Original Article

A novel prognostic signature for idiopathic pulmonary fibrosis based on five-immune-related genes

Lingxiao Qiu[1,2,3]^, Gencheng Gong[4], Wenjuan Wu[5], Nana Li[1]^, Zhaonan Li[6]^, Shanshan Chen[1,3], Ping Li[1,3,7], Tengfei Chen[1], Huasi Zhao[1], Chunling Hu[8], Zeming Fang[9], Yan Wang[1], Hongping Liu[1], Panpan Cui[10]^, Guojun Zhang[1,3,7]^

[1]Department of Respiratory Medicine, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; [2]Academy of Medical Sciences, Zhengzhou University, Zhengzhou, China; [3]Henan Provincial Medical Key Laboratory for Interstitial Lung Disease and Lung Transplantation, Zhengzhou, China; [4]State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China; [5]Department of Geriatric Medicine, Henan Provincial People's Hospital, Zhengzhou, China; [6]Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; [7]Zhengzhou Key Laboratory for Chronic Respiratory Disease, Zhengzhou, China; [8]Department of Respiratory Intensive Care Unit, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; [9]Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; [10]School of Nursing and Heath, Zhengzhou University, Zhengzhou, China

Contributions: (I) Conception and design: L Qiu, G Gong; (II) Administrative support: G Zhang; (III) Provision of study materials or patients: L Qiu, C Hu; (IV) Collection and assembly of data: L Qiu, W Wu, Z Li; (V) Data analysis and interpretation: L Qiu, G Gong, N Li, P Cui; (VI) Manuscript writing: All authors; (VII) Final approval of the manuscript: All authors.

Correspondence to: Guojun Zhang. The First Affiliated Hospital of Zhengzhou University, No. 1 Jianshe East Road, Zhengzhou 450000, China. Email: zgj@zzu.edu.cn.

Background: Idiopathic pulmonary fibrosis (IPF) is a highly fatal lung disease of unknown etiology with a median survival after diagnosis of only 2–3 years. Its poor prognosis is due to the limited therapy options available as well as the lack of effective prognostic indicators. This study aimed to construct a novel prognostic signature for IPF to assist in the personalized management of IPF patients during treatment.

Methods: Differentially-expressed genes (DEGs) in IPF patients versus healthy individuals were analyzed using the "limma" package of R software. Immune-related genes (IRGs) were obtained from the ImmPort database. Univariate Cox regression analysis was adopted to screen significantly prognostic IRGs for IPF patients. Multiple Cox regression analysis was used to identify optimal prognostic IRGs and construct a prognostic signature.

Results: Compared with healthy individuals, there were a total of 52 prognosis-related DEGs in the bronchoalveolar lavage (BAL) samples of IPF patients, of which 37 genes were identified as IRGs. Of these, five genes (*CXCL14*, *SLC40A1*, *RNASE3*, *CCR3*, and *RORA*) were significantly associated with overall survival (OS) in IPF patients, and were utilized for establishment of the prognostic signature. IPF patients were divided into high- and low-risk groups based on the prognostic signature. Marked differences in the OS probability were observed between high- and low-risk IPF patients. The area under curves (AUCs) of the receiver operating characteristic (ROC) curve for the prognostic signature in the training and validation cohorts were 0.858 and 0.837, respectively. The expression levels between *RNASE3* and *SLC40A1* (P<0.01, r=0.394), between *RORA* and *CXCL14* (P<0.01, r=–0.355), between *CCR3* and *CXCL14* (P<0.01, r=0.258), as well as between *RNASE3* and *CCR3* (P<0.01, r=0.293) were significantly correlated.

Conclusions: We developed a validated and reproducible IRG-based prognostic signature that should be helpful in the personalized management of patients with IPF, providing new insights into the relationship between the immune system and IPF.

^ ORCID: Lingxiao Qiu, 0000-0002-1703-0797; Nana Li, 0000-0003-2882-7045; Zhaonan Li, 0000-0003-3055-8817; Panpan Cui, 0000-0002-1587-7863; Guojun Zhang, 0000-0003-1756-0082.

Page 2 of 13

Qiu et al. A prognostic signature for idiopathic pulmonary fibrosis

Introduction

Idiopathic pulmonary fibrosis (IPF) is a deadly interstitial lung disorder of unknown etiology (1). It is characterized by irreversible fibrogenesis in the lung parenchyma, leading to progressive respiratory function failure and eventually death (2,3). IPF is the most common interstitial lung disease and has the worst prognosis in pulmonary fibrosis (4). Nearly half of IPF patients die within 2–3 years after diagnosis (3,4), and the 5-year survival rate is less than 30% (5). IPF is a highly heterogeneous disease with a greatly variable natural history (6,7). The course of this disease in an individual patient is difficult to predict (4,8); some patients with IPF experience rapid decline, while others experience much slower development (3,8). For a long time, the lack of effective prognostic indicators has made it difficult to accurately track and evaluate the prognosis of IPF, which has led to the poor prognosis of IPF to a certain extent. Hence, the development of applicable prognostic signatures is urgently needed for the clinical treatment of IPF.

The pathophysiological pathogenesis of IPF involves aberrant transcription and gene expression (9-14). Molecular genomic features based on lung tissue have been used to predict the development of IPF (15,16). Though previous studies have identified some genes and pathways may play an important role in the occurrence and development of IPF, and may be expected to be biomarkers or therapeutic targets for the diagnosis of IPF (17,18). However, the lack of verification of survival information is the biggest short board in these papers. Meanwhile, the resources required to perform a lung biopsy and the risks associated with the procedure limit the applicability of such genomic features. Molecular models have also been established based on peripheral blood mononuclear cell (PBMC) transcription profile data to predict the disease state of IPF (19,20). However, in the absence of lung biopsies, it is difficult to explain the correlation between abnormal PBMC transcription and pulmonary fibrosis course. Bronchoalveolar lavage (BAL) is a method of obtaining alveolar surface lining fluid with fiberoptic bronchoscopy for evaluating inflammation, immune cells, and soluble substances. BAL plays a vital role in assisting IPF diagnosis and has been recommended as the auxiliary diagnostic reference by the American Thoracic Society (ATS) (21). The advantages of utilizing the gene expression profiles of BAL cells to depict the molecular features of IPF include lung localization, ease of accessibility, and dynamic assessment of disease status through longitudinal sample collection. Previous studies have revealed that Innate and adaptive immune responses disorders possess an important role in the pathogenesis of lung fibrosis (22). The differentially-expressed immune-related genes (IRGs) also have been reported associated with the development of IPF (23,24). The immPort database is funded by the National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID), Health and Human Services (HHS) in support of the NIH mission to share data with the public. It provides information about the immune-related genes of humans. Therefore, using the GSE70866 gene expression data set of the Gene Expression Synthesis (GEO) database and the IRGs list of the ImmPort database, we aim to combine the survival information of IPF patients to establish a new molecular genome feature screening from IRGs, to predict the prognosis of IPF patient. We present the following article in accordance with the STARD reporting checklist (available at https://dx.doi.org/10.21037/atm-21-4545).

Methods

Acquisition and analysis of datasets

Microarray profile data from the GSE70866 gene expression dataset were downloaded from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database. The platform was a GPL14550 Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Agilent Technologies Inc., California, U.S.). A total of 132 BALF samples, including 20 samples from healthy individuals and 112 samples from IPF patients, were used to analyze the microarray data. All 112 IPF patients had detailed sociodemographic characteristics and complete survival information. The study was conducted in accordance with

the Declaration of Helsinki (as revised in 2013).

The criteria of differentially-expressed genes (DEGs) and differentially-expressed immune-related genes (IRGs)

The filtration of DEGs was performed in 112 IPF patients versus healthy individuals. In this study, DEGs between IPF and healthy individuals were defined using a log2 fold change (FC) >1 and an adjusted P value (adj. P) <0.05 as thresholds. A total of 1,793 IRGs were obtained from the ImmPort (https://www.immport.org/shared/genelists) database. Taking the intersection through the Venn algorithm (http://bioinformatics.psb.ugent.be/webtools/Venn/), 52 differentially-expressed IRGs were filtered, which remained and were used as candidates for subsequent analysis.

Construction and validation of the prognostic IRG-based signature

The 112 included patients were randomly divided into a training cohort (50%) and validation cohort (50%) using the random numbers method. The construction of prognostic gene-based signatures was carried out in the training cohort, and verification was performed in the verification cohort. Univariate Cox regression analysis was used to screen for immune genes that were significantly associated with prognosis, with a cut-off of P<0.05. Next, multivariate Cox-regression analysis was performed on the training cohort to further determine the best prognostic IRG signature using the "survival" package (URL: https://github.com/therneau/survival) in R software (version 4.0.3) (URL: https://cran.r-project.org/mirrors.html), with a cut-off of P<0.05. The formula of IPF patient's risk score was established as follows: score = sum (each gene's expression × corresponding coefficient). The patients were stratified into high-risk and low-risk groups based on the median value of the risk score. Based on the risk score groups, survival differences between high-risk and low-risk groups were carried out with the "survival" R package (URL: https://github.com/therneau/survival).

Statistical analysis

Baseline characteristics such as age, sex, race, days to death, and vital status were collected. Continuous variables were reported as the mean (± standard deviation) and compared using the Student's *t*-test. Categorical variables were reported as counts n (%) and compared using the chi-square test. The comparison of sociodemographic features between the training and validation cohorts was carried out using GraphPad Prism (version 7.0; GraphPad Software, La Jolla, CA, USA).

The other statistical analyses were carried out using R software (version 4.0.3) (URL: https://cran.r-project.org/mirrors.html) and considered significant when the corresponding P<0.05. The adjusted P<0.05 was used for screening DEGs, and P<0.05 was used as a significance threshold in the remaining statistical analyses. The analysis of DEGs was conducted by utilizing the "limma" package (URL: http://www.strimmerlab.org/software/st/). Univariate Cox regression analysis was used to screen for DEGs that were significantly associated with overall survival (OS). Multivariate Cox regression analysis was performed on the training cohort to further determine the best prognostic IRG model. A multivariate Cox regression model was conducted for the variables with P<0.05 in the univariate analyses. A gene-based signature was built with the coefficients of each factor in the multivariate Cox analysis. The "survival" package (URL: https://github.com/therneau/survival) calculated the survival curve function, and the "survminer" package (URL: https://mirror.lzu.edu.cn/CRAN/bin/windows/contrib/4.0/survminer_0.4.9.zip) executed the visualization. The heat map was drawn using the "pheatmap" (pretty heatmap) package (URL: https://mirror.lzu.edu.cn/CRAN/bin/windows/contrib/4.0/pheatmap_1.0.12.zip). The volcano map was drawn using the "ggplot2" package (URL: https://cran.r-project.org/web/packages/ggplot2movies/index.html).

Results

Baseline characteristics of patient with IPF

Table 1 summarizes the sociodemographic information of the included IPF patients. A total of 112 IPF patients were identified, with a median age of 69.5 (±10.1) years. IPF was more common in older populations (67.0% of patients were older than 65 years versus 33.0% of patients less than 6 years). The incidence of IPF was higher in men than in women (83.0% male patients versus 17.0% female patients).

These 112 IPF patients were randomly divided into training (50%) and validation (50%) cohorts, with 56 patients in each group. No significant differences between the two cohorts were observed in terms of age, sex, days to death, and vital status (P>0.05). Qualified

Page 4 of 13

Qiu et al. A prognostic signature for idiopathic pulmonary fibrosis

Table 1 The sociodemographic information of patients

Characteristics	Total (n=112)	Training cohort (n=56)	Validation cohort (n=56)	P value
Age, mean (± SD)	67.97 (±10.1)	67.0 (±10.4)	69.0 (±9.7)	0.300
Age, n (%)				
<65	37 (33.0)	18 (32.1)	19 (34.0)	
≥65	75 (67.0)	38 (67.9)	37 (66.0)	0.841
Gender, n (%)				
Female	19 (17.0)	7 (12.5)	12 (21.4)	
Male	93 (83.0)	49 (87.5)	44 (78.6)	0.208
Days to death, mean (± SD)	698.1 (±555.9)	656.7 (±551.9)	739.5 (±561.7)	0.433
Vital status, n (%)				
Alive	36 (32.1)	20 (35.7)	16 (28.6)	
Dead	76 (67.9)	36 (64.3)	40 (71.4)	0.418
Sample contact country, n (%)				
Germany	112 (100.0)	56 (100.0)	56 (100.0)	NA

SD, standard deviation.

survival information for all of the included IPF patients was available for further analysis.

Identification of DEGs

DEGs of the IPF and healthy individuals from the GPL14550 platform of the GSE70866 gene expression dataset were analysed using the "limma" package. In this dataset, a total of 379 DEGs met the criteria, of which 207 genes were upregulated and 172 genes were downregulated (Table S1). *Figure 1A* is a volcano map of 379 DEGs in the IPF group compared to the healthy individuals group. The profiling of all the DEGs is shown in *Figure 1B* and presented in the form of a non-cluster analysis expression heatmap. SPP1, PPBP, and MMP7 were the top three most significantly upregulated genes in the IPF group, while NALCN, C8B, and ITIH5 were the three most downregulated genes in the IPF group.

Identification of differential expression IRGs

Combining the results of DEGs (Table S1) and the IRGs from the ImmPort database, 52 differentially expressed IRGs were identified. A volcano map was constructed to present the differential expression of all IRGs (*Figure 2A*). *Figure 2B* shows the expression of the 52 differential IRGs

in the form of a heatmap. SPP1, PPBP, TUBB3, CCL2, and S100A12 were the five most significantly upregulated IRGs, while the top five downregulated IRGs were PTGER3, CD40LG, CAMP, IGF1, and CXCL9.

Prognostically relevant IRGs filtration

Prognostically relevant IRGs for IPF were selected based on the results of univariate Cox regression analysis. A forest plot was drawn to show the 37 obtained prognostically relevant IRGs, including prognostically protective IRGs such as *RORA* [hazard ratio (HR): 0.613, 95% confidence interval (CI): (0.474–0.794)] and ICOS [HR: 0.672, 95% CI: (0.560–0.809)] (*Figure 3*). Conversely, MPO [HR: 1.287, 95% CI: (1.139–1.454)], *RNASE3* [HR: 1.711, 95% CI: (1.338-2.188)], PDGFA [HR: 1.228, 95% CI: (1.030–1.465)], PPBP [HR: 1.154, 95% CI: (1.002–1.330)], and FABP3 [HR: 1.522, 95% CI: (1.216–1.905)] were prognostic factors of worse survival (*Figure 3*).

An IRGs prognostic model of IPF

Multivariate Cox regression analysis was performed based on 37 prognostic factors of OS to establish a model to predict the outcomes of IPF patients. *CXCL14, SLC40A1, RNASE3, CCR3*, and *RORA* were ultimately identified to
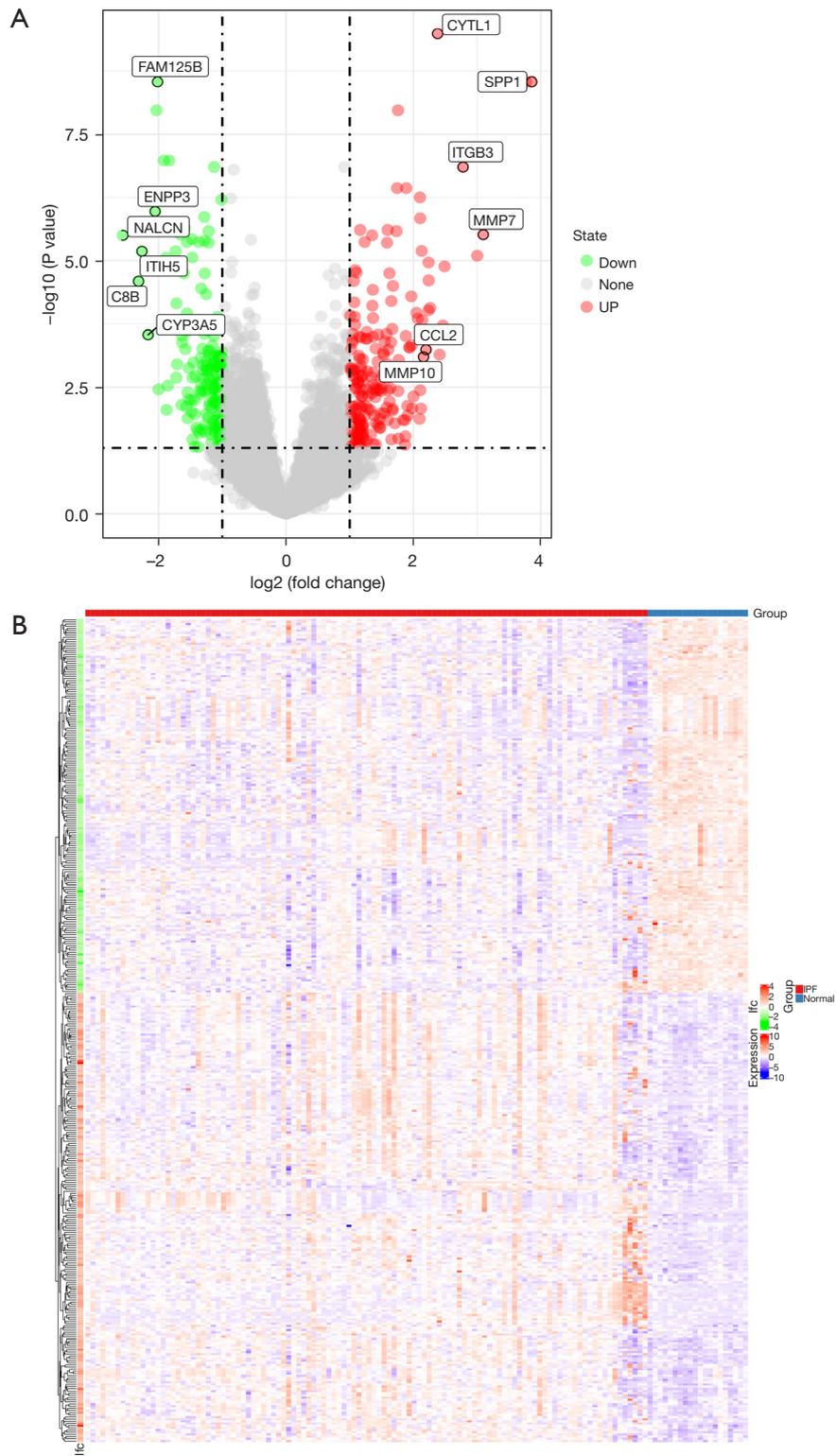
Figure 1 Comparison of the gene expression profile between the IPF group and the healthy individuals group. (A) Heatmap of significantly DEGs. (B) Volcano map of DEGs; red dots represent upregulated DEGs, grey dots represent non-differentially expressed genes, and green dots represent downregulated DEGs. IPF, idiopathic pulmonary fibrosis; DEGs, differentially-expressed genes.

Page 6 of 13

Qiu et al. A prognostic signature for idiopathic pulmonary fibrosis
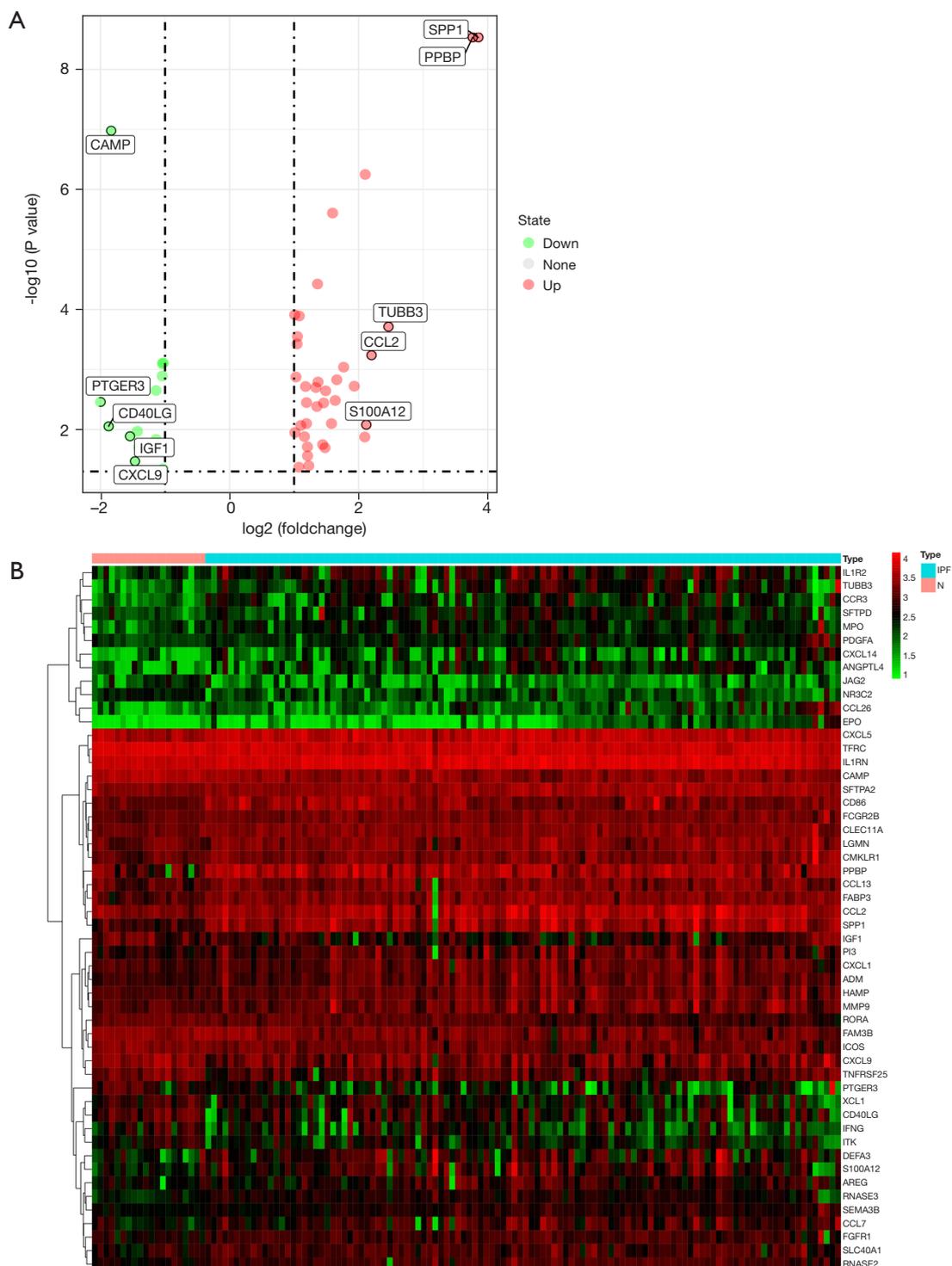
Figure 2 Comparison of the IRG expression profile between the IPF group and the healthy individuals group. (A) Heatmap of significantly differentially-expressed IRGs. (B) Volcano map of IRGs; red dots represent upregulated differentially expressed IRGs, grey dots represent non-differentially expressed IRGs, and green dots represent downregulated differentially expressed IRGs. IRG, immune-related gene; IPF, idiopathic pulmonary fibrosis.

	P value	Hazard ratio
LGMN	<0.001	1.482(1.262–1.740)
HAMP	<0.001	1.493(1.274–1.750)
PI3	<0.001	1.398(1.220–1.602)
CAMP	<0.001	0.640(0.514–0.797)
PPBP	0.047	1.154(1.002–1.330)
CXCL14	<0.001	1.491(1.320–1.686)
CXCL1	<0.001	1.711(1.422–2.059)
CCL13	<0.001	1.412(1.172–1.702)
DEFA3	<0.001	1.207(1.117–1.305)
S100A12	<0.001	1.325(1.212–1.449)
MMP9	<0.001	1.277(1.125–1.450)
SLC40A1	<0.001	1.426(1.183–1.720)
SFTPA2	<0.001	1.759(1.394–2.218)
FABP3	<0.001	1.522(1.216–1.905)
MPO	<0.001	1.287(1.139–1.454)
RNASE3	<0.001	1.730(1.361–2.199)
CCL26	0.015	1.154(1.028–1.296)
CCR3	<0.001	1.394(1.200–1.620)
CCL2	<0.001	1.662(1.432–1.929)
CCL7	<0.001	1.660(1.445–1.906)
RNASE2	<0.001	1.711(1.338–2.188)
FCGR2B	0.020	1.339(1.047–1.713)
SEMA3B	<0.001	1.626(1.316–2.011)
CMKLR1	<0.001	1.356(1.210–1.520)
ADM	<0.001	1.818(1.550–2.132)
AREG	<0.001	1.282(1.106–1.485)
FAM3B	<0.001	0.761(0.654–0.885)
IL1RN	<0.001	1.755(1.270–2.426)
PDGFA	0.022	1.228(1.030–1.465)
SPP1	<0.001	1.401(1.240–1.582)
ANGPTL4	<0.001	1.535(1.301–1.811)
IL1R2	<0.001	1.475(1.301–1.671)
NR3C2	0.041	0.800(0.646–0.991)
PTGER3	0.020	0.852(0.745–0.975)
RORA	<0.001	0.614(0.474–0.795)
TUBB3	<0.001	1.294(1.171–1.431)
ICOS	<0.001	0.673(0.559–0.809)

Figure 3 Forest plot of the differentially-expressed IRGs related to prognosis. IRGs, immune-related genes.

build a five-IRG-based prognostic signature to predict the survival time of patients with IPF in the training cohort.

Figure 4A-4E shows the survival outcomes of IPF patients stratified by *CXCL14*, *SLC40A1*, *RNASE3*, *CCR3*, and *RORA*. The survival curve revealed that IPF patients with higher expression levels of *CXCL14*, *SLC40A1*, *RNASE3*, and *CCR3* had much worse survival outcomes. Patients with a relatively lower expression of *RORA* had markedly longer OS.

Detailed results of the multivariate Cox regression analysis, including coefficients, P values, hazard ratios, etc., are provided in Table S2. Accordingly, the patient's risk score representing the risk for OS was calculated as follows: risk score = 0.1970 × expression value of *CXCL14* + 0.3280 × expression value of *SLC40A1* + 0.5852× expression value of *RNASE3* + 0.2802 × expression value of *CCR3* – 0.6504 × expression value of *RORA*. According to the median risk score, IPF patients were divided into high- and low-risk groups. Individuals with risk scores beyond 0.711 were recognized as high-risk; otherwise, they were considered low-risk (*Figure 5A*, Table S3). There was a significant decrease in the OS of IPF patients as the risk score increased (*Figure 5B*). *Figure 5C* displays the expression level of the five IRGs

between the high- and low-risk groups. As shown in *Figure 5C*, *CXCL14*, *SLC40A1*, *RNASE3*, and *CCR3* were more highly expressed, while *RORA* expression exhibited relatively lower expression in the high-risk IPF patients than in the low-risk individuals. The survival curve constructed by the five-IRG-based prognostic signature in the training cohort showed that there was an extremely significant difference between the high- and low-risk groups (*Figure 6A*). A validation cohort was utilized to verify the five-IRG-based signature, and notable differential survival outcomes were observed between the high- and low-risk groups (*Figure 6B*). The area under curves (AUC) of the five-IRG-based prognostic signature for IPF in the training model was 0.858 (*Figure 6C*). The AUC of this predictive five-gene-based signature in the validation was 0.837 (*Figure 6D*), indicating that this predictive signature could be trusted.

Correlation expression map of the five genes included in the predictive signature

A correlation map of the five included prognostic IRGs expression levels is described in *Figure 7*. The strongest expression correlations were observed between *RNASE3*

Page 8 of 13

Qiu et al. A prognostic signature for idiopathic pulmonary fibrosis
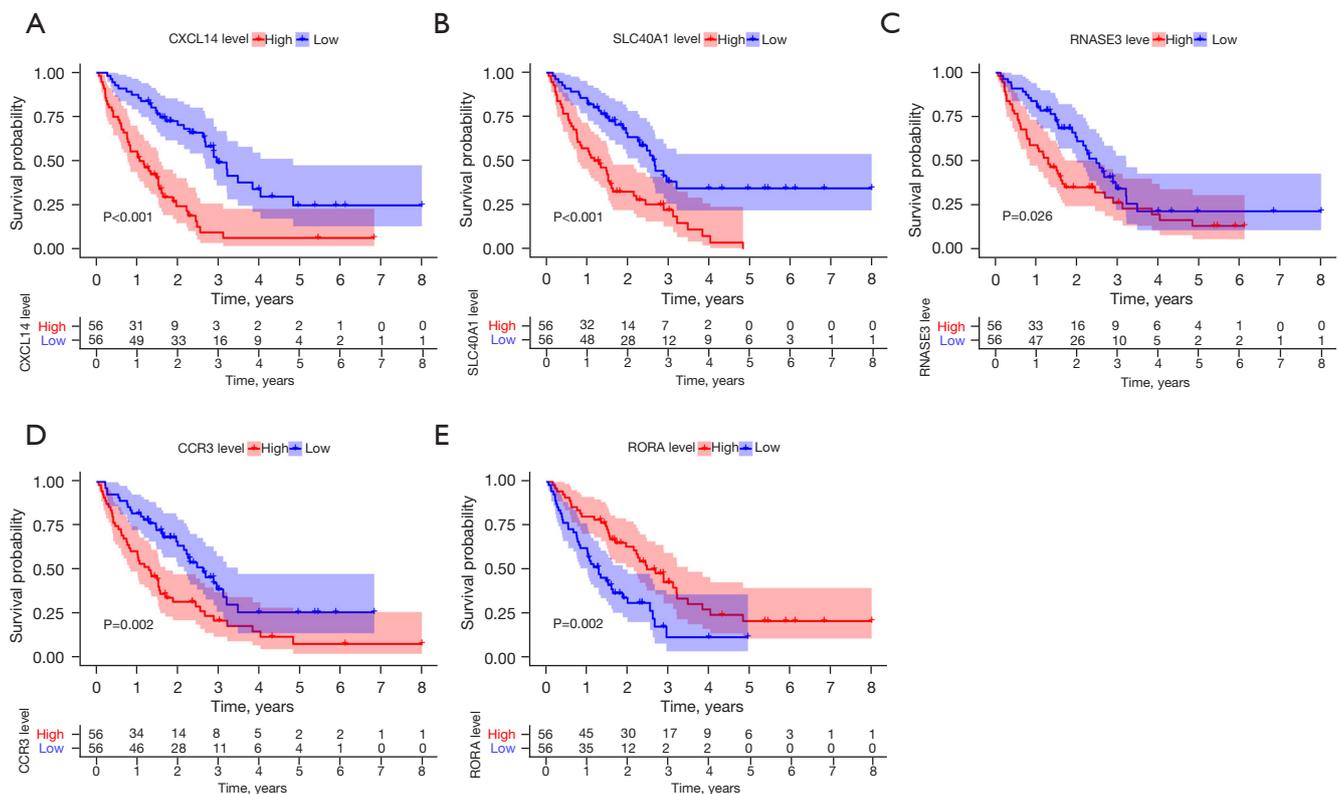
Figure 4 OS of patients with IPF stratified by the genes included in our novel signature, including (A) *CXCL14*, (B) *SLC40A1*, (C) *RNASE3*, (D) *CCR3*, and (E) *RORA*. OS, overall survival; IPF, idiopathic pulmonary fibrosis.

and *SLC40A1* (P<0.01, r=0.394), as well as between *RORA* and *CXCL14* (P<0.01, r=−0.355). Meanwhile, the expression level of *CCR3* was significantly positively correlated with the expression of *CXCL14* (P<0.01, r=0.258). There was an intimate positive association between *RNASE3* and *CCR3* (P<0.01, r=0.293).

Discussion

IPF is the most prevalent subtype of interstitial lung disease (ILD) worldwide (25). However, it has the poorest prognosis among the various ILD subtypes, with a median survival of 2–3 years after diagnosis (3,4). Lung transplantation is the only intervention that has been shown to prolong survival for patients with IPF (26). Pirfenidone and nintedanib have emerged as effective therapies that can significantly slow the decline in forced vital capacity (FVC) and disease progression in IPF patients (27,28). However, the prognosis of IPF remains unfavourable. The poor prognosis of IPF is partly due to a lack of effective prognostic biomarkers

to guide treatment. Without the ability to forecast disease progression, it is difficult to determine which IPF patients are likely to benefit from new therapies or lung transplantation. Therefore, we constructed a molecular genomic signature to predict the prognosis of IPF patients using the GSE70866 gene expression dataset from the GEO database.

Previous studies have revealed that the immune system possesses an actual effect on the IPF process (22,29,30). All stages of fibrogenesis are accompanied by innate and adaptive immune responses (22). More importantly, increasing evidence has appeared over the last few years establishing the meaningful role of IRGs in the pathogenesis and treatment of lung fibrosis (23,24,31,32). It has been shown that regulating the expression of IRGs can ameliorate pulmonary fibrogenesis in bleomycin-induced (BLM-induced) mouse models (31,32). Furthermore, data from clinical trials of newly developed drugs for the treatment of IPF have demonstrated the active role of IRG-targeting drugs in slowing disease progression. For instance, IRG-
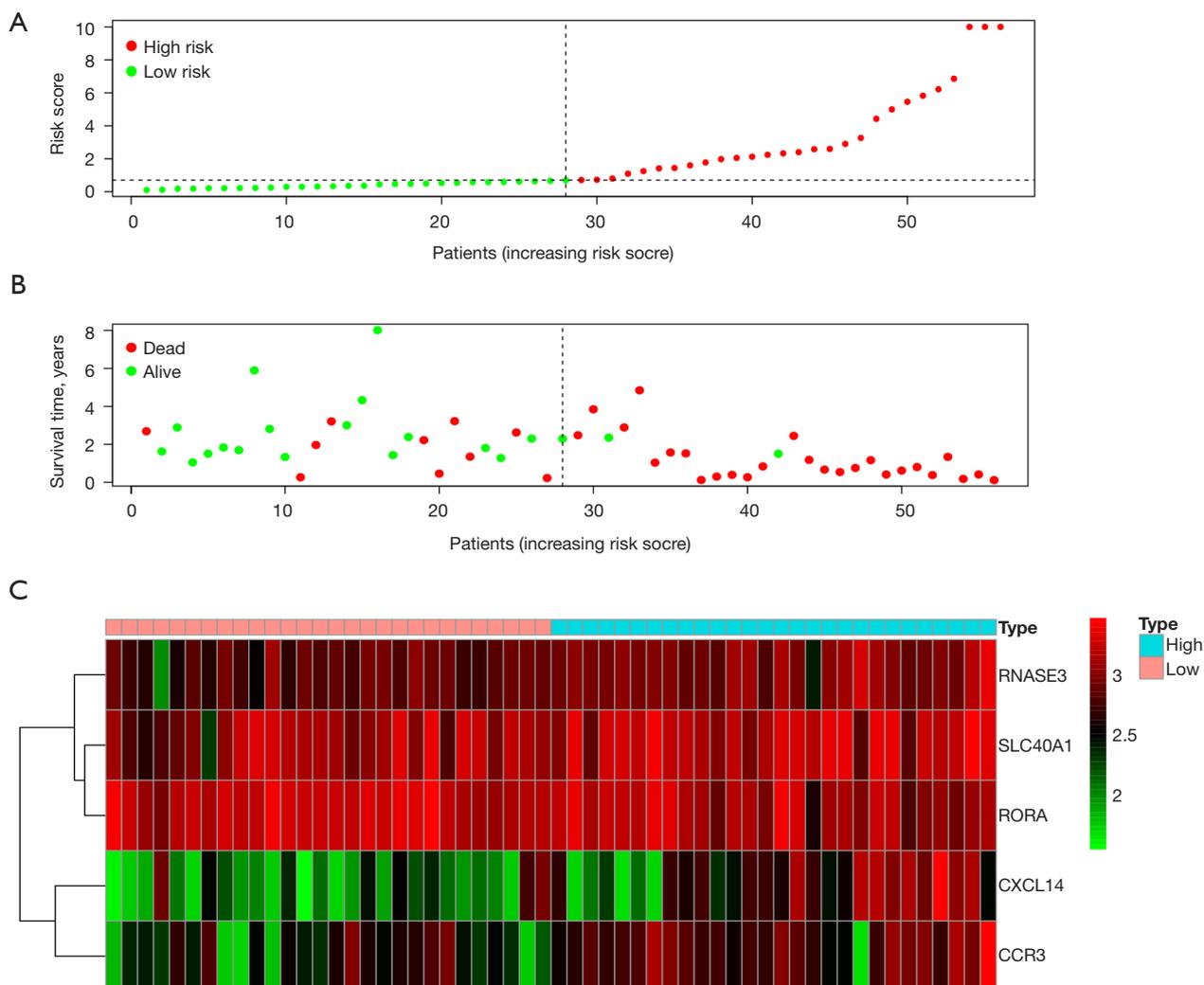
Figure 5 The risk score could effectively predict IPF patient prognosis. (A) Scatter plot of the risk score distribution of the samples. One point refers to a sample, red points are samples with higher risk scores, green points are samples with lower risk scores, and the intersecting point represents the median risk score. (B) Scatter plot of the survival outcome distribution of the samples. One point refers to a sample, red points represent live samples, green points represent dead samples with lower risk scores, and the intersecting point represents the median risk score. (C) Heatmap of signature-based genes (*CXCL14*, *SLC40A1*, *RNASE3*, *CCR3*, and *RORA*) between the high- and low-risk groups. IPF, idiopathic pulmonary fibrosis.

targeting drugs have been shown to play a positive role in reducing fibrogenesis (33). These previous studies highlight the importance of IRGs in the pathophysiological mechanism of IPF. In the present study, we were interested in the role of IRGs in the prognosis of IPF.

In total, 112 IPF patients and 20 healthy individuals were included in our study. The included IPF patients were predominantly older males (aged >65 years old). This demographic feature, as well as the fact that the prevalence

of IPF is higher in men than in women, are consistent with previous studies (1,3). In this comparative microarray profile of an IPF cohort versus a healthy individual cohort, a total of 379 DEGs were identified. The genes involved in encoding extracellular matrix (ECM) components, tissue architecture remodeling, and ECM accumulation (SPP1, MMP7, MMP10, CCL2, and ITGB3) were observed to be significantly upregulated (34-37). Of the 379 DEGs, 52 were filtered as IRGs based on the ImmPort database.

Page 10 of 13

Qiu et al. A prognostic signature for idiopathic pulmonary fibrosis
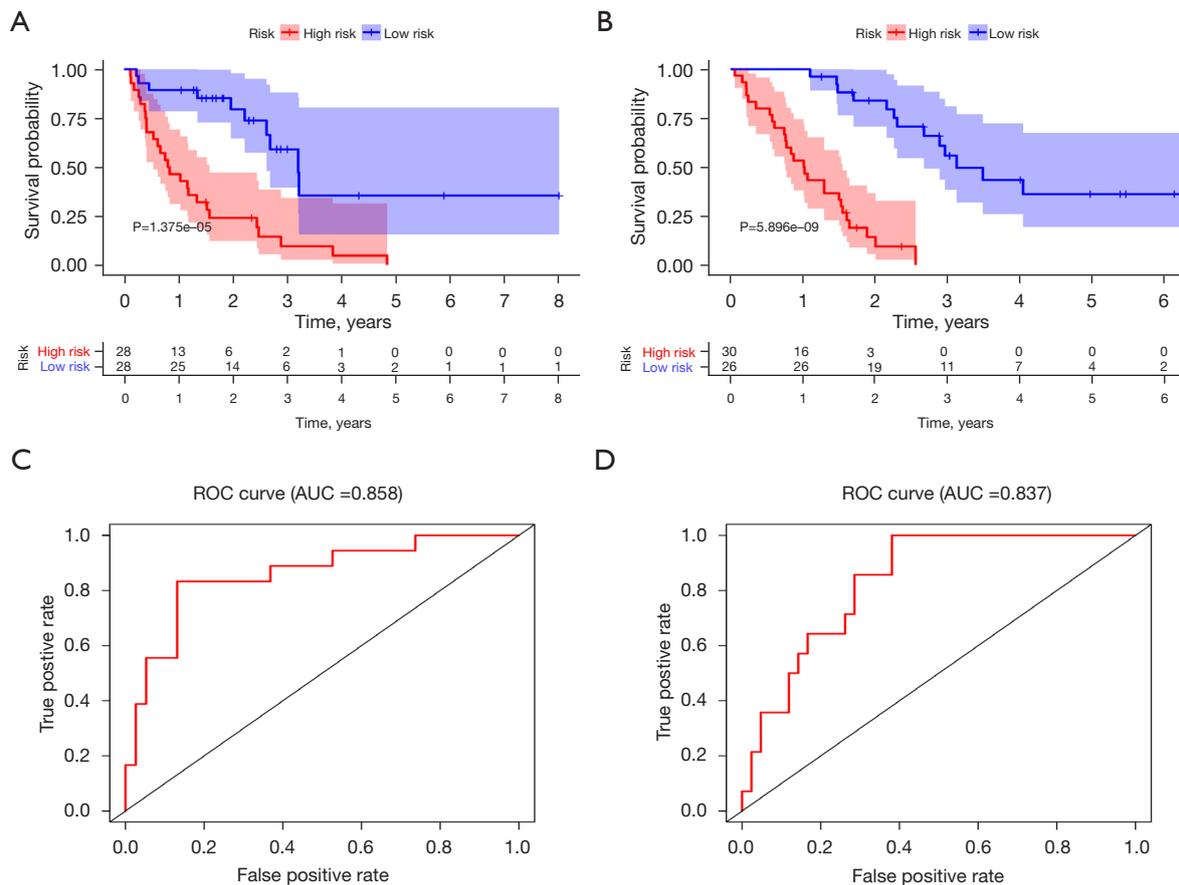
Figure 6 Signature of predicting survival probability for IPF patients. (A) Survival curve of the risk score distribution of the training cohort, which also shows the 1-, 2-, 3-, 4-, 5-, and 6-year survival rates of IPF patients. (B) Survival curve of the risk score distribution of the validation cohort, which also shows the 1-, 2-, 3-, 4-, 5-, and 6-year survival rates of IPF patients. (C) ROC curve of the signature in the training cohort. (D) ROC curve of the signature in the validation cohort. IPF, idiopathic pulmonary fibrosis; ROC, receiver operating characteristic.

Next, 37 of these 52 differentially-expressed IRGs were recognized as significant prognostic biomarkers for patients with IPF. More than 70% of the differentially-expressed IRGs had notable associations with survival. Our results further suggested that there was a close association between IRGs and the progression of IPF, which was consistent with previous studies. Based on these findings, a five IRG-based prognostic signature (*CXCL14*, *SLC40A1*, *RNASE3*, *CCR3*, and *RORA*), was built in the training cohort in this study. This signature presented an excellent predictive prognostic effect, with an AUC value of 0.858. In addition, the risk score was significantly different between the high- and low-risk groups. Meanwhile, the risk score was significantly correlated with the OS of IPF patients. *CXCL14*, *SLC40A1*,

CXCL14, and *CCR3* were differentially-upregulated genes between IPF patients and healthy individuals. The expression levels of these four genes in the high-risk IPF group were significantly higher than those in the low-risk group. *RORA* was detected at a lower expression level in the healthy individuals group compared to the IPF group. Consistently, the expression level of *RORA* was lower in the high-risk IPF group than in the low-risk group.

Fibroblast foci represent the main pathogenic lesions of IPF, including abnormally activated fibroblasts and myofibroblasts. Myofibroblasts are the main effector cells of IPF. They can secrete a large amount of ECM protein and promote the abnormal hardening of ECM, which leads to the remodeling of lung structure and the
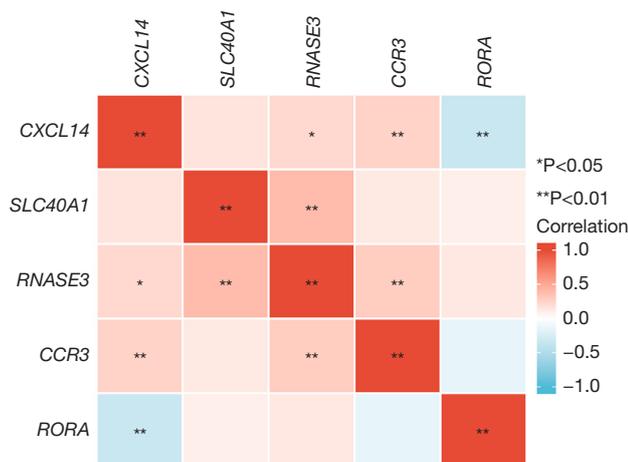
Figure 7 Gene co-expression network of 5 genes: *CXCL14*, *SLC40A1*, *RNASE3*, *CCR3*, and *RORA*.

gradual loss of lung function (38-40). Previous studies have confirmed that knockdown of *CXCL14* could inhibit lung fibrogenesis by suppressing lung fibroblasts proliferation and downregulating *MMP2/9* (31). Zagai *et al.* found eosinophil cationic protein (*ECP*, also known as *RNASE3*) could stimulate human lung fibroblasts to secrete extracellular matrix, thereby leads to airway fibrosis (41). The concentration of *RNASE3* in bronchoalveolar lavage fluid (BALF) is markedly increased in IPF patients compared with healthy individuals and is highly correlated with acute exacerbation during the preceding 3- to 6-month period (42,43). *CCR3* can increase the activation, migration and proliferation ability of lung fibroblasts, and the ability of myofibroblasts to secrete ECM protein (44,45). In addition, *CCR3* is notably expressed in the lungs of BLM-induced mice and is expressed not only by eosinophils but also by neutrophils (44). *CCR3* plays a key role in the recruitment of granulocytes and is an important suppressor of fibrogenesis in BLM-treated lungs (44). These studies on the pathophysiological mechanisms between IPF and *CXCL14*, *RNASE3*, and *CCR3* increase the credibility of the signature constructed in our study. Our research also showed that there is a meaningful correlation between the expression of *RNASE3* and *CCR3*. Meanwhile, a significant expression correlation between *CXCL14* and *CCR3* was also observed in this study. For the *SLC40A1* and *RORA*, no relevant studies have been conducted to determine the association with lung fibrosis. We first reported that there may be some potential associations between the pathological mechanism of IPF and *SLC40A1* along with *RORA*. The

specific pathophysiological mechanism is worthy of further study.

Finally, we evaluated the performance of the genomic signature in the validation cohort. The signature showed an equally excellent ability to distinguish between high- and low-risk patient groups. The AUC value of the Receiver Operating Characteristic Curve (ROC) curve was 0.837, demonstrating the potential applicability of our findings for real-world use.

While the genomic model developed in this study was successfully validated, there were still some potential limitations that should be noted. Firstly, this research was based on the gene expression profiles from the GEO database. Due to the difficult of recruitment of a large number of IPF patients, no validation of the 5 genes in real world data in this paper. Also, the IPF patients included in this study were all from Germany. Thus, our results might only represent patients in Germany and might not applicable to all IPF patients worldwide. Finally, due to limited data on treatment, our study did not subgroup IPF patients according to the different treatment choices. Consequently, the reliability and accuracy of our results might be affected and needs to be re-evaluated by future studies.

Conclusions

In conclusion, our study identified a novel five-IRG-based signature that is a reproducible predictor of outcome in IPF patients. This novel signature benefits the personalized management of patients with IPF. Furthermore, this finding provides new insights into the relationship between the immune system and IPF, offering incremental clinical value for IPF prognosis and therapy.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://dx.doi.

Page 12 of 13

Qiu et al. A prognostic signature for idiopathic pulmonary fibrosis

org/10.21037/atm-21-4545

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/atm-21-4545). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The data used in this study was derived from a public database, and thus, no ethical approval was needed.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med 2018;198:e44-68.
2. Selman M, King TE, Pardo A, et al. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. Ann Intern Med 2001;134:136-51.
3. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183:788-824.
4. Mura M, Porretta MA, Bargagli E, et al. Predicting survival in newly diagnosed idiopathic pulmonary fibrosis: a 3-year prospective study. Eur Respir J 2012;40:101-9.
5. Vancheri C, du Bois RM. A progression-free end-point for idiopathic pulmonary fibrosis trials: lessons from cancer. Eur Respir J 2013;41:262-9.
6. Kim HJ, Perlman D, Tomic R. Natural history of idiopathic pulmonary fibrosis. Respir Med 2015;109:661-70.
7. Kim DS, Collard HR, King TE Jr. Classification and natural history of the idiopathic interstitial pneumonias. Proc Am Thorac Soc 2006;3:285-92.
8. Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2011;183:431-40.
9. Selman M, López-Otín C, Pardo A. Age-driven developmental drift in the pathogenesis of idiopathic pulmonary fibrosis. Eur Respir J 2016;48:538-52.
10. Bueno M, Lai YC, Romero Y, et al. PINK1 deficiency impairs mitochondrial homeostasis and promotes lung fibrosis. J Clin Invest 2015;125:521-38.
11. Prasse A, Binder H, Schupp JC, et al. BAL Cell Gene Expression Is Indicative of Outcome and Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med 2019;199:622-30.
12. Caporarello N, Meridew JA, Jones DL, et al. PGC1α repression in IPF fibroblasts drives a pathologic metabolic, secretory and fibrogenic state. Thorax 2019;74:749-60.
13. McDonough JE, Kaminski N, Thienpont B, et al. Gene correlation network analysis to identify regulatory factors in idiopathic pulmonary fibrosis. Thorax 2019;74:132-40.
14. DePianto DJ, Chandriani S, Abbas AR, et al. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. Thorax 2015;70:48-56.
15. Boon K, Bailey NW, Yang J, et al. Molecular phenotypes distinguish patients with relatively stable from progressive idiopathic pulmonary fibrosis (IPF). PLoS One 2009;4:e5134.
16. Korfei M, von der Beck D, Henneke I, et al. Comparative proteome analysis of lung tissue from patients with idiopathic pulmonary fibrosis (IPF), non-specific interstitial pneumonia (NSIP) and organ donors. J Proteomics 2013;85:109-28.
17. Xu Z, Mo L, Feng X, et al. Using bioinformatics approach identifies key genes and pathways in idiopathic pulmonary fibrosis. Medicine (Baltimore) 2020;99:e22099.
18. Min F, Gao F, Liu Z. Screening and further analyzing differentially expressed genes in acute idiopathic pulmonary fibrosis with DNA microarray. Eur Rev Med Pharmacol Sci 2013;17:2784-90.
19. Li FJ, Surolia R, Li H, et al. Autoimmunity to Vimentin Is Associated with Outcomes of Patients with Idiopathic Pulmonary Fibrosis. J Immunol 2017;199:1596-605.
20. Huang Y, Ma SF, Vij R, et al. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. BMC Pulm Med 2015;15:147.
21. Meyer KC, Raghu G, Baughman RP, et al. An official

American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. Am J Respir Crit Care Med 2012;185:1004-14.

22. Heukels P, Moor CC, von der Thüsen JH, et al. Inflammation and immunity in IPF pathogenesis and treatment. Respir Med 2019;147:79-91.

23. Cecchini MJ, Hosein K, Howlett CJ, et al. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. Respir Res 2018;19:153.

24. Walsh SM, Worrell JC, Fabre A, et al. Novel differences in gene expression and functional capabilities of myofibroblast populations in idiopathic pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol 2018;315:L697-710.

25. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. N Engl J Med 2018;378:1811-23.

26. George PM, Patterson CM, Reed AK, et al. Lung transplantation for idiopathic pulmonary fibrosis. Lancet Respir Med 2019;7:271-82.

27. Richeldi L, Costabel U, Selman M, et al. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. N Engl J Med 2011;365:1079-87.

28. Taniguchi H, Ebina M, Kondoh Y, et al. Pirfenidone in idiopathic pulmonary fibrosis. Eur Respir J 2010;35:821-9.

29. Drakopanagiotakis F, Wujak L, Wygrecka M, et al. Biomarkers in idiopathic pulmonary fibrosis. Matrix Biol 2018;68-69:404-21.

30. O'Dwyer DN, Ashley SL, Moore BB. Influences of innate immunity, autophagy, and fibroblast activation in the pathogenesis of lung fibrosis. Am J Physiol Lung Cell Mol Physiol 2016;311:L590-601.

31. Li L, Li Q, Wei L, et al. Chemokine (C-X-C motif) ligand 14 contributes to lipopolysaccharide-induced fibrogenesis in mouse L929 fibroblasts via modulating PPM1A. J Cell Biochem 2019;120:13372-81.

32. Mehrad B, Burdick MD, Strieter RM. Fibrocyte CXCR4 regulation as a therapeutic target in pulmonary fibrosis. Int J Biochem Cell Biol 2009;41:1708-18.

33. Lukey PT, Harrison SA, Yang S, et al. A randomised, placebo-controlled study of omipalisib (PI3K/mTOR) in idiopathic pulmonary fibrosis. Eur Respir J 2019;53:1801992.

34. Insua-Rodríguez J, Pein M, Hongu T, et al. Stress

signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis. EMBO Mol Med 2018;10:e9003.

35. Rohani MG, Parks WC. Matrix remodeling by MMPs during wound repair. Matrix Biol 2015;44-46:113-21.

36. Kuo CS, Pavlidis S, Zhu J, et al. Contribution of airway eosinophils in airway wall remodeling in asthma: Role of MMP-10 and MET. Allergy 2019;74:1102-12.

37. Liu A, Liu Y, Li B, et al. Role of miR-223-3p in pulmonary arterial hypertension via targeting ITGB3 in the ECM pathway. Cell Prolif 2019;52:e12550.

38. Guillotin D, Taylor AR, Platé M, et al. Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis. Thorax 2021;76:73-82.

39. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. Lancet 2017;389:1941-52.

40. Martinez FJ, Collard HR, Pardo A, et al. Idiopathic pulmonary fibrosis. Nat Rev Dis Primers 2017;3:17074.

41. Zagai U, Dadfar E, Lundahl J, et al. Eosinophil cationic protein stimulates TGF-beta1 release by human lung fibroblasts in vitro. Inflammation 2007;30:153-60.

42. Fujimoto K, Kubo K, Yamaguchi S, et al. Eosinophil activation in patients with pulmonary fibrosis. Chest 1995;108:48-54.

43. Birring SS, Parker D, McKenna S, et al. Sputum eosinophilia in idiopathic pulmonary fibrosis. Inflamm Res 2005;54:51-6.

44. Huaux F, Gharaee-Kermani M, Liu T, et al. Role of Eotaxin-1 (CCL11) and CC chemokine receptor 3 (CCR3) in bleomycin-induced lung injury and fibrosis. Am J Pathol 2005;167:1485-96.

45. Puxeddu I, Bader R, Piliponsky AM, et al. The CC chemokine eotaxin/CCL11 has a selective profibrogenic effect on human lung fibroblasts. J Allergy Clin Immunol 2006;117:103-10.

(English Language Editor: B. Draper)

Table S1 The information of logFC for differential expression genes between IPF patients and the control group

id	logFC
SPP1	3.856003
PPBP	3.767079
MMP7	3.094781
SFTPB	3.001383
ITGB3	2.777328
CYP1B1	2.489336
TUBB3	2.463291
LRRC2	2.409921
CYTL1	2.379117
VSNL1	2.262473
HTRA1	2.239659
OLIG1	2.238058
TIMP3	2.238012
FFAR3	2.235872
CCL2	2.19958
MMP10	2.157871
MERTK	2.138287
C14orf34	2.129425
S100A12	2.119374
BICC1	2.10942
GPR179	2.106344
PLA2G7	2.106114
FABP3	2.102401
DEFA3	2.095526
SPINK1	2.068789
TPSAB1	2.048006
FNDC5	2.003073
IL1R2	2.000731
AQP4	1.980216
SOD3	1.972751
STAB1	1.966629
SFTPC	1.939547
TPSD1	1.936252
CCL7	1.932093

Table S1 (*continued*)

Table S1 (*continued*)

id	logFC
VSTM1	1.917281
TPST1	1.913652
PROM2	1.905313
CST6	1.888228
SDS	1.879609
WNT2B	1.874928
GPR182	1.869439
HBD	1.86808
HS3ST2	1.837265
ANGPTL4	1.76816
LOC729040	1.764131
TM4SF1	1.762285
GFRA2	1.757664
MATK	1.744541
EMP1	1.734211
AANAT	1.710304
RNASE1	1.678622
DACH1	1.675864
COL22A1	1.675523
NRAP	1.67189
PID1	1.667208
CCR3	1.661399
KIAA0125	1.653681
F13A1	1.652466
CPA3	1.649287
CD86	1.637381
RPA4	1.625746
ARAP3	1.624116
SLC28A3	1.605387
NT5DC2	1.602076
RNASE2	1.596596
EHD2	1.594722
B3GNT8	1.593013
SPRY2	1.585458

Table S1 (*continued*)

id	logFC
MGC24103	1.580791
CXCL14	1.5796
CH25H	1.575003
RGL1	1.550247
MRVI1	1.542015
RAB3IL1	1.53331
SEPP1	1.528336
OR13H1	1.516111
KRT79	1.498058
MALL	1.495627
IBSP	1.492416
ADM	1.489362
PI3	1.479423
STEAP4	1.46926
CLC	1.467909
CCL13	1.458057
CDA	1.445885
CCL26	1.442975
ARNT2	1.43881
DIRAS1	1.426543
HDC	1.416884
CLGN	1.408777
HS3ST1	1.408676
PRSS8	1.406913
HIST2H3A	1.403166
GPT	1.382998
C10orf116	1.376584
IL8	1.373865
CNIH3	1.373499
CMKLR1	1.37083
ACOX2	1.367016
SH3RF1	1.365489
RNASE3	1.362714
MRC2	1.355777

Table S1 (*continued*)

id	logFC
LGMN	1.352665
CD36	1.350796
MPO	1.337994
CYR61	1.330317
ASPHD1	1.325651
KRT14	1.291002
TM4SF19	1.288525
RGS2	1.27315
CACNA1G	1.272916
OR8G5	1.27172
FCN1	1.267912
IER3	1.264788
KIT	1.254733
TDRD10	1.254632
PRKAR1B	1.240034
VCAN	1.234562
MMP9	1.227406
PCSK9	1.212719
MS4A2	1.211344
AREG	1.20871
SFTPD	1.206889
FAM20A	1.20638
ECM1	1.206065
CEACAM7	1.202738
SNAI1	1.199136
HRK	1.196395
KCNG1	1.194507
CLDN18	1.193072
CXCL5	1.192979
SLC40A1	1.19176
DIRC3	1.190271
ATP9A	1.182874
FBXO15	1.181945
P2RY2	1.180127

Table S1 (*continued*)

id	logFC
FGFR1	1.177801
PGA3	1.176969
LOC100132368	1.170005
SFN	1.16869
MUC21	1.166485
HOMER3	1.164147
S1PR3	1.162823
HAMP	1.160114
SPTLC3	1.159878
ABLIM3	1.156865
ENHO	1.155081
AQP2	1.154688
SLC16A10	1.152375
SEC14L2	1.148491
SLC24A3	1.145543
LTC4S	1.145343
TAAR2	1.143163
LRG1	1.139364
C6orf108	1.139162
HIST1H3B	1.137506
GAS6	1.134418
SULT1C2	1.131406
DYSF	1.126893
C1orf111	1.126254
LOC283050	1.123883
HES4	1.119226
KRT17	1.1121
CALB2	1.110349
MUC1	1.110008
NRGN	1.106171
EPO	1.103573
PAX6	1.100351
FAM198B	1.098049
NIPAL4	1.092226

id	logFC
GAL3ST4	1.086711
NOV	1.086149
CYBRD1	1.086021
SNCA	1.085304
SPTB	1.08173
FCGR2B	1.080386
CLEC5A	1.075267
CXCL1	1.074218
QPCT	1.072253
C14orf162	1.070768
OR52E8	1.066605
FAM124B	1.06621
UCK2	1.064365
MGC14436	1.063247
SLC16A8	1.05739
FCER2	1.056397
PPP1R14C	1.053759
IL1RN	1.051554
CLEC11A	1.046712
PMP22	1.041398
SFRP1	1.03858
SFTA2	1.034844
MYL9	1.034835
NPAS2	1.030934
CD24	1.030668
LEPREL1	1.030111
LOC284263	1.02944
SFTPA2	1.029373
MGP	1.024552
CEBPE	1.023772
MYO7A	1.022615
FAM20C	1.020749
KRTAP4-11	1.020715
LOC100130480	1.017798

id	logFC
PDGFA	1.012284
SEMA3B	1.00929
KIF4A	1.005653
SLC47A1	-1.00141
ERN2	-1.00479
MPP7	-1.00695
HOXC4	-1.00729
GATA3	-1.00734
MYB	-1.00767
RANBP3L	-1.00783
RNF183	-1.01242
C11orf80	-1.01247
CD6	-1.01563
JAG2	-1.01796
AQP7P3	-1.02239
LOC283392	-1.02313
LOC100270804	-1.02639
RORA	-1.02654
SNTN	-1.02656
HRASLS	-1.0302
ITK	-1.03125
SNAI2	-1.03637
SLC7A2	-1.03645
C8A	-1.03662
CSPG4	-1.03949
LOC650293	-1.04049
TFRC	-1.04068
RIC3	-1.04191
ZNF404	-1.04298
FOLR3	-1.04313
NR3C2	-1.04324
CC2D2A	-1.04924
THAP2	-1.05016
ZNF239	-1.05069

id	logFC
ZNF589	-1.05218
SNORA12	-1.05973
PM20D1	-1.06203
TCF7	-1.06253
EPB41L4A	-1.06331
TNNT1	-1.06454
ZNF610	-1.06512
SCN8A	-1.06542
ARMC3	-1.06592
LOC256880	-1.06671
D4S234E	-1.06734
LARP6	-1.06889
IFT81	-1.06903
SERPINI2	-1.07138
LOC400655	-1.07223
GPR85	-1.07447
DLEC1	-1.07639
ITGB8	-1.0784
MYO7B	-1.08026
CDC42EP3	-1.08253
LOC728218	-1.08572
ABHD1	-1.08959
MAL	-1.09019
MAGI3	-1.09278
COL9A2	-1.09321
KPNA5	-1.10405
GRIN3B	-1.10515
DSP	-1.11168
KLK11	-1.11497
LOC729867	-1.11537
C7orf58	-1.11548
TMEM130	-1.11625
EPM2AIP1	-1.11628
NDN	-1.12473

Table S1 (*continued*)

id	logFC
ODZ4	-1.1275
TPBG	-1.1293
CAPS2	-1.12947
OR2A7	-1.13147
RAP1GAP2	-1.13156
FLJ46875	-1.13275
ENPP5	-1.13735
FAM3B	-1.13918
ICOS	-1.13931
C20orf46	-1.14512
RAB39B	-1.15124
DNAH5	-1.15318
FBXL16	-1.15713
SLC4A8	-1.1666
CAPN11	-1.16854
ANK3	-1.16942
SERPINB4	-1.17668
GPRASP1	-1.18083
LOC100131289	-1.19614
NHS	-1.19903
MAP9	-1.20022
CES1	-1.20284
ZBP1	-1.20287
ACSS3	-1.20422
HLF	-1.2142
ZNF251	-1.21519
AKR1E2	-1.21823
FAM70A	-1.21997
CHRM2	-1.22546
PDCD1LG2	-1.23139
NBEA	-1.2346
TRIB2	-1.23661
LOC400891	-1.23772
SEC16B	-1.24531

id	logFC
IQCA1	-1.24651
ZMAT3	-1.252
ZFP14	-1.25298
MFAP3L	-1.25472
SYNE2	-1.2653
KLF12	-1.26866
KIAA0408	-1.27245
FAM47E	-1.27394
LPAR3	-1.27435
MYO1A	-1.27997
C17orf69	-1.28199
EPB41L5	-1.28237
TJP1	-1.29275
ODF3L1	-1.29365
RFPL4A	-1.30527
GDA	-1.31169
C9orf30-TMEFF1	-1.31377
C9orf171	-1.32157
SLITRK4	-1.3288
TC2N	-1.33076
ACSM1	-1.33541
PLEKHA6	-1.33918
CPLX3	-1.35549
KLRB1	-1.35674
BEX5	-1.35708
ZNF540	-1.37252
IFNG	-1.37571
TRAT1	-1.38275
XCL1	-1.40778
DLX3	-1.41913
TNFRSF25	-1.42826
PIGR	-1.43529
LAMB1	-1.43589
SAMD12	-1.45227

Table S1 (*continued*)

id	logFC
CXCL9	-1.46689
SHROOM3	-1.46744
RBM11	-1.47051
ARHGAP24	-1.47717
GSTA5	-1.48294
C1orf194	-1.48387
DMD	-1.52499
RSPH1	-1.52543
IGF1	-1.54242
TMEM200A	-1.54474
PRRT4	-1.55266
DLX4	-1.55278
LOC645206	-1.55438
FAM183A	-1.56155
LOC100128252	-1.57934
TMEM56	-1.58744
EFCAB1	-1.60353
MURC	-1.62903
LOC400043	-1.63097
DNAI2	-1.63951
CXCR7	-1.65679
ZNF702P	-1.66273
KCNAB1	-1.70793
GBP7	-1.72011
GABRE	-1.72196
CYP3A7	-1.73123
CAMP	-1.83452
LEF1	-1.85639
CD40LG	-1.87393
AOC3	-1.88126
TCEA3	-1.91778
PTGER3	-2.00014
FAM125B	-2.01546
TCF7L1	-2.03268

id	logFC
ENPP3	-2.05402
CYP3A5	-2.16341
ITIH5	-2.26283
C8B	-2.31597
NALCN	-2.56164

Table S2 Detailed results of prognostic model using the multivariate Cox regression

id	coef	HR	HR.95L	HR.95H	pvalue
CXCL14	0.197048	1.217802	1.003692	1.477588	0.045791
SLC40A1	0.328027	1.388227	0.9675	1.991911	0.074976
RNASE3	0.585181	1.795316	1.141325	2.824052	0.011344
CCR3	0.280172	1.323357	1.006682	1.73965	0.044676
RORA	-0.65037	0.521853	0.322487	0.844468	0.008089

Table S3 The grouping information of IPF patients stratified by risk scores

id	sex	futime	fustat	CXCL14	SLC40A1	RNASE3	CCR3	RORA	riskScore	risk
GSM1820750	1	2.690411	1	2.176255	7.529534	6.677045	2.666712	9.647956	0.107123	low
GSM1820791	1	1.627397	0	2.499763	6.017312	5.653176	4.215465	8.547031	0.120597	low
GSM1820810	1	2.887671	0	2.777443	5.255918	5.328032	4.453929	7.37939	0.18742	low
GSM1820848	1	1.049315	0	6.6777	5.989944	3.099479	4.009592	6.701181	0.191552	low
GSM1820802	1	1.509589	0	3.305012	6.31339	5.10529	5.519865	8.129156	0.213762	low
GSM1820787	1	1.835616	0	2.367557	6.906236	6.111136	4.424728	8.566003	0.215352	low
GSM1820837	1	1.69589	0	4.760112	3.958506	5.382541	6.043815	7.780152	0.224785	low
GSM1820752	1	5.89589	0	3.719579	7.266238	6.78585	2.44916	8.809874	0.230328	low
GSM1820832	1	2.813699	0	2.967795	8.591801	5.829121	2.416916	8.248287	0.250268	low
GSM1820842	1	1.334247	0	3.176554	9.216768	4.82854	4.67816	8.422183	0.299943	low
GSM1820755	1	0.265753	1	2.500494	8.716575	7.674821	2.596618	9.628112	0.300185	low
GSM1820744	1	1.967123	1	4.268577	8.020283	5.474395	4.999664	8.791301	0.315529	low
GSM1820819	1	3.208219	1	2.021657	7.910024	6.498425	4.388959	8.62793	0.333527	low
GSM1820828	0	3.005479	0	3.344809	7.928118	6.094522	4.151996	8.469966	0.356513	low
GSM1820775	1	4.328767	0	2.391152	7.643999	6.374089	5.155236	8.687885	0.364383	low
GSM1820739	1	8.016438	0	2.979139	6.655104	5.899136	6.985888	8.39758	0.451896	low
GSM1820796	1	1.432877	0	4.52326	7.191799	6.691145	5.24078	9.028555	0.472453	low
GSM1820808	0	2.389041	0	2.869742	7.561899	6.878794	4.727886	8.639491	0.479462	low
GSM1820753	1	2.221918	1	4.899363	9.158801	6.081315	5.120282	9.461191	0.49536	low
GSM1820745	1	0.457534	1	3.816168	7.090138	7.198446	5.399707	9.108913	0.530801	low
GSM1820764	1	3.221918	1	4.261457	9.340405	6.533717	6.070303	10.04573	0.539097	low
GSM1820814	1	1.350685	1	3.363133	6.093123	6.612222	6.740906	8.376596	0.582347	low
GSM1820834	1	1.808219	0	2.980574	8.905558	5.695522	4.508208	7.88783	0.584142	low
GSM1820804	1	1.282192	0	3.386925	8.595365	5.624132	4.024725	7.564871	0.590677	low
GSM1820815	1	2.624658	1	3.149784	6.814047	6.241344	4.733438	7.386996	0.617481	low
GSM1820823	0	2.30411	0	2.451211	8.230981	6.625342	4.1469	7.932164	0.63817	low

Table S3 (*continued*)

Table S3 (*continued*)

id	sex	futime	fustat	CXCL14	SLC40A1	RNASE3	CCR3	RORA	riskScore	risk
GSM1820756	1	0.224658	1	5.772672	7.865604	6.66736	2.443293	8.000723	0.662443	low
GSM1820797	1	2.29863	0	6.766446	7.397493	6.317389	3.591284	8.178159	0.692039	low
GSM1820773	1	2.479452	1	5.537316	6.992163	6.55567	4.955777	8.361049	0.711424	high
GSM1820743	1	3.846575	1	2.379045	9.403569	7.207628	5.371375	9.353161	0.726805	high
GSM1820838	0	2.353425	0	3.310096	6.397982	6.739718	6.18868	7.900863	0.80104	high
GSM1820777	1	2.890411	1	4.054571	8.720728	6.731595	5.445271	8.492374	1.093124	high
GSM1820768	1	4.846575	1	2.335034	8.767706	7.256159	5.791397	8.407448	1.252062	high
GSM1820812	1	1.035616	1	3.374072	8.411273	6.745853	6.078712	8.018581	1.415376	high
GSM1820747	1	1.572603	1	2.388742	9.728437	7.015746	7.851603	9.364179	1.440252	high
GSM1820741	1	1.526027	1	5.634221	8.213857	6.932814	7.047001	9.000967	1.599634	high
GSM1820778	0	0.123288	1	5.134296	8.305977	6.392059	6.207405	7.886328	1.776669	high
GSM1820820	1	0.30137	1	5.788532	7.812034	6.59018	5.74561	7.650992	1.976326	high
GSM1820835	1	0.394521	1	4.093968	6.971222	6.175657	5.836972	6.320748	2.05381	high
GSM1820788	0	0.268493	1	4.526021	8.377087	7.615885	4.987702	8.040077	2.122609	high
GSM1820792	1	0.838356	1	5.897792	6.782039	7.60049	5.909519	7.95029	2.242037	high
GSM1820841	1	1.50411	0	5.629299	7.880296	5.949937	5.529668	6.715361	2.32938	high
GSM1820774	1	2.446575	1	5.136586	9.032086	7.895386	7.036578	9.499056	2.402559	high
GSM1820772	1	1.183562	1	7.822168	8.932191	6.751911	6.252478	8.786222	2.579864	high
GSM1820850	1	0.668493	1	5.646344	8.099367	4.359378	5.454	5.20076	2.595834	high
GSM1820806	1	0.539726	1	4.565593	9.405672	7.212026	4.734887	7.618223	2.900974	high
GSM1820830	1	0.756164	1	4.953695	9.462666	7.525471	4.695753	7.848243	3.264402	high
GSM1820846	1	1.167123	1	8.11578	6.291916	8.865949	2.279307	6.90457	4.424709	high
GSM1820795	0	0.410959	1	8.151472	9.225448	7.652115	5.740017	8.606792	4.996426	high
GSM1820786	1	0.619178	1	7.022599	9.257297	6.975988	7.36407	8.236266	5.457902	high
GSM1820824	1	0.805479	1	7.757735	6.122578	6.455595	6.865361	6.094522	5.825976	high
GSM1820822	1	0.380822	1	6.64152	8.456156	6.611661	7.035951	7.047561	6.216139	high
GSM1820799	1	1.339726	1	10.04684	8.190663	6.953897	5.593384	7.481033	6.85691	high
GSM1820798	1	0.180822	1	7.108656	8.28654	6.55567	7.94744	6.529698	11.27931	high
GSM1820742	1	0.413699	1	7.905854	9.771968	8.206341	6.754999	7.36587	23.46121	high
GSM1820829	1	0.115068	1	4.772996	9.145321	9.25043	9.855494	7.935954	31.22926	high