Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Comprehensive Study on Molecular Characteristics and Clinical Prognosis of Immune-Related Genes in Idiopathic Pulmonary Fibrosis

Jie Xuan
Zhejiang Cancer Hospital
Wenyuan Niu
Zhejiang Chinese Medical University
Gaofeng Hu
Zhejiang Cancer Hospital
Yupeng Zhang
Zhejiang Chinese Medical University
Zhen Zhang
Zhejiang Cancer Hospital
Junjie Xia
Zhejiang Chinese Medical University
Zhihao Lin
Zhejiang Cancer Hospital
Changqing Xu
Affiliated Hospital of Hangzhou Normal University
Qinglin Li
Zhejiang Cancer Hospital
Yuanqiang Li
yuanqiangli@hotmail.com

Zhejiang Cancer Hospital

Research Article

Keywords: IPF, Immune-related genes, Prognostic risk model, PPBP

Posted Date: October 30th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7616111/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background:

Idiopathic pulmonary fibrosis (IPF) is a progressive, fatal interstitial lung disease with unclear pathogenesis. Immune-related genes (IRGs) are increasingly recognized as key players in its development, but their prognostic value remains underexplored.

Methods:

Using the GSE70866 dataset, we identified differentially expressed IRGs (DE-IRGs) by intersecting IRGs (from ImmPort) with differentially expressed genes. Prognostic IRGs were refined via univariate Cox, least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression, yielding four independent factors (PPBP, CCL7, ADM, SFTPD) to construct a risk model. The model was validated in training and validation cohorts, with immune infiltration analyzed via Single-sample gene set enrichment analysis (ssGSEA). In vitro and in vivo experiments verified gene expression, particularly PPBP's role in fibrosis.

Conclusions:

The IRG-based risk model effectively stratified IPF prognosis. PPBP was abnormally elevated in IPF, and its knockdown inhibited fibroblast activation and fibrosis progression. These findings highlight PPBP as a critical pathogenic factor, offering novel prognostic and therapeutic insights for IPF.

1. Introduction

IPF is a progressive, life-threatening interstitial lung disease of unknown etiology, characterized by persistent cough, dyspnea, and progressive deterioration of pulmonary function[1]. The disease carries a grave prognosis, with median survival ranging from 2–3 years post-diagnosis and unpredictable progression patterns[1, 2]. Currently, pirfenidone and nintedanib can be used to slow the progression of IPF, but both drugs have tolerability issues and cannot improve or stabilize lung function, nor can they significantly enhance the patients' quality of life[3]. Therefore, identifying early biomarkers to predict the prognosis of IPF holds significant clinical importance.

Growing evidence underscores the crucial role of immune responses in IPF. Immune cells play a crucial role in the pathogenesis of IPF, and many IPF patients exhibit unexplained elevations in autoantibodies. Certain autoantibodies demonstrate significant associations with acute exacerbation events in IPF patients[3, 4]. Some immune cells, particularly macrophages, drive fibrosis by releasing cytokines and growth factors, interacting with fibroblasts, and responding to ECM mechanical properties, thus promoting fibroblast activation and proliferation[5]. Morse et al. utilized single-cell sequencing to

demonstrate that SPP1 high macrophages undergo significant proliferation during the progression of IPF[6]. Additionally, a separate study indicated that alveolar macrophages derived from monocytes contribute to the advancement of pulmonary fibrosis[7].Recent investigations have identified substantial immune cell accumulation in the bronchoalveolar lavage fluid of IPF patients, predominantly comprising activated natural killer (NK) cells, monocytes, and mast cells[8]. In a bleomycin (BLM)-induced mouse model, the number of Th2 cells and PD-1^{+} CD4^{+} cells increased in mice with IPF, and the augmentation of these cells is closely associated with the fibrosis process[9].

IRGs are a class of genes that directly or indirectly participate in the regulation of immune system function. Their encoded products play pivotal roles in immune responses, inflammatory reactions, immune cell differentiation, and the regulation of immune system equilibrium. Although the exact cause of IPF remains unclear, substantial evidence suggests that IRGs may play a pivotal role in the progression of fibrosis[2, 10]. IRGs influence the progression of fibrosis through multiple mechanisms, including cytokine production, regulation of macrophage function, cell-matrix interactions, spatio-temporal coordination of the healing process, and mechanical signal transduction[11]. Fan et al. developed a prognostic model for pulmonary fibrosis using autophagy-related genes, achieving an Area Under Curve (AUC) of 0.864 for 3-year survival rates[12]. Tan et al. identified key immune- and inflammation-related genes for IPF diagnosis, with an AUC of 0.7[13]. Additionally, Elevated concentrations of pro-inflammatory mediators, including TNF-α and IL-8, are consistently observed in the pulmonary microenvironment of patients with idiopathic pulmonary fibrosis[14]. In conclusion, further exploring the action pathways of IRGs and related proteins may help elucidate the pathogenesis of IPF and improve the accuracy of prognosis prediction.

This investigation seeks to elucidate the functional involvement of immune-related genes IRGs in IPF pathogenesis and systematically and identify IRGs associated with IPF prognosis through machine learning. We established molecular subtypes of IPF and used expression profiles and multivariate analysis coefficients of four independent prognostic factors (PPBP, SFTPD, CCL7, and ADM) to develop and validate prognostic signatures for IPF. Furthermore, we re-verified the abnormal expression of these four independent prognostic factors in both BLM-induced in vitro and in vivo models of pulmonary fibrosis. Our results elucidate novel mechanistic insights into the crucial involvement of immune-related genes in IPF pathogenesis and offer potential targets for prognosis prediction and future clinical applications.

2. Materials and methods

2.1. IPF dataset collection

In this study, we obtained transcriptomic sequencing data for IPF and normal samples from the GSE70866 dataset, which included 20 normal samples and 176 IPF samples for subsequent analysis. To mitigate batch effects, we employed the "SVA" R package, and for data normalization, we utilized the

"limma" package during the processing of matrix data derived from these platforms. Clinical data for both IPF and normal samples were extracted from the matrices and included key information such as survival time, survival status, and various clinical pathological parameters.

2.2. Identification of Prognostic-Associated Immune-Related Genes and Unsupervised Consensus Clustering Analysis

To identify potential prognostic IRGs, we downloaded the complete set of IRGs from ImmPort (https://doi.org/10.21430/9pd6-z002) and used a "venn" script to extract the intersecting genes between differentially expressed genes (DEGs) and IRGs. Subsequently, we employed the "survminer" package to perform univariate Cox regression analysis, calculating the hazard ratios (HRs) and corresponding *P*-values for the DE-IRGs, with a significance threshold set at $P < 0.05$. Subsequently, we performed LASSO regression to refine the prognostic variables, followed by multivariate Cox regression analysis to identify independent prognostic factors. To explore the molecular classification of IPF utilizing these independent prognostic determinants, we used the "ConsensusClusterPlus" package for unsupervised consensus clustering analysis. The optimal cluster configuration was established through comprehensive assessment of clustering robustness, incorporating consensus values, cumulative distribution function (CDF) analysis, and quantification of ambiguous clustering (PAC) proportions, with k ranging from 2 to 9. Based on this analysis, IPF samples were divided into distinct IRG-based subtypes. Ultimately, we employed the "survival" package to evaluate prognostic differences among the identified molecular subtypes.

2.3. Establishment and Validation of an IRG Prognostic Signature for IPF

We calculated the IRG score for each IPF sample employing expression profiles and weighted regression coefficients to PPBP, SFTPD, CCL7, and ADM. Subsequently, we stratified the IPF samples into low- and high-IRG score subgroups using the median IRG score as a cutoff. To evaluate the robustness of the IRG score, The IPF cohort was randomly stratified into training and validation subsets in a 7:3 ratio through implementation of the "caret" package.. The clinical outcomes of patients within these IRG score subgroups were then analyzed using the "survival" package. Time-dependent ROC curves were generated, and the AUC values for 1-, 3-, and 5-year survival were calculated using the "survivalROC" package. Finally, the "ggalluvial" package was utilized to investigate potential associations between clinical survival outcomes, IRG scores, and IRG-based subtypes.

2.4. Immune infiltration analysis

ssGSEA is an extension of the gene set enrichment analysis (GSEA) method and is now widely used in bioinformatics studies related to immune infiltration[15]. We utilized the "GSVA" R package to perform ssGSEA, calculating the enrichment scores of 28 immune cell signature gene sets for each sample. The enrichment scores reflect the relative abundance of specific immune cell types in the sample, indicating

the infiltration levels of different immune cell types in the tissue[16]. The gene sets for the 28 tumor-infiltrating lymphocytes were downloaded from the TISIDB database (http://cis.hku.hk/TISIDB/data/download/CellReports.txt).

2.5. Prognostic independence evaluation and predictive nomogram development

We performed both univariate and multivariate Cox regression analyses to assess the independent prognostic value of the IRG score, age, and gender in IPF. Through multivariate integration of these parameters, we assessed their prognostic independence and quantified their individual contributions to clinical outcomes.. Additionally, Utilizing the "rms" package, we developed a predictive nomogram to quantify 1-, 3-, and 5-year survival probabilities in IPF patients according to their clinical and molecular profiles We utilized the "pROC" package to evaluate the diagnostic performance of the IRG score, age, and gender.

2.6. Cell culture and treatment

Human embryonic lung fibroblasts (MRC-5) were purchased from the Cell Resource Center and cultured in MRC-5 special culture medium at 37°C and 5% CO_2. MRC-5 cells induced by 20 μg/mL BLM (Cat. 9041-93-4, MedChemExpress, NJ, USA) were used as a model of pulmonary fibrosis cells.

2.7. Western blot

Protein extraction was performed with RIPA lysis buffer, after which the samples were subjected to SDS-PAGE electrophoresis and electroblotted onto PVDF membranes. The membranes were then probed with target-specific primary antibodies., followed by secondary antibodies, and then observed using the Odyssey CLx imaging system (Li-Cor, America). The specific primary antibodies used in this study, including PPBP (Cat No. 13313-1-AP), SFTPD (Cat No. 11839-1-AP), ADM (Cat No. 10778-1-AP), and GAPDH (Cat No. 10494-1-AP) antibody were purchased from Proteintech (Chicago, IL, USA). Band intensities were quantified using Quantity One V 4.62 software (Bio-Rad, Life Science, California, America).

2.8. Enzyme-linked immunosorbent assay (ELISA) Analysis

The protein levels in cell or tissue homogenates were measured using the PPBP ELISA Kit(Cat.EHPPBP, Thermofisher, Waltham, MA, USA) following the manufacturer's protocol. In brief, the 96-well ELISA plates were coated with a capture antibody specific to PPBP and incubated at 4°C overnight. Following three washes with Phosphate Buffered Saline (PBS) containing 0.05% Tween-20 (PBST), the microplates were incubated with 1% bovine serum BSA in PBST as a blocking buffer for 60 minutes at ambient temperature. Samples and reference standards were aliquoted into the microplate wells and subjected to a 2-hour incubation period at ambient temperature. Subsequently, five PBST washing cycles were performed to remove unbound components, followed by the addition of a biotinylated detection antibody specific to PPBP and incubation for 1 hour at room temperature. Following subsequent washes, the wells were treated with horseradish peroxidase (HRP)-conjugated streptavidin for 30 minutes. The

substrate solution was added to each well and the reaction was stopped with 2M sulfuric acid after 15 minutes. Optical density measurements were obtained at 450 nm wavelength using a microplate spectrophotometer. Sample PPBP concentrations were quantified through interpolation from a standard curve derived from reference solutions.

2.9. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis

The RNA was extracted from cell samples using the TRIzol reagent (Cat. 15596026, Invitrogen, Carlsbad, CA, USA) based on the manufacturer's protocols. The cDNA was synthesized by utilizing the PrimeScriptTM RT reagent kit (Cat. RR047A, Takara, Tokyo, Japan). The qRT-PCR analysis was performed by executing triplicate PCRs for each sample in an Mx3000P Real-Time Thermal Cycler (Stratagene, La Jolla, CA, USA). The 18S ribosomal RNA gene served as an endogenous control for reaction normalization. Relative mRNA expression levels were determined using the comparative threshold cycle method as previously established by Xue et al. et al [(17)].

2.10. Animal Experiment and IPF Mouse Model Construction

SPF-grade C57BL/6J male mice (Strain NO. N000013), aged 6–8 weeks, were purchased from Jiangsu Jicui Pharmaceutical Biotechnology Co., Ltd. (Animal License No. SCXK (Su) 2023-0009). The experimental cohort comprised 20 murine subjects. All experimental animals were maintained in the animal facility of the Institute of Medical Sciences, Chinese Academy of Sciences, under controlled environmental conditions (ambient temperature: 20–25°C; relative humidity: 40–70%) with a 12-hour photoperiod cycle and ad libitum access to standard rodent chow and water. After a 7-day acclimatization period, the experiment was initiated. The experimental protocol received ethical approval from the Institutional Animal Care and Use Committee of the Institute of Medical Sciences, Chinese Academy of Sciences (Approval No. AP2024-09-0188).

The experimental timeline was initiated (Day 0) with bleomycin administration. At baseline, mice were weight-stratified and randomly allocated into two experimental groups (n = 10/group): untreated controls and BLM-induced model animals. Model group subjects first underwent anesthetic procedures, and after the disappearance of hind-limb reflex and stabilization of respiration, BLM was intratracheally instilled to induce pulmonary fibrosis. The volume of instillation was 50 μl, and the induction dose was 1.5 U/kg body weight (BW). Murine body weights were recorded thrice weekly with concurrent survival monitoring. Following a 4-week induction period, pulmonary function assessments were conducted through invasive measurements using the flexiVent SCIREQ platform (SCIREQ Inc., Montreal, Canada). Subsequently, the mice were euthanized, and lung tissues were collected for histopathological analysis, protein concentration determination, immunohistochemical analysis, and qRT-PCR.

The collected tissue samples were fixed in 4% paraformaldehyde (Cat. 441244-1KG, St. Louis, MO, USA) for 48 hours, followed by routine dehydration, clarification, and paraffin embedding. The embedded

tissue blocks were placed on the Leica HistoCore AUTOCUT (Leica, Wetzlar, Germany) for continuous sectioning. The sections were baked at 60°C for 1–2 hours to enhance adhesion and stored at room temperature for further use.

2.11. Pulmonary Function Testing

Mice were anesthetized through intraperitoneal administration of pentobarbital sodium to achieve respiratory suppression. Subsequent to tracheal cannulation, subjects were mechanically ventilated using the flexiVent apparatus (Cat. FV-FXM2/4, SCIREQ, Montreal, QC, Canada). The computer-regulated rodent ventilator was programmed to deliver standardized ventilation parameters: tidal volume of 10 mL/kg body weight, respiratory frequency of 150 breaths/min, and positive end-expiratory pressure maintained at 2 cm H_2O, ventilating the mice in a quasi-sinusoidal manner to approximate the mean lung volume during spontaneous breathing. On the flexiVent, a Snapshot perturbation was performed. Preceding each perturbation, a total lung capacity (TLC) recruitment maneuver was performed to standardize pulmonary volume history. The Snapshot perturbation protocol was iteratively administered until three technically satisfactory measurements (coefficient of determination > 0.95) were acquired per subject, with subsequent derivation of mean values for analysis. Quantified respiratory parameters encompassed inspiratory capacity (IC), pulmonary elastance, and the forced expiratory volume at 0.1 seconds to forced vital capacity ratio ($FEV_{0.1}$/FVC).

2.12. Hematoxylin and Eosin Staining (HE)

Hematoxylin and Eosin staining was meticulously performed in strict accordance with the manufacturer's protocol using Hematoxylin and Eosin (Cat. ab245880, Cambridge, MA, USA). Post-staining, the sections were mounted with Tissue-Tek® Film (Sakura Finetek, Tokyo, Japan). High-resolution images were subsequently captured using the Olympus SLIDEVIEW VS200 microscope (Olympus, Tokyo, Japan) to evaluate the fundamental tissue architecture, cellular morphological characteristics, and potential pathological alterations.

2.13. Immunohistochemistry Staining (IHC)

Tissue sections embedded in paraffin were first subjected to deparaffinization followed by sequential rehydration. Antigen retrieval was then conducted using sodium citrate buffer (pH 6) via heat-mediated antigen retrieval to fully expose antigenic sites. The sections were treated with 3% bovine serum albumin (BSA) at 4°C for blocking nonspecific interactions, followed by overnight incubation at 4°C with the primary antibody against PPBP (Cat. No. 13313-1-AP, Proteintech, Chicago, IL, USA). On the following day, the sections were subsequently treated with a HRP-conjugated secondary antibody for 60 minutes at ambient temperature. Finally, signal amplification was achieved using a Diaminobenzidine (DAB) substrate kit (Cat. 34002, Thermo Fisher Scientific, Waltham, MA, USA), and nuclei were counterstained with hematoxylin to enhance contrast.

2.14. Masson's Trichrome Staining

Masson's Trichrome staining was performed in strict adherence to the manufacturer's instructions using the Masson's Trichrome Stain Kit (Cat. HT15, Sigma-Aldrich, St. Louis, MO, USA). Following staining, collagen fibers were distinctly visualized in blue, facilitating clear observation of collagen fiber distribution within the tissue. High-resolution images were captured using the Olympus SLIDEVIEW VS200 microscope (Olympus, Tokyo, Japan) for subsequent histological analysis.

2.15. Statistical analysis

All statistical computations were conducted using R software (v4.1.1), GraphPad Prism (v8.0.1), and SPSS 18.0 (SPSS Inc., Chicago, IL, USA). Quantitative PCR results, derived from three independent experimental replicates, were analyzed via one-way ANOVA, Student's t-test, and Wilcoxon rank-sum test. A threshold of $P < 0.05$ (adjusted for multiple comparisons) was applied to determine statistical significance.

3. Results

3.1 The identification of IRGs related to prognosis in IPF

To examine the involvement of IRGs in IPF, differential gene expression analysis between healthy controls and IPF patients was conducted using the "limma" package in R. Our analysis revealed that 20,187 genes were differentially expressed between these two groups at a significant threshold of $P < 0.05$ (Fig. 1A). A heatmap was constructed to illustrate the expression of the top 25 upregulated and downregulated DEGs in healthy controls (HC) and IPF groups (Fig. 1B). Comprehensive data are presented in Supplementary Table S1. A Venn diagram was employed to identify the intersection between IRGs and DEGs, resulting in a set of 48 genes selected for further study (Fig. 1C). To assess the potential prognostic value of these IRGs in IPF, we conducted univariate Cox regression analysis incorporating clinical survival data from IPF samples, determining prognostic factors correlated with IPF disease outcomes (Fig. 1D, HR > 1, $P < 0.05$). The complete results of the univariate Cox regression analysis for all genes are available in Supplementary Table S2. Additionally, we performed LASSO analysis to further refine the feature selection, discovering that the minimum log lambda value occurred when the number of variables was eight (Figs. 1E-F). For details, see Supplementary Tables S3 and S4. Ultimately, our multivariate Cox regression analysis indicated that four IRGs—PPBP, SFTPD, CCL7, and ADM—may serve as independent prognostic factors for IPF

3.2 The molecular subtyping characteristics of IPF

We employed unsupervised consensus clustering analysis to characterize molecular heterogeneity of IPF based on the expression profiles of four independent prognostic factors. The consensus clustering model was most reliable when k = 2, as IPF samples were divided into two IRGs molecular subgroups (Fig. 2A). The clinical prognostic survival curves demonstrated that IPF samples in IRGs subgroup A had significantly better survival outcomes compared to those in subgroup B (Fig. 2B). The raw data underlying the survival analysis are available in Supplementary Table S5. The principal component

analysis (PCA) plot revealed two distinct distribution patterns for IRGs subgroups A and B (Fig. 2C). Source data are in Supplementary Table S6. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed substantial enrichment of multiple immune-associated pathways, particularly cytokine-cytokine receptor interactions, focal adhesion, and exocytosis (Fig. 2D). The complete list of enriched pathways is available in Supplementary Table S7. Comparative immune profiling revealed significantly elevated infiltration levels of multiple immune cell populations (including activated B lymphocytes, $CD4^+$ and $CD8^+$ T cell subsets, eosinophils, and neutrophils) in the high-risk IRG cluster B compared to cluster A (Fig. 2E). Detailed quantification is provided in Supplementary Table S8. Detailed quantification is provided in Supplementary Table S8.We also observed significant correlations between PPBP, SFTPD, CCL7, ADM, and most immune cells. The immune correlation analysis revealed potential associations between the four independent prognostic factors and 23 immune cells (Fig. 2F).Complete correlation data are in Supplementary Table S9. In summary, our results suggest that the high IRGs risk subgroup with poor prognosis has a more pronounced immune status compared to the low IRGs risk subgroup, which may be associated with adverse clinical outcomes.

3.3 Construction of a prognostic model based on IRGs

Based on the expression values of four independent prognostic IRGs (PPBP, SFTPD, CCL7, and ADM) and their corresponding Cox regression coefficients, we developed a novel prognostic risk model for IRGs. By stratifying IPF samples into risk subgroups according to the optimal cutoff value derived from clinical prognostic outcomes, we found that the high-IRG risk subgroup was more likely to experience mortality risk (Fig. 3A). Detailed subgroup stratification data are provided in Supplementary Table S10. Additionally, clinical prognostic outcomes analysis revealed that the low IRGs risk subgroup had significantly better clinical prognoses than the high IRGs risk subgroup (Fig. 3B, $P < 0.001$). Complete prognostic data are in Supplementary Table S11. These findings suggest that the risk model, developed using four independent prognostic factors, can predict the clinical survival prognosis of IPF risk subgroups. The time-related ROC curve results for 1-, 3-, and 5-year predictions showed AUCs of 0.798, 0.604, and 0.724 (Fig. 3C). PCA demonstrated distinct clustering patterns between IRG-based risk subgroups, confirming the discriminative capacity of these prognostic signatures (Fig. 3D). Moreover, we also observed that the expression of PPBP SFTPD CCL7 and ADM was significantly upregulated in the high IRGs risk subgroup when compared with the low IRGs risk subgroup (Fig. 3E) Furthermore, we explored the distribution of IRG score in the MiAGs-based subtypes and found that the IRG score of cluster B with poor 28-day clinical prognosis was significantly higher than that in IRG cluster A (Figure. 3F). The Sankey diagram revealed the potential association between different IRGs clustering subgroups, IRGs risk subgroups, and clinical survival outcomes (Fig. 3G). These findings suggest a potential mechanistic role for the risk score constructed based on PPBP, SFTPD, CCL7, and ADM stratifies IPF samples into risk subgroups and may be associated with clinical prognosis.

3.4 Validate the accuracy of IRGs risk score in predicting the clinical prognosis of IPF

Using the "caret" R package, IPF samples from GSE70866 were randomly allocated into two independent datasets: a training set and a validation set. An IRG-based risk score model was developed to assess the independence and predictive accuracy of this model for IPF clinical prognosis. Using the clinically validated prognostic threshold, IPF specimens in both cohorts were categorized into high- and low-risk categories according to their immune-related gene risk profiles. Survival analysis revealed that in both datasets, IPF patients exhibiting elevated IRG risk scores showed markedly reduced survival probabilities relative to their low-risk counterparts(Fig. 4A, B). Moreover, time-dependent ROC curve analysis indicated that the area under the curve (AUC) values for 1-, 3-, and 5-year survival in the training set were 0.805, 0.806, and 0.845, respectively, while the corresponding values in the validation set were 0.798, 0.604, and 0.724 (Fig. 4C, D). Additionally, unsupervised principal component analysis (PCA) of the two independent datasets revealed distinct clustering patterns between the high- and low-risk IRG subgroups (Fig. 4E, F). Collectively, these results demonstrate that the IRG-based risk model provides an accurate assessment of clinical survival outcomes in IPF.

3.5 Independent prognostic analysis of IRGs score and the nomogram establishment of the IRGs score and clinical features

Given that the IRGs score can accurately assess the clinical survival outcomes of IPF, we performed additional analyses to assess its prognostic significance independent of other clinical variables. We subsequently used univariate and multivariate Cox analyses to comprehensively evaluate the HR values of the IRGs score and other clinical pathological features, as demonstrated in Fig. 5A-B. Complete statistical results are provided in Supplementary Tables S11 and S12. The results of the univariate Cox analysis suggested that gap (HR = 1.395(1.273–1.574), $P < 0.001$) and riskscore (HR = 1.301 (1.228–1.378), $P < 0.001$) were closely associated with adverse prognosis in IPF (Fig. 5A). Multivariate Cox regression analysis revealed that both the gap and risk score served as independent predictors of clinical outcomes in IPF. (Fig. 5B). The ROC curve results revealed that the AUC of the risk score was 0.807, which was significantly higher than other clinical pathological features of IPF, demonstrating a high model diagnostic ability (Fig. 5C). Furthermore, by integrating clinical feature parameters and the risk score, we developed a nomogram model to evaluate the 1-, 3-, and 5-year prognostic probabilities of IPF samples (Fig. 5D). Construction details are in Supplementary Table S14. In summary, our findings suggest that the risk model developed based on IRGs prognostic features is an independent prognostic factor for IPF that distinguishes it from clinical pathological features and can be used to accurately predict the survival outcomes of IPF.

3.6 qRT-PCR and ELISA analysis of 4 IRGs in MRC-5 cells

We further verify whether there are abnormal expressions of PPBP, SFTPD, ADM,and CCL7 in IPF. To establish the IPF model, MRC-5 cells were exposed to BLM (20 μg/mL) for 48 hours[(18)]. qRT-PCR results revealed a significant upregulation of PPBP, CCL7, and SFTPD mRNA expression in MRC-5 cells, which is consistent with previous studies (Fig. 6A). As the proportion of PPBP rising was the highest, it was selected for the following experiment. ELISA was then used to assess PPBP protein levels, and the results confirmed a marked increase in PPBP protein expression in MRC-5 cells, which aligned with the

qRT-PCR findings (Fig. 6B). Subsequently, small interfering RNA (siRNA) was used to further investigate the role of PPBP in fibroblast biology. MRC-5 fibroblast cultures were systematically allocated into two experimental conditions: untreated controls and PPBP-silenced counterparts. Cell proliferation was assessed using CCK-8 (Fig. 6C) and Wound healing assay (Fig. 6E), which showed that PPBP knockdown inhibited fibroblast proliferation. The western blot results further confirmed that the expression levels of COL1A1, α-SMA, and COL3A1 were significantly reduced following PPBP knockdown(Fig. 6D). Scratch assays revealed that PPBP knockdown also impeded fibroblast migration (Fig. 6E). qRT-PCR analysis confirmed successful knockdown of PPBP and demonstrated a corresponding reduction in fibrosis markers, including COL6A1, TGF-β1, α-SMA, SNAL1L, COL6A3, COL1A2, and COL1A1(Fig. 6F).

3.7 The expression level of PPBP was elevated in bleomycin-induced mice.

The BLM-induced murine pulmonary fibrosis model represents the most representative and widely utilized experimental system for studying IPF[18]. We further validated gene expression profiles in this established model. Pulmonary fibrosis was induced via intratracheal administration of BLM (2.5mg/kg). Mice were euthanized at day 21 post-induction for lung tissue collection. Masson's trichrome staining revealed significant collagen deposition enhancement in BLM-treated specimens (Fig. 7A). Pulmonary function assessments demonstrated substantial reductions in lung compliance and respiratory volumes in the BLM modeling group (Fig. 7B). Comparative analysis revealed that bleomycin-treated mice exhibited marked upregulation of PPBP gene expression and significant elevation of other fibrosis markers, including COL5A2, SNA1A1, TGF-β1, α-SMA, COL6A1, COL3A1, and COL1A1, compared to controls. (Fig. 7C). Western blot analysis corroborated the increased expression of these fibrotic biomarkers post-bleomycin challenge (Fig. 7D), collectively confirming successful model establishment. Subsequent immunohistochemical analysis revealed pronounced PPBP overexpression in alveolar regions of fibrotic lungs (Fig. 7E). ELISA quantification of protein levels further validated elevated PPBP expression in BLM-treated mice (Fig. 7F), consistent with our preliminary model characterization.

4. Discussion

In recent years, lung transplantation remains the sole disease-modifying intervention available for IPF[19]. Nevertheless, this therapeutic approach remains accessible to a limited subset of patients owing to the critical shortage of suitable donor organs, surgical complexity, high costs, and the advanced age of most IPF patients. Althoughsome new therapeutic targets for IPF have been developed, such as nintedanib attenuates pulmonary fibrosis by inhibiting tyrosine kinase receptors, thereby suppressing the secretion of fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), and vascular endothelial growth factor (VEGF)[20]. However, most clinical trials ultimately failed in patients with IPF, suggesting the need for further exploration of therapeutic targets[21]. In addition, the accurate stratification of the risk of IPF is one of the difficulties of treatment [22]. To address this issue, many researchers have developed numerous models for IPF risk stratification and prognosis prediction. Zheng et al. developed EMT and

immune-related gene signatures using alveolar lavage fluid cells from IPF patients to predict the prognosis of IPF, however, the limited sample size restricted the generalizability of their findings[23]. Yang et al. developed an IPF risk stratification system utilizing senescence-associated gene signatures[24]. Pokhreal et al. demonstrated that a predictive model for IPF onset and progression, constructed using pyroptosis-related genes and their underlying immunological features, achieved an AUC value of 0.91[25]. The immune-related prognostic model developed in this investigation elucidates the involvement of immune mechanisms in IPF pathogenesis, while simultaneously identifying four candidate therapeutic targets that may facilitate clinical outcome prediction in affected patients.

Our immune profiling analysis revealed substantially elevated immune cell infiltration across multiple subsets in high-IRG-score patients relative to their low-scoring counterparts. In addition, when comparing the results of pathway enrichment analysis in patients with high and low scores, cytokine-cytokine receptor interaction, focal adhesion, and exocytosis are highly enriched in patients with high scores. Combined with the poor prognosis of patients with higher IRG scores, these findings implicate immune cell involvement in IPF pathogenesis. We also found that β-alanine metabolism had lower enrichment in patients with higher IRG scores. β-alanine and its metabolite carnosine play an important role in the body as a buffer and antioxidant[26]. Reestablishment of redox homeostasis represents a critical therapeutic target in IPF management[27]. The potential protective effect of β-alanine on IPF deserves further study.

PPBP, as a chemokine, exhibits specific expression across multiple infectious diseases[28], and the dysregulated expression of PPBP is closely associated with various inflammatory disorders[29]. Functionally, PPBP serves as a key immunoregulatory mediator, facilitating humoral immunity and promoting neutrophil activation.[30]. PPBP facilitates the progression of pulmonary fibrosis through multiple mechanisms. PPBP can recruit neutrophils to the inflammatory area in the lungs. By activating these neutrophils, pro-inflammatory mediators and reactive oxygen species (ROS) are released, damaging lung epithelial cells and fibroblasts, thereby creating a pro-fibrotic environment[31].Single-cell transcriptomic analyses have revealed that within the pulmonary microenvironment of IPF patients, the CCR2 + pro-inflammatory macrophage subset exhibits high PPBP expression, with an expression level 3.8 times significantly elevated compared to healthy control subjects[6]. In the BLM-induced pulmonary fibrosis model, collagen deposition in PPBP knockout mice decreased by 42%, along with reduced neutrophil infiltration[32]. Notably, immunofluorescence analysis indicated that the PPBP protein was predominantly localized in vascular endothelial cells and activated alveolar macrophages surrounding fibrotic lesions, and its spatial distribution exhibited significant co-localization with α-SMA + myofibroblasts[33]. At the molecular mechanism level, PPBP activates a dual signaling pathway by binding to the CXCR2 receptor. Specifically, it promotes fibroblast proliferation via the PI3K/Akt/mTOR pathway[33]. Our study confirmed through *in vivo* and *in vitro* experiments that PPBP is abnormally overexpressed in IPF, and that knockdown of PPBP significantly inhibits fibroblast proliferation and migration. This proves the close association between PPBP and IPF.

Due to the limited conditions, we lack of in-depth mechanism research on the selected targets. Further study of the molecular mechanism will help deepen current knowledge regarding IPF disease mechanisms. In addition, the obtained results are concentrated in bioinformatics databases and cell line studies, and lack of multi-center and large sample analysis. Due to the complexity and heterogeneity of IPF, the targets selected in this paper need to be further verified in multi-centers.

Conclusion

We developed and rigorously validated an immune-related gene signature for prognostic risk stratification in IPF, enabling enhanced clinical utility of immune markers for both diagnostic and prognostic assessment in IPF management.. Furthermore, this work elucidates the pivotal role of PPBP in IPF pathogenesis, offering novel perspectives and potential therapeutic targets for prognosis prediction and future clinical interventions. Experimental validation using ELISA and IHC confirmed the expression levels of core genes in both bleomycin-induced pulmonary fibrosis models and PPBP-knockdown cellular models. However, further investigations are warranted to comprehensively unravel the biological significance and underlying mechanisms of these findings.

Declarations

Data availability statement

The datasets used and analyzed during the current study are available from thecorresponding author upon reasonable request.

Acknowledgments

We express our gratitude to all individuals who participated in this study.

Funding

This work was supported by the Natural Science Foundation of Zhejiang province, China (LHDMZ25H29003) and the Cultivation Fund of the National Natural Science Foundation of Zhejiang Cancer Hospital (No. PY2023008).

Author information

Jie Xuan, Wenyuan Niu, Gaofeng Hu, Yupeng Zhang have contributed equally to this work.

Authors and Affiliations

School of Pharmaceutical Sciences, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China.

Wenyuan Niu, Yupeng Zhang, Junjie Xia

Zhejiang Cancer Hospital, Hangzhou, Zhejiang, China

Jie Xuan Gaofeng Hu Zhen Zhang Zhihao Lin Qinglin Li Yuanqiang Li

Respiratory Department, Affiliated Hospital of Hangzhou Normal University

Changqing Xu

Contributions

Jie Xuan, Wenyuan Niu, Gaofeng Hu, Yupeng Zhang: Writing–original draft, Visualization, Investigation. Zhen Zhang, Junjie Xia and Shang Ma: Investigation, Data curation. Zhihao Li: Methodology, Data curation. Changqing Xu: Writing–original draft, Project administration. Qinglin Li: Conceptualization, Writing–original draft. Yuanqiang Li: Writing–review & editing, Supervision.

Corresponding author

Correspondence to Changqing Xu, Qinglin Li or Yuanqiang Li.

Ethics declarations

Ethics Approval and Consent to Participate

This study has obtained the approval for animal research from the Institutional Animal Care and Use Committee (IACUC) of the Hangzhou Institute of Medicine, Chinese Academy of Sciences (Approval No.: AP2024-09-0188).

Consent for Publication

Consent for publication has been obtained from all authors. Transcriptomic sequencing data of this paper were obtained from the public GEO database GSE70866. The source papers publishing this dataset were approved by the respective local ethics committees and registered with the German Clinical Trials Register.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Noble PW, Barkauskas CE, Jiang D. Pulmonary fibrosis: patterns and perpetrators. J Clin Invest. 2012;122(8):2756-62.
2. Wang Q, Xie ZL, Wu Q, Jin ZX, Yang C, Feng J. Role of various imbalances centered on alveolar epithelial cell/fibroblast apoptosis imbalance in the pathogenesis of idiopathic pulmonary fibrosis.

Chin Med J (Engl). 2021;134(3):261-74.
3. Kahloon RA, Xue J, Bhargava A, Csizmadia E, Otterbein L, Kass DJ, et al. Patients with idiopathic pulmonary fibrosis with antibodies to heat shock protein 70 have poor prognoses. Am J Respir Crit Care Med. 2013;187(7):768-75.
4. Taillé C, Grootenboer-Mignot S, Boursier C, Michel L, Debray MP, Fagart J, et al. Identification of periplakin as a new target for autoreactivity in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;183(6):759-66.
5. Jin C, Li J, Li Q, Zhang L, Zheng S, Feng Q, et al. Contribution of cuproptosis and immune-related genes to idiopathic pulmonary fibrosis disease. Front Immunol. 2025;16:1458341.
6. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. Eur Respir J. 2019;54(2).
7. Misharin AV, Morales-Nebreda L, Reyfman PA, Cuda CM, Walter JM, McQuattie-Pimentel AC, et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. J Exp Med. 2017;214(8):2387-404.
8. Lyu Y, Guo C, Zhang H. Fatty acid metabolism-related genes in bronchoalveolar lavage fluid unveil prognostic and immune infiltration in idiopathic pulmonary fibrosis. Front Endocrinol (Lausanne). 2022;13:1001563.
9. Yu Y, Liu X, Xue Y, Li Y. Identification of immune-related genes for the diagnosis of ischemic heart failure based on bioinformatics. iScience. 2023;26(11):108121.
10. Deng L, Huang T, Zhang L. T cells in idiopathic pulmonary fibrosis: crucial but controversial. Cell Death Discov. 2023;9(1):62.
11. Pakshir P, Hinz B. The big five in fibrosis: Macrophages, myofibroblasts, matrix, mechanics, and miscommunication. Matrix Biol. 2018;68-69:81-93.
12. Fan G, Liu J, Wu Z, Li C, Zhang Y. Development and validation of the prognostic model based on autophagy-associated genes in idiopathic pulmonary fibrosis. Front Immunol. 2022;13:1049361.
13. Tan Y, Qian B, Ma Q, Xiang K, Wang S. Identification and Analysis of Key Immune- and Inflammation-Related Genes in Idiopathic Pulmonary Fibrosis. J Inflamm Res. 2025;18:1993-2009.
14. Kapanci Y, Desmouliere A, Pache JC, Redard M, Gabbiani G. Cytoskeletal protein modulation in pulmonary alveolar myofibroblasts during idiopathic pulmonary fibrosis. Possible role of transforming growth factor beta and tumor necrosis factor alpha. Am J Respir Crit Care Med. 1995;152(6 Pt 1):2163-9.
15. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108-12.
16. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. Nat Commun. 2018;9(1):5361.
17. Xue GP, Kooiker M, Drenth J, McIntyre CL. TaMYB13 is a transcriptional activator of fructosyltransferase genes involved in beta-2,6-linked fructan synthesis in wheat. Plant J.

2011;68(5):857-70.
18. Gul A, Yang F, Xie C, Du W, Mohammadtursun N, Wang B, et al. Pulmonary fibrosis model of mice induced by different administration methods of bleomycin. BMC Pulm Med. 2023;23(1):91.
19. Glass DS, Grossfeld D, Renna HA, Agarwala P, Spiegler P, DeLeon J, et al. Idiopathic pulmonary fibrosis: Current and future treatment. Clin Respir J. 2022;16(2):84-96.
20. Roach KM, Castells E, Dixon K, Mason S, Elliott G, Marshall H, et al. Evaluation of Pirfenidone and Nintedanib in a Human Lung Model of Fibrogenesis. Front Pharmacol. 2021;12:679388.
21. Spagnolo P, Maher TM. Clinical trial research in focus: why do so many clinical trials fail in IPF? Lancet Respir Med. 2017;5(5):372-4.
22. Albera C, Ferrero C, Rindone E, Zanotto S, Rizza E. Where do we stand with IPF treatment? Respir Res. 2013;14 Suppl 1(Suppl 1):S7.
23. Zheng J, Dong H, Zhang T, Ning J, Xu Y, Cai C. Development and Validation of a Novel Gene Signature for Predicting the Prognosis of Idiopathic Pulmonary Fibrosis Based on Three Epithelial-Mesenchymal Transition and Immune-Related Genes. Front Genet. 2022;13:865052.
24. Yang C, Han Z, Zhan W, Wang Y, Feng J. Predictive investigation of idiopathic pulmonary fibrosis subtypes based on cellular senescence-related genes for disease treatment and management. Front Genet. 2023;14:1157258.
25. Pokhreal D, Crestani B, Helou DG. Macrophage Implication in IPF: Updates on Immune, Epigenetic, and Metabolic Pathways. Cells. 2023;12(17).
26. Brosnan ME, Brosnan JT. Histidine Metabolism and Function. J Nutr. 2020;150(Suppl 1):2570s-5s.
27. Pan L, Cheng Y, Yang W, Wu X, Zhu H, Hu M, et al. Nintedanib Ameliorates Bleomycin-Induced Pulmonary Fibrosis, Inflammation, Apoptosis, and Oxidative Stress by Modulating PI3K/Akt/mTOR Pathway in Mice. Inflammation. 2023;46(4):1531-42.
28. Glibetic N, Shvetsov YB, Aan FJ, Peplowska K, Hernandez BY, Matter ML. Transcriptome profiling of colorectal tumors from patients with sepsis reveals an ethnic basis for viral infection risk and sepsis progression. Sci Rep. 2022;12(1):20646.
29. Bdeir K, Gollomp K, Stasiak M, Mei J, Papiewska-Pajak I, Zhao G, et al. Platelet-Specific Chemokines Contribute to the Pathogenesis of Acute Lung Injury. Am J Respir Cell Mol Biol. 2017;56(2):261-70.
30. Wang D, Zhang Z, Cui S, Zhao Y, Craft S, Fikrig E, et al. ELF4 facilitates innate host defenses against Plasmodium by activating transcription of Pf4 and Ppbp. J Biol Chem. 2019;294(19):7787-96.
31. Graca FA, Stephan A, Minden-Birkenmaier BA, Shirinifard A, Wang YD, Demontis F, et al. Platelet-derived chemokines promote skeletal muscle regeneration by guiding neutrophil recruitment to injured muscles. Nat Commun. 2023;14(1):2900.
32. Lin PH, Liang CY, Yao BY, Chen HW, Pan CF, Wu LL, et al. Robust induction of T(RM)s by combinatorial nanoshells confers cross-strain sterilizing immunity against lethal influenza viruses. Mol Ther Methods Clin Dev. 2021;21:299-314.

33. Skendros P, Mitsios A, Chrysanthopoulou A, Mastellos DC, Metallidis S, Rafailidis P, et al. Complement and tissue factor-enriched neutrophil extracellular traps are key drivers in COVID-19 immunothrombosis. J Clin Invest. 2020;130(11):6151-7.

Table

Table1: In the multivariate Cox regression analysis, the coef of four IRGs.

Id	Coef
PPBP	1.61413163246783
SFTPD	1.64277774988535
CCL7	1.12052997761835
ADM	2.90513312628096

Figures

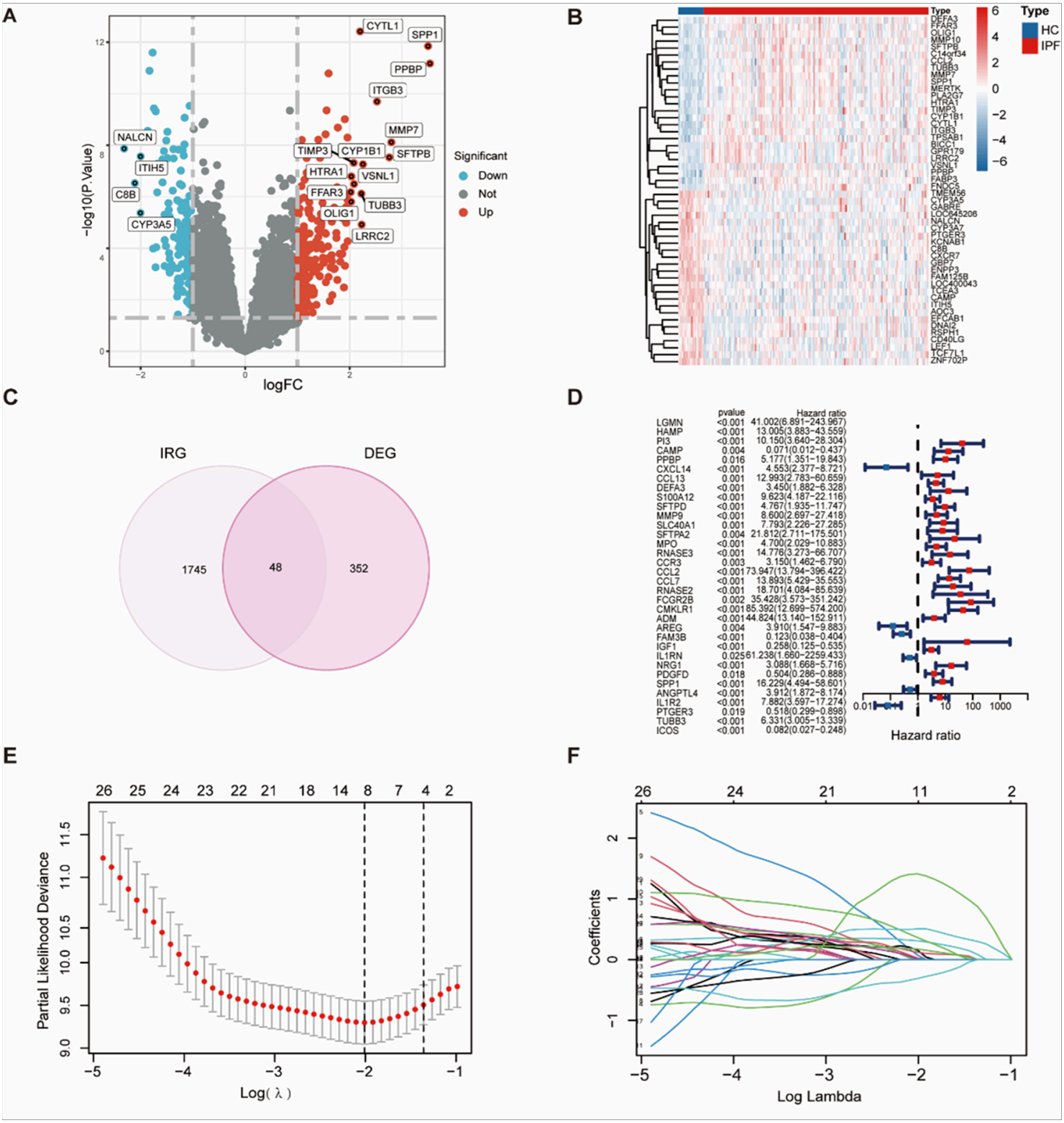

Figure 1

Identification of Prognostically Associated IRGs. (A) Volcano plot depicting DEGs between HC and IPF groups. DEGs were filtered with thresholds set at |fold change| ≥2 and *P* (adjusted) <0.05. Teal dots denote downregulated genes, and red dots indicate upregulated genes. (B) Analysis of the top 25 upregulated and downregulated genes in the HC and IPF groups. (C) Venn diagram showing IRGs differentially expressed as selected by WGCNAanalysis. (D) Univariate Cox analysis indicating the

association of differentially expressed IRGs with prognosis. (E) Performance evaluation of LASSO regression through cross-validation. (F) Feature selection trajectory demonstrating coefficient shrinkage in LASSO regression.

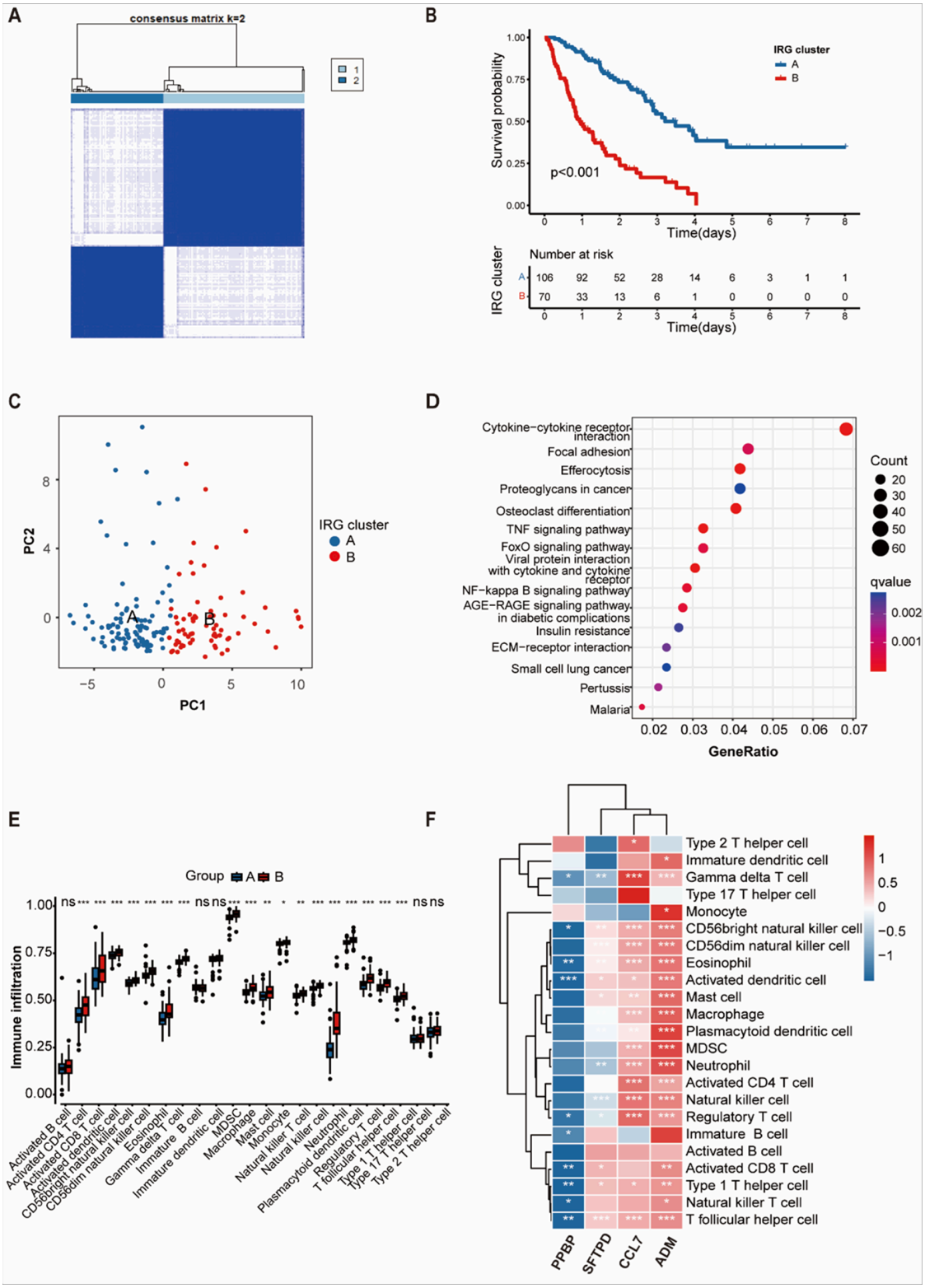

Figure 2

Characterization and pathway enrichment profiling of immune-related gene clusters in IPF cohorts. (A) Analysis of subtypes based on IRGs. (B) Clinical prognostic analysis of IRGs subtypes. (C) Unsupervised

PCA analysis. (D) KEGG analysis among differentially expressed immune-related genes across molecular subtypes. (E) Immune infiltration features of IRGs subtypes. (F) Correlation analysis between prognostic gene signatures and immune cell infiltration patterns.

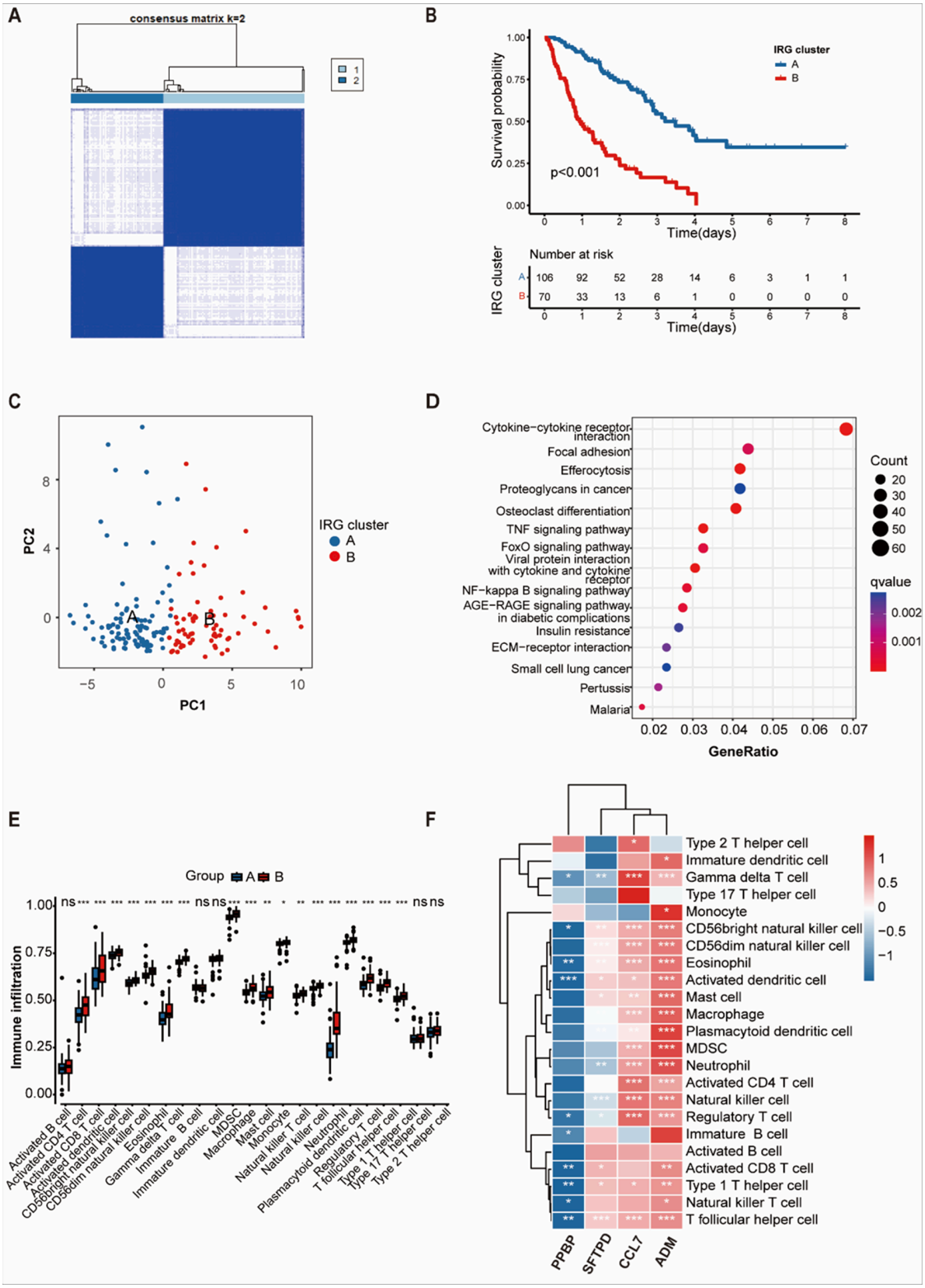

Figure 3

IRG risk score evaluation and prognostic model construction. (A) Stratification of IRGs into risk subgroups. (B) Clinical prognosis outcome analysis of IRGs risk subgroups. (C) Time-related ROC curve

analysis. (D) PCA analysis based on IRG prognostic features. (E) Expression profile analysis of four independent prognostic factors in IRGs risk subgroups. (F) Analysis based on IRG prognostic features. (G) Sankey diagram showing the association between IRG clustering subgroups, risk subgroups, and clinical prognosis.

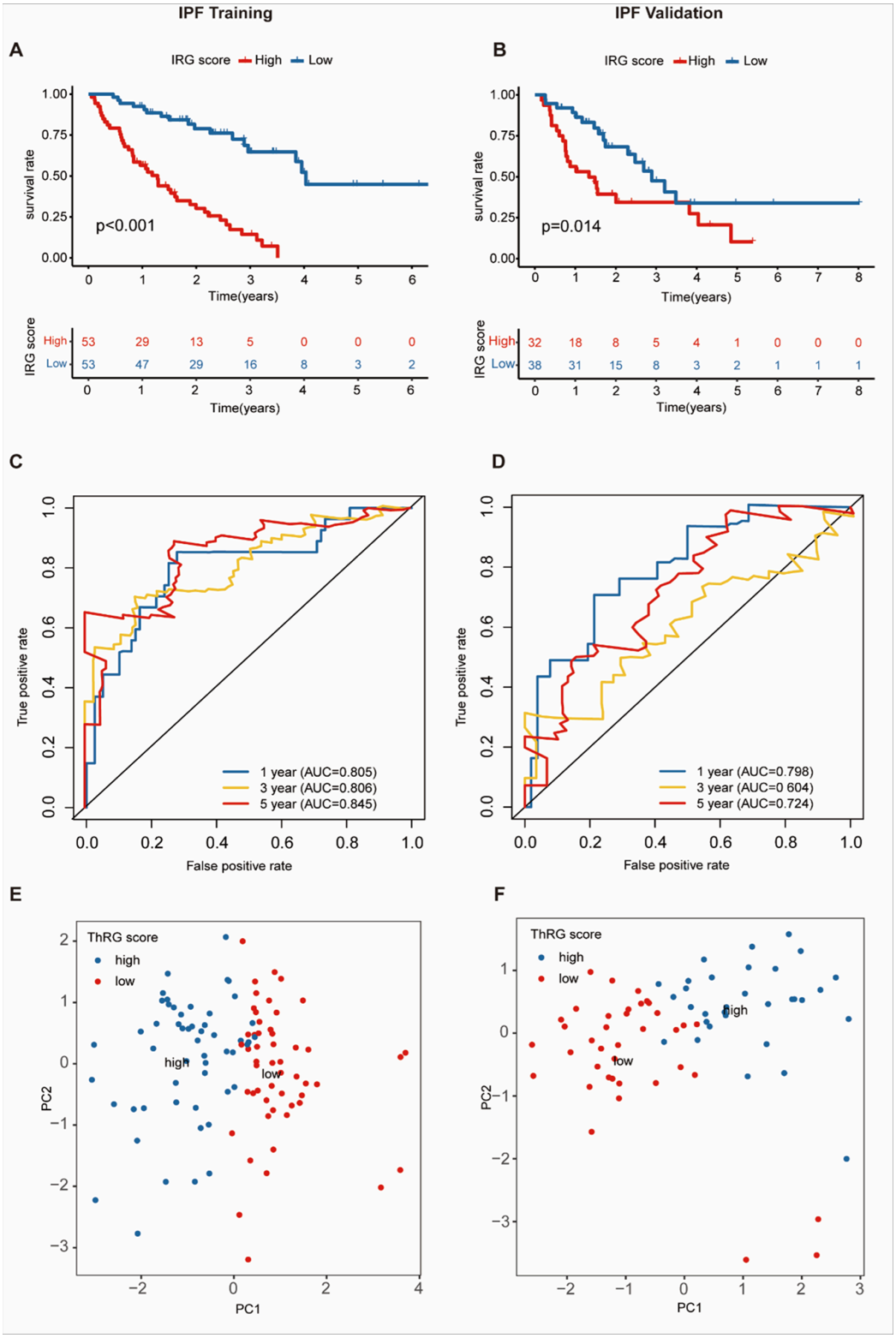

Figure 4

Risk model construction for IRGs in independent cohorts. (A, B) Survival outcome assessment between high- and low-risk IPF subgroups in the training (A) and validation (B) cohorts. (C, D) Analysis of time-dependent ROC curves in two independent cohorts. (E, F) PCA analysis of the training and validation sets.

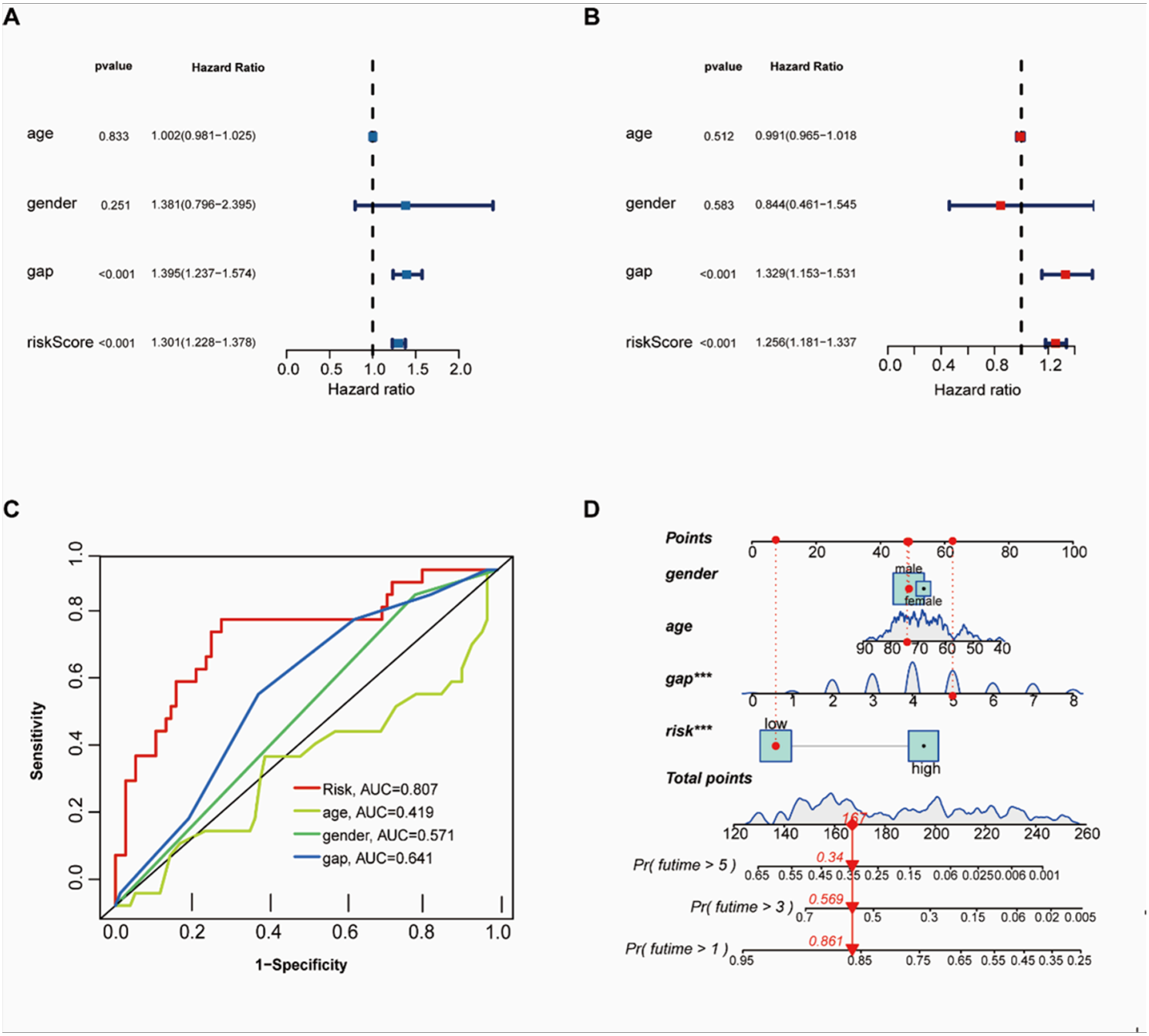

Figure 5

Independent prognostic analysis of risk score and various clinical pathological features. (A, B) HR and P values for the risk score and various clinicopathological parameters were derived from univariate and multivariate Cox proportional hazards regression analyses. (C) ROC curve analysis of riskscore, age, gender, and gap. (D) Development of a nomogram model based on riskscore and clinical pathological parameters.

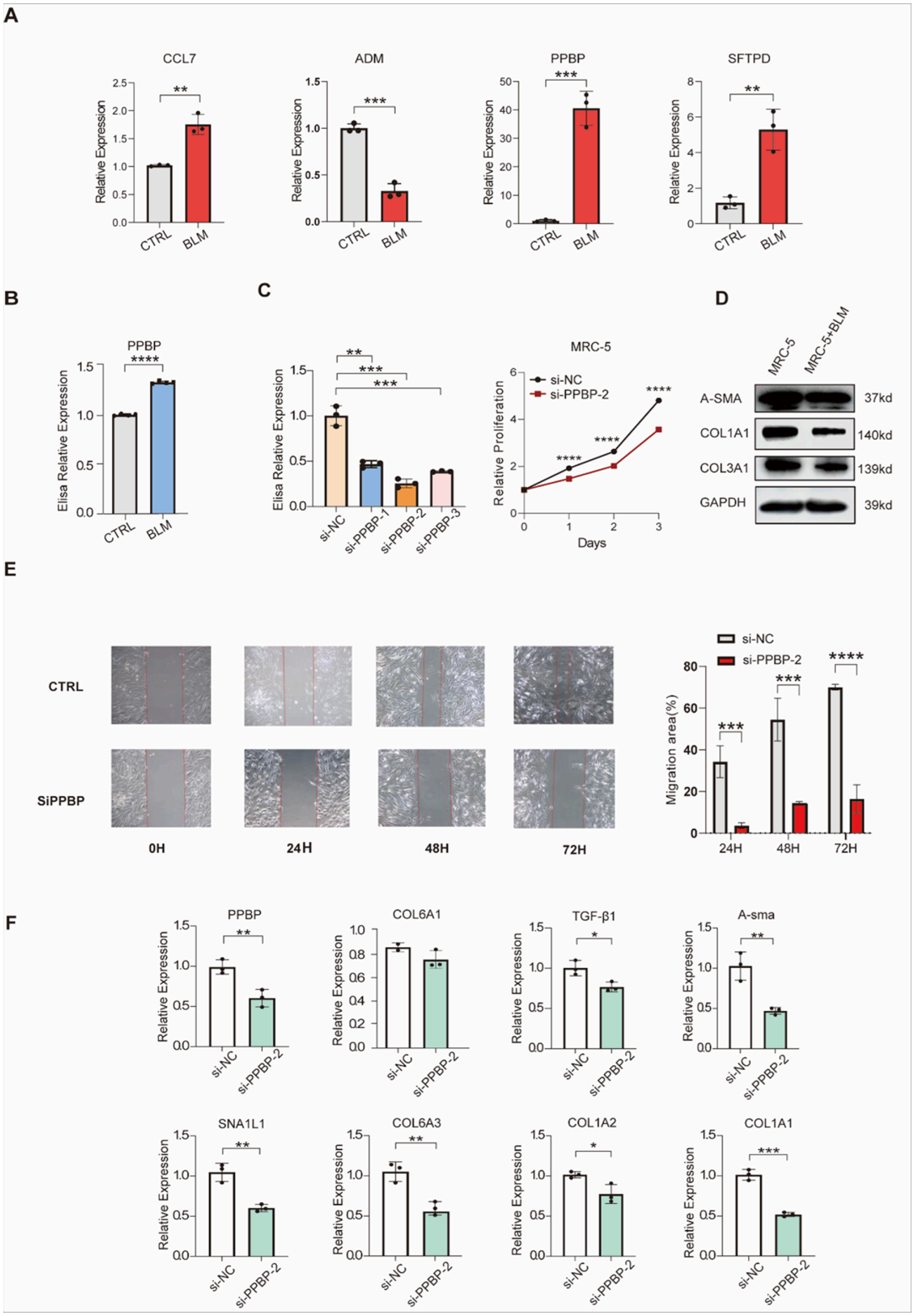

Figure 6

The knockdown of PPBP reversed the fibrotic phenotype in the MRC-5 cells. (A) mRNA expression of CCL7, ADM, PPBP, and SFTPD in MRC-5 cells (n=3). (B) ELISA analysis of PPBP expression in MRC-5 cells (n=4). (C) CCK-8 assay indicated that knockdown of PPBP inhibited fibroblast proliferation. (D) Western blot results corroborated the RT-qPCR data, demonstrating reduced production of profibrotic mediators in PPBP-deficient MRC-5 fibroblasts..(E) The cell migration and Cell motility was evaluated

through scratch wound assays, demonstrating that PPBP knockdown significantly reduced the migratory potential of MRC-5 fibroblasts. (F) Reduced expression of PPBP impacted the expression of fibrosis markers. *P<0.05, **P<0.01, ***P<0.001, ****P < 0.0001.

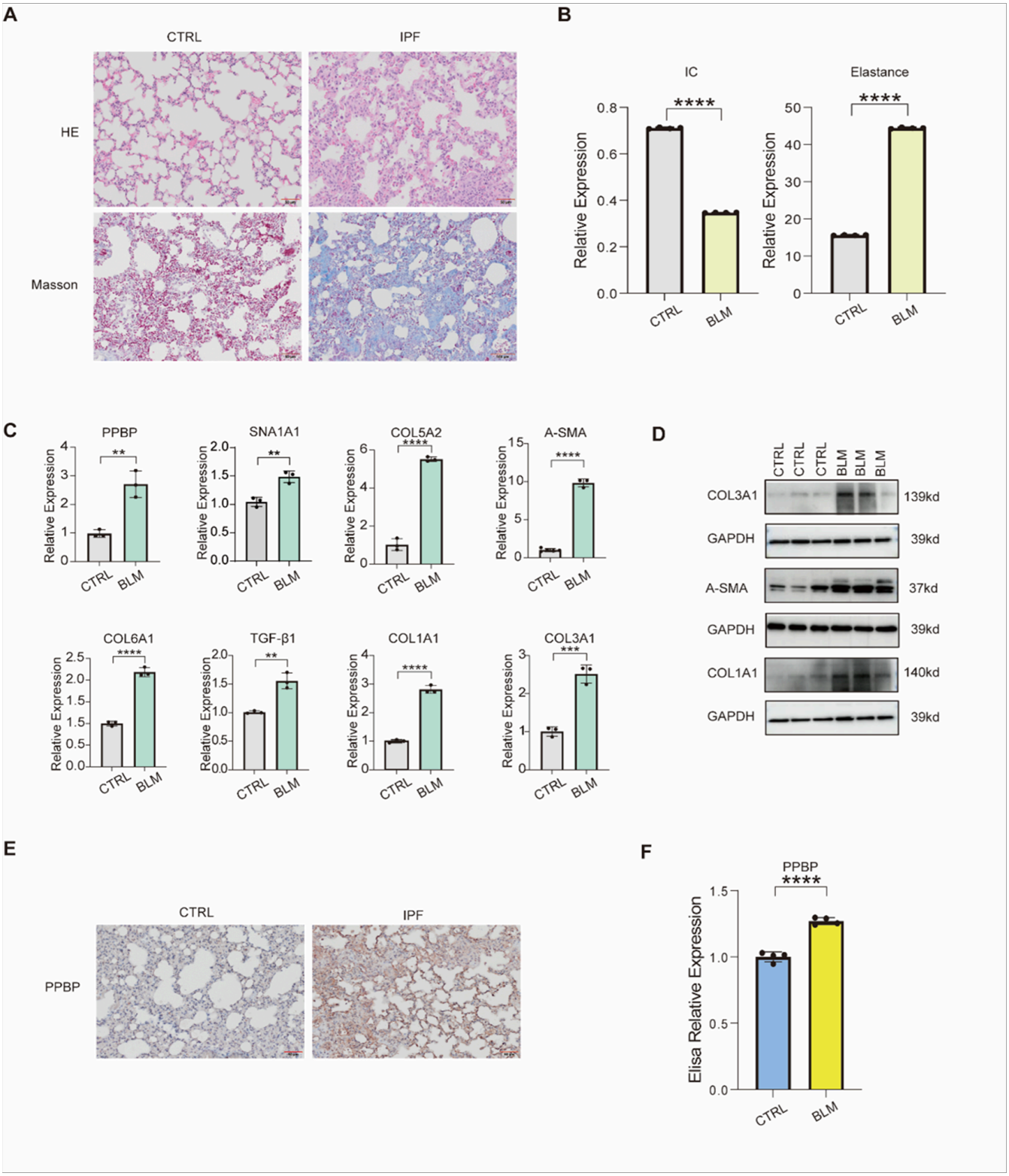

Figure 7

Changes in biomarkers in animal models. (A) Histopathological evaluation of BLM-treated murine lung tissues using hematoxylin-eosin (H&E) and Masson's trichrome staining techniques. (B) Pulmonary

function tests demonstrated significant differences in lung compliance and volume between the BLM model group and the control group. (C) RT-qPCR analysis revealed increased expression levels of fibrosis markers in the BLM model group compared to the control group. (D) Western blot analysis confirmed the RT-qPCR findings, showing upregulated fibrosis markers in the BLM model group. (E) Immunohistochemical staining illustrated the distribution of PPBP protein in lung samples from BLM-treated mice compared to controls.(F) ELISA analysis measured PPBP protein levels in lung tissues from control and BLM-treated mice. *P<0.05, **P<0.01, ***P<0.001, ****P < 0.0001.

Supplementary Files

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

- TableS1.xls
- TableS12.xlsx
- TableS5.xlsx
- TableS3.xlsx
- TableS9.xlsx
- TableS7.xlsx
- TableS13.xlsx
- TableS8.xlsx
- TableS11.xlsx
- TableS4.xlsx
- TableS10.xlsx
- TableS6.xlsx
- TableS14.xlsx
- TableS2.xls