell Dev Biol. 2020;8(null):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(null):107798.

7. Kreuter M, Bonella F, Wijsenbeek M, Maher TM, Spagnolo P. Pharmacological Treatment of Idiopathic Pulmonary Fibrosis: Current Approaches, Unsolved Issues, and Future Perspectives. Biomed res int. 2015;2015(null):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. Respirology. 2017;22(6):1171–8.
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.
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.
11. Sunaga H, Matsui H, Ueno M, Maeno T, Iso T, Syamsunarno MR, et al. Deranged fatty acid composition causes pulmonary fibrosis in Elovl6-deficient mice. Nat Commun. 2013;4(null):2563.
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(null):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(null):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.
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.
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.
17. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. Lancet. 2017;389(10082):1941–52.
18. Lin Y, Xu Z. Fibroblast Senescence in Idiopathic Pulmonary Fibrosis. Front Cell Dev Biol. 2020;8(null):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.
20. Liu W, Meridew JA, Aravamudhan A, Ligresti G, Tschumperlin DJ, Tan Q. Targeted regulation of fibroblast state by CRISPR-mediated CEBPA expression. Respir Res. 2019;20(1):281.
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.
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.
23. Zhao H, Chan-Li Y, Collins SL, Zhang Y, Hallowell RW, Mitzner W, et al. Pulmonary delivery of docosahexaenoic acid mitigates bleomycin-induced pulmonary fibrosis. BMC Pulm Med. 2014;14(null):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.
25. Kim HS, Yoo HJ, Lee KM, Song HE, Kim SJ, Lee JO, et al. Stearic acid attenuates profibrotic signaling in idiopathic pulmonary fibrosis. Respirology. 2021;26(3):255–63.
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.
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 Bioinformatics. 2012;13(null):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.
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.
30. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3.
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.
32. Martinez FJ, Safrin 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(12 Pt 1):963–7.
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.
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.
35. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor-Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol. 2018;1711(null):243–59.
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): null.
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.

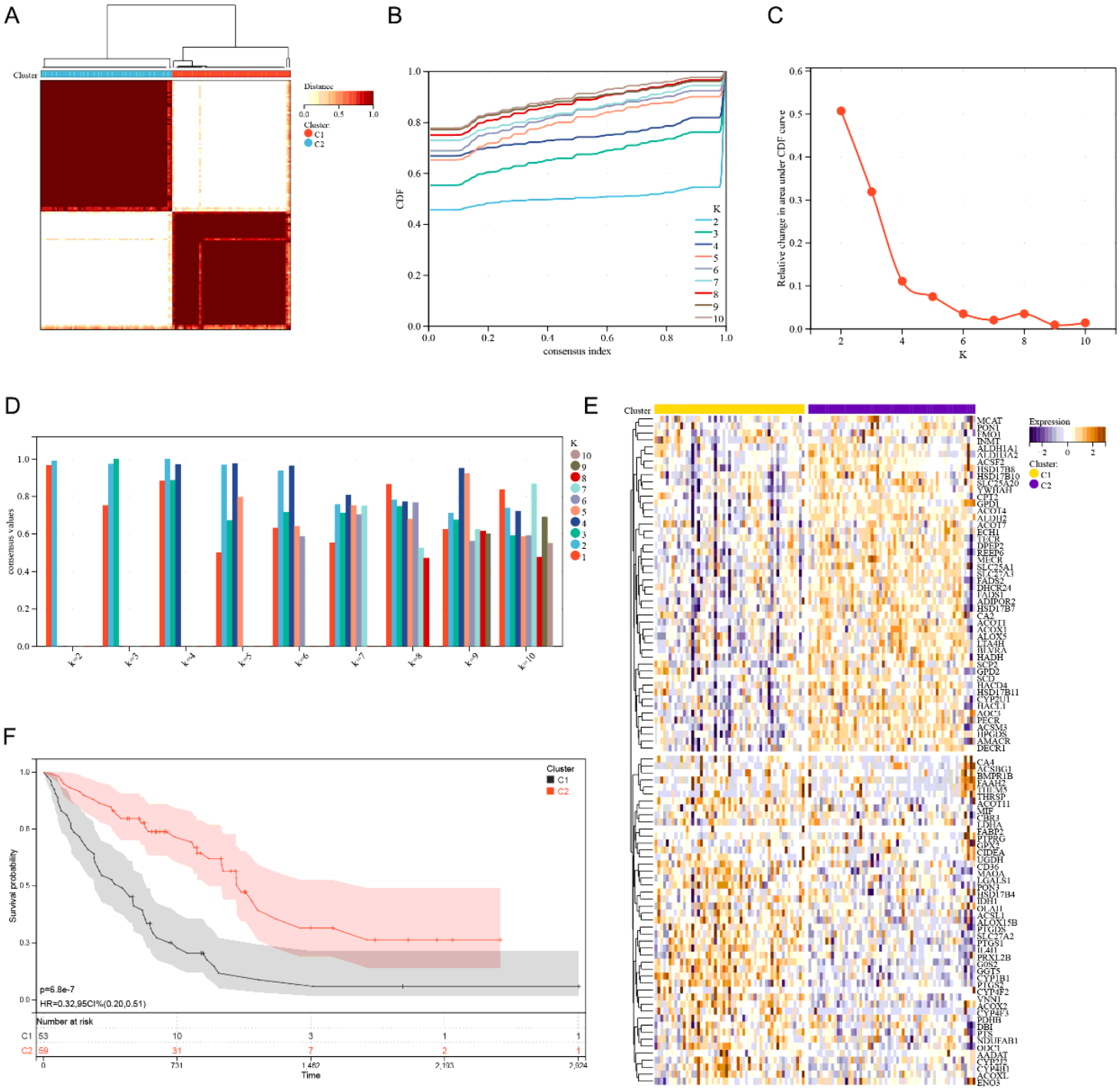
38. Li X, Cai H, Cai Y, Zhang Q, Ding Y, Zhuang Q. Investigation of a Hypoxia-Immune-Related Microenvironment Gene Signature and Prediction Model for Idiopathic Pulmonary Fibrosis. Front Immunol. 2021;12(null):629854.
39. Bergantini L, Cameli P, d'Alessandro M, Vagaggini C, Refini RM, Landi C, et al. NK and NKT-like cells in granulomatous and fibrotic lung diseases. Clin exp med. 2019;19(4):487–94.
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.
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.
43. Pardo A, Ruiz V, Arreola JL, Ramírez 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.
44. Vilarinho S, Sari S, Mazzacuva 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.
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.
46. Powell WS, Rokach J. Biosynthesis, biological effects, and receptors of hydroxyeicosatetraenoic acids (HETEs) and oxoeicosatetraenoic acids (oxo-ETEs) derived from arachidonic acid. Biochim biophys acta. 2015;1851(4):340–55.
47. Scaioli 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): null.

# Figures



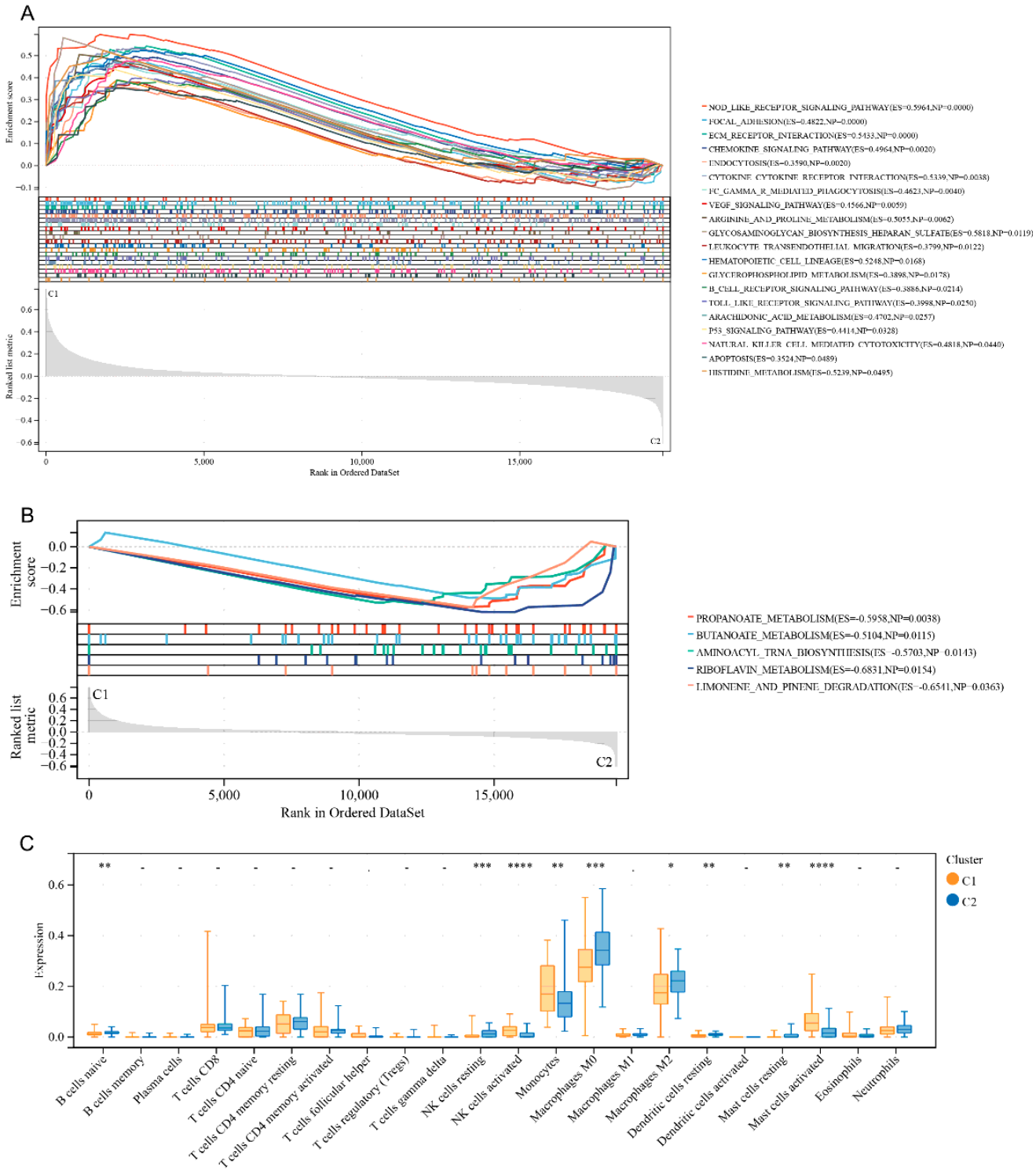

## Figure 1

**Comparison of expression data distribution and 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.



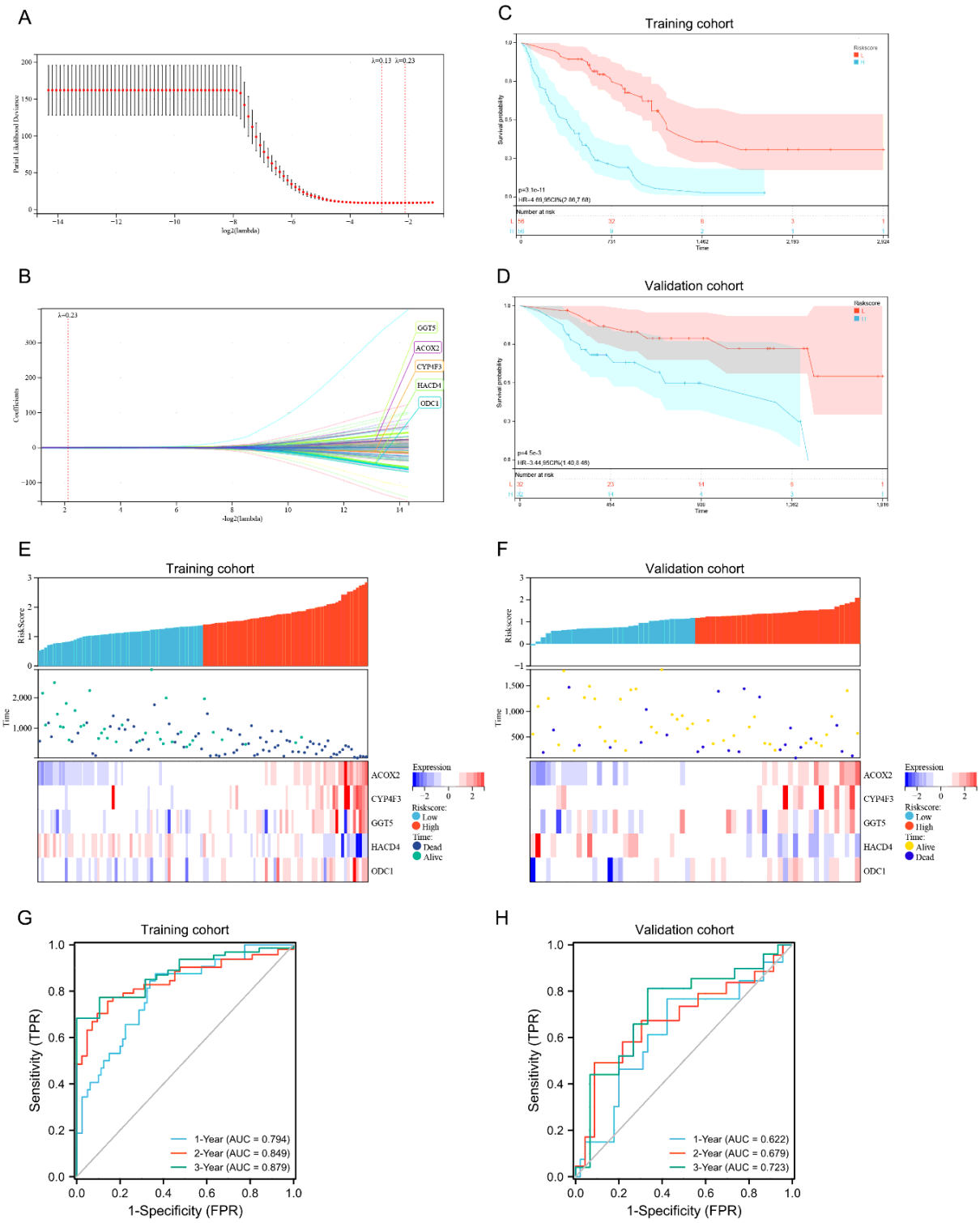

## Figure 2

**Consensus cluster. (A–D):** K = 2 was identified as the optimal value for consensus clustering. **(E)** Heatmap showing the expression of fatty acid metabolism-related genes in the two clusters. **(F)** Survival curve of the patients in the two clusters.




## Figure 3

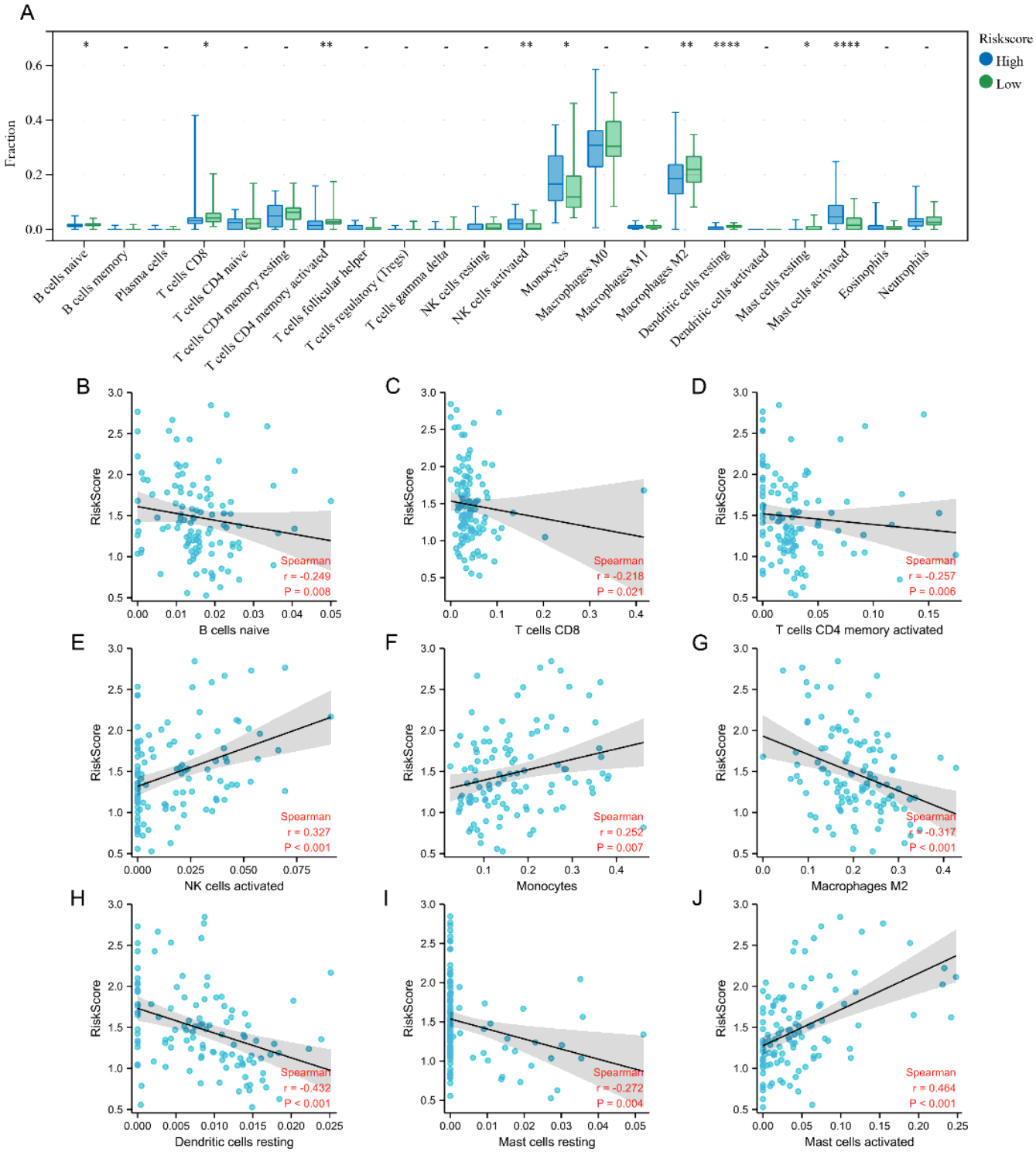
**Enrichment analyses and immune infiltration characteristics of the two identified subgroups with distinct prognoses.** (**A**) GSEA for cluster 1 (C1) and cluster 2 (C2). Immune and metabolism-related pathways were significantly enriched in cluster 1, (**B**) and another five metabolism-related pathways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




## Figure 4

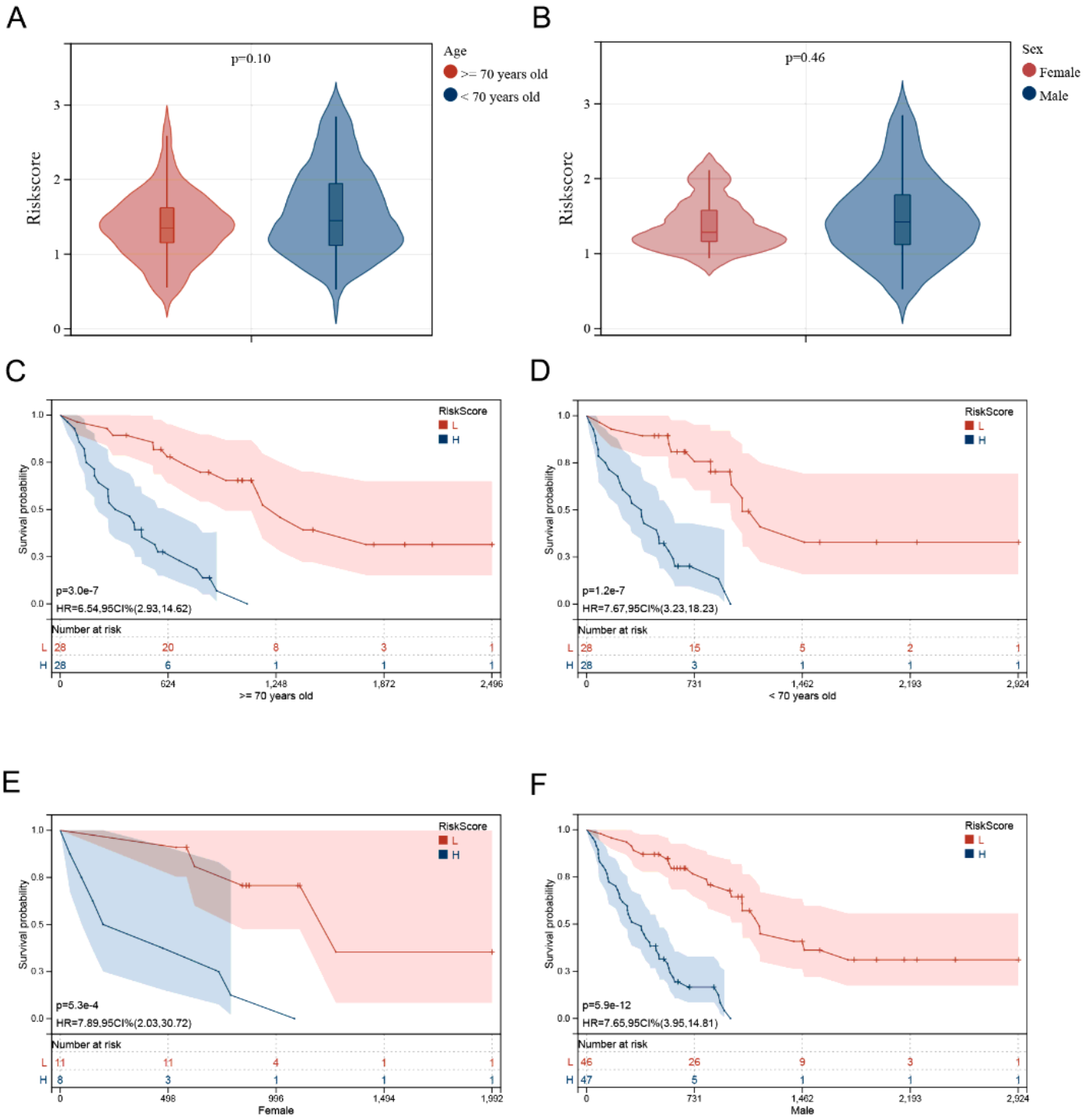
**Development and assessment of a prognostic fatty acid metabolism-related signature for patients with IPF. (A)** and **(B)** LASSO analysis with optimal lambda. **(C)** and **(D)** Kaplan–Meier survival analysis for patients with high- and low- Riskscore in the training (C) and validation cohorts (D), patients with IPF in

the high group had worse prognostic survival than those in the low group. **(E)** and **(F)** Distribution of the Riskscore, survival status, and the expression of the five candidate genes in the high- and low- risk groups in the training (E) and validation (F) cohorts. **(G)** and **(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.



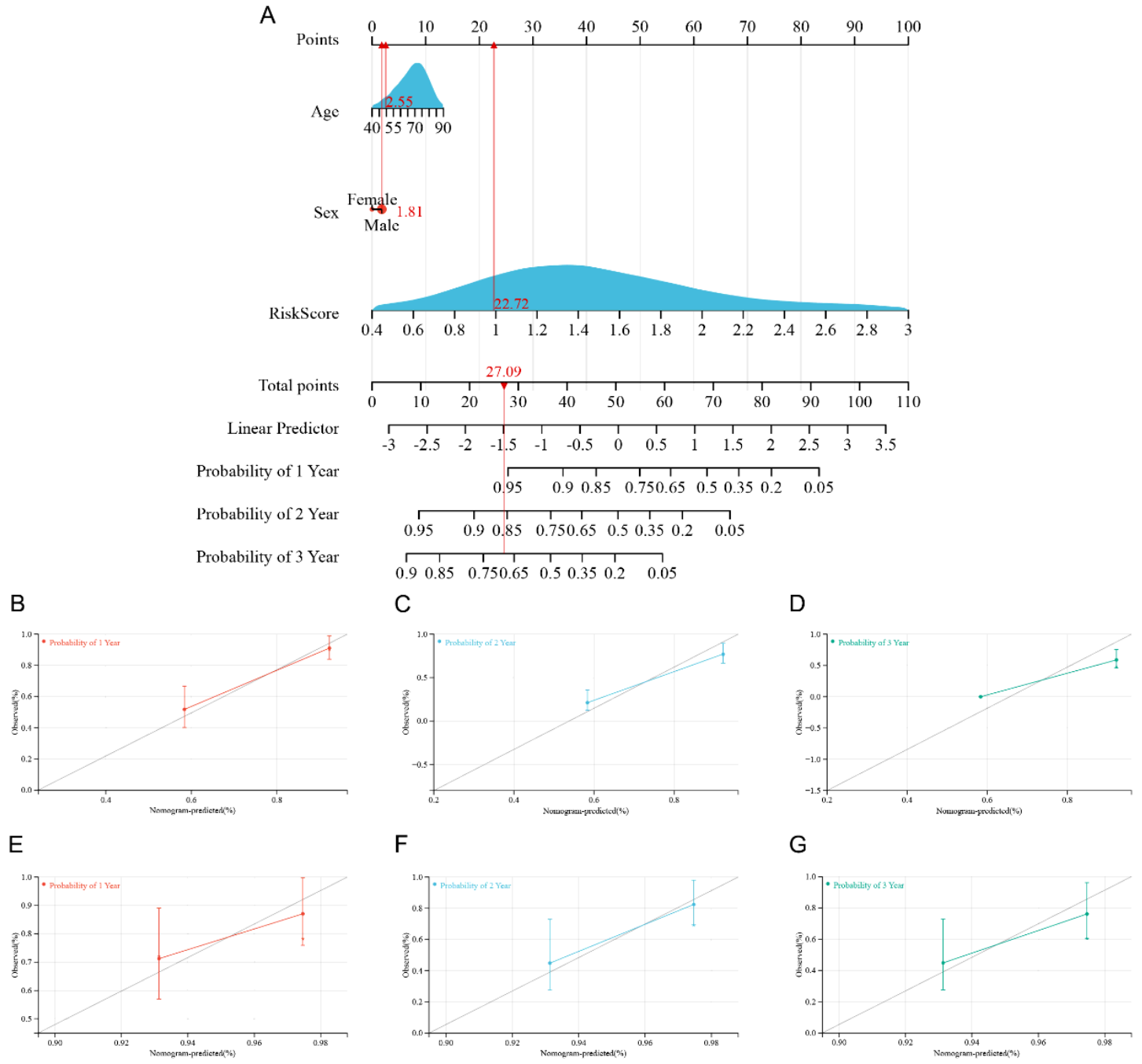
## Figure 5

**Estimation of bronchoalveolar lavage fluid immune cell infiltration by the prognostic signature. (A)** Infiltrating levels in the high- and low- Riskscore groups using CIBERSORT algorithms. The plots show the differences in immune infiltration score between the high- and low- Riskscore groups. **(B–J)** The Spearman correlation between Riskscore and the fraction of BALF immune cells is shown. Scatterplots show that the Riskscore is 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 it is positively correlated with activated NK cells (E), monocytes (F), and activated mast cells (J).




## Figure 6

**Association between Riskscore and clinical characteristics.** No significant difference was identified in patients with different ages **(A)** and sex **(B)**. Independence analysis of the risk model **(C–F)**. Survival curve of patients regrouped according to age (C, D) and sex (E, F)




## Figure 7

**Construction and calibration of the nomogram. (A)** Nomogram integrating risk score and clinical features, **(B–D)** calibration of the nomogram at 1, 2, and 3 years in the training cohort, and **(E–G)** calibration of the nomogram at 1, 2, and 3 years in the verification cohort.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx