

## Article

# Comprehensive Landscape of Matrix Metalloproteinases in the Pathogenesis of Idiopathic Pulmonary Fibrosis

Zhongzheng Li, Huabin Zhao, Shenghui Wang, Peishuo Yan, Hongmei Yuan, Mengxia Zhao, Ruyan Wan, Jiawei Wang, Zhao Zhu, Zhenwei Li, Juntang Yang, Yajun Li, Lan Wang \*

State Key Laboratory of Cell Differentiation and Regulation, Henan International Joint Laboratory of Pulmonary Fibrosis, Henan Center for Outstanding Overseas Scientists of Pulmonary Fibrosis, College of Life Science, Institute of Biomedical Science, Henan Normal University, Xinxiang 453007, China; lizhongzheng@stu.htu.edu.cn (Z.L.); zhaohuabin@stu.htu.edu.cn (H.Z.); wangshenghui@stu.htu.edu.cn (S.W.); yanpeishuo@stu.htu.edu.cn (P.Y.); yuanhongmei@stu.htu.edu.cn (H.Y.); zhaomengxia@stu.htu.edu.cn (M.Z.); wanruyan@stu.htu.edu.cn (R.W.); wangjiawei@stu.htu.edu.cn (J.W.); zhuzhao@stu.htu.edu.cn (Z.Z.); lizhenwei@stu.htu.edu.cn (Z.L.); yangjuntang@stu.htu.edu.cn (J.Y.); liyajun@stu.htu.edu.cn (Y.L.)

\* Corresponding author. Email: wanglan@stu.htu.edu.cn (L.W.); Tel: +86-373-3326340

Received: 30 March 2023; Revised: 19 April 2023; Accepted: 28 April 2023; Available online: 19 May 2023

**ABSTRACT:** Idiopathic pulmonary fibrosis (IPF) is a progressive, chronic interstitial lung disease with unknown etiology. Matrix metalloproteinases (MMPs) are involved in fibrotic lung tissues, contributing to the initiation, progression, or resolution of chronic inflammatory disease. In present study, comprehensive changes of MMPs expressions were investigated in IPF by integrative analysis of single-cell transcriptome and bulk transcriptome data. 24 of MMPs were altered and the changes could significantly distinguish IPF from normal subjects and other lung diseases. Among them, MMP1, MMP7 and MMP19 were closely associated to lung functions, susceptibility and alveolar surface density. MMP1 and MMP7 as potential diagnostic indicators, MMP1 and MMP19 as prognostic markers in IPF could accurately predict disease progression. Devolution of MMPs at single-cell resolution, MMP19 was highly expressed in macrophages and markedly interfered with TNF signaling pathway which synchronizes fibrotic microenvironment. MMP19+ macrophages were significantly different from MMP19-macrophages in energy metabolism and immune function. The interaction of MMP19+ macrophages with hyperplastic AT2 was mediated by TNFSF12-TNFRSF12A, and further activated the TNFRSF12A receptor to affect cell glucose metabolism and mitochondrial function. In summary, MMPs has great application potential in the diagnosis, treatment, and prognosis of IPF.

**Keywords:** Matrix metalloproteinases, Idiopathic pulmonary fibrosis, RNA-seq, Macrophage



© 2023 The authors. This is an open access article under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive chronic interstitial lung disease characterized by irreversible scarring of the distal lung, leading to respiratory failure and death [1]. Mortality in IPF is high, with a reported median survival of 2–4 years from diagnosis based on historical data and the pirfenidone clinical trials reported an on-treatment median survival of 77.2 months [2]. The underlying causes are multifactorial, including late diagnosis due to lack of sensitive biomarkers, inadequate treatment options, and unclear mechanism [3]. One of the earliest and most explored events in IPF disease is fibroblast hyperactivation, further leading to excessive accumulation of extracellular matrix [1]. This process is evident in the aberrant expression of matrix metalloproteinases (MMPs) in IPF [4,5].

MMPs are zinc-dependent endopeptidases that collectively degrade all components of base membrane and the extracellular matrix (ECM) [6]. However, recent studies have shown that MMPs can modulate the activity of proteins other than ECM proteins, including inflammatory mediators, latent growth factors, antifibrotic growth factors, and dividing cell surface molecules and receptors [4,7,8]. The function of the MMP family has been implicated in various aspects of IPF biology, including abnormal cell proliferation, epithelial-mesenchymal transition and interaction, and immune system modulation [4]. However, to date, not enough studies have been conducted to comprehensively investigate MMPs expression patterns and their association with clinical features in IPF.

The transcriptome is the collection of all transcribed gene sequences by a particular tissue or cell at a certain developmental stage or functional state. Transcriptome studies can study gene function and gene structure at the overall level and reveal the

molecular mechanism of specific biological processes and disease occurrence [9]. It is a powerful tool for in-depth study of transcriptome complexity and has been widely used in the mechanism study of IPF [10]. RNA-seq can be used to construct gene regulation networks and analyze MMPs or MMPs-related pathways conveniently. It is also extremely suitable for large-scaled cohort studies. RNA-seq and its related methods will promote the development of new personalized therapies for lung diseases [11].

Herein we pooled the current mainstream IPF datasets from the GEO database to form a multicenter large-scaled cohort study to profile the expression of MMPs, and investigated robust changes of MMPs specific in IPF that were not subject to experimental error. The correlation between MMPs and clinical features, association of genotypes with phenotypes and mechanism underlying were determined. In addition, MMP19+ macrophages play an important regulation role in the fibrotic microenvironment through cell-cell interactions by combining the single cell transcriptome to observe the cell-specific expression of MMPs.

## 2. Materials and Methods

### 2.1. Study Design

All data were obtained from GEO datasets GSE32537 including 50 healthy donors and 119 IPF patients [12], GSE47460 including 108 healthy donors and 254 IPF patients [13], GSE150910 including 82 CHP patients, 103 IPF patients and 103 healthy donors [14], GSE124685 including 35 healthy donors and 49 IPF patients [15], GSE134692 including 8 ALI patients, 46 IPF patients and 26 healthy donors [16], GSE135893 including 20 PF and 10 healthy donors [17] and GSE70867 including 62 IPF patients from Freiburg, 50 IPF patients from Siena and 64 IPF patients from Leuven [18] ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)). The lung tissue samples used for bulk RNA sequencing contained 296 normal and 431 IPF, respectively from GEO datasets (GSE32537, GSE47460, GSE150910, GSE124695). Lung tissue used for single-cell sequencing was mainly derived from dataset GSE135893. The sample of alveolar lavage fluid for IPF was obtained from the dataset GSE70867.

### 2.2. Bulk RNA Sequencing Analysis

The differentially expressed genes (DEGs) were identified by the “limma” package using the normalized counts (FPKM data) [19]. The R packages of “heatmap” and “ggplot2” were utilized to generate heat maps, violin plots, and box plots.

### 2.3. Microarray Data Analysis

The microarray data from GSE32537 and GSE47460 were subjected to normalization by a cyclic loess approach, with the probe exhibiting the highest average signal per gene being selected. The significance of differential gene expression was assessed using the Wilcoxon rank sum test, with a p-value threshold of less than 0.05 and an absolute value of log<sub>2</sub> (fold change [FC]) greater than 1. Principal component analysis (PCA) demonstrated the clustering of disease and control samples.

### 2.4. Construction of Prognostic Signatures and Survival Analysis

The GSE70876 dataset with survival data were used to assess the clinical relevance of bronchoalveolar lavage fluid samples from IPF patients. A univariate Cox regression analysis was performed to investigate the association between the levels of matrix metalloproteinase (MMP) and the survival of IPF patients. Significant MMPs were subsequently incorporated into a multivariate Cox regression analysis. Additionally, LASSO and ridge regression techniques were employed to refine the model parameters [20,21]. Subsequently, patients were categorized into high-risk and low-risk groups by the optimal risk score cut-off value. The overall survival (OS) was determined using the Kaplan-Meier method and log-rank test. The sensitivity vs specificity was plotted by the Receiver operating characteristic (ROC) curves.

### 2.5. scRNA-seq Data Processing

The Seurat R package [22] was utilized for quality control, analysis, and examination of single-cell RNA-seq data. Initially, cells with unique feature counts exceeding 5000, unique feature counts below 300, or greater than 5% mitochondrial counts were filtered out. We normalized and stabilized the variance of molecular count data derived from scRNA-seq experiments using “sctransform” [23]. The UMAP visualization and Louvain clustering and subclustering were performed by “RunUMAP” and “FindClusters” in the Seurat. Cluster markers were identified by the “FindAllMarkers”, and cells with a high hemoglobin ratio were removed.

### 2.6. Cell Communication Analysis

CellChat, in conjunction with the CellChatDB database, was employed to investigate cell-cell communication networks through ligand-receptor interactions [24]. Additionally, we projected gene expression data onto a protein-protein interaction (PPI) network to refine gene expression values. We utilized a statistically robust mean method called trimean, which generated fewer interactions compared to other methods.

## 2.7. Gene Expression Regulation Network Analysis

The R package bigScale2 [19] was used to construct the GRNs. Cells from mouse lung tissues were treated with Saline, BLM, and BLM+T3 respectively, and all cell count matrices were generated using Seurat's GetAssayData. The cell matrices were then filtered to remove genes with ensemble identifiers and passed to bigScale2, where networks were constructed under the default parameter with a cutoff of 0.8 for the correlation coefficient. Networks were visualized using software Cytoscopes [20] (version 3.8.9), and each network's layout parameter is "Edge-weighted Spring Embedded Layout".

## 3. Result

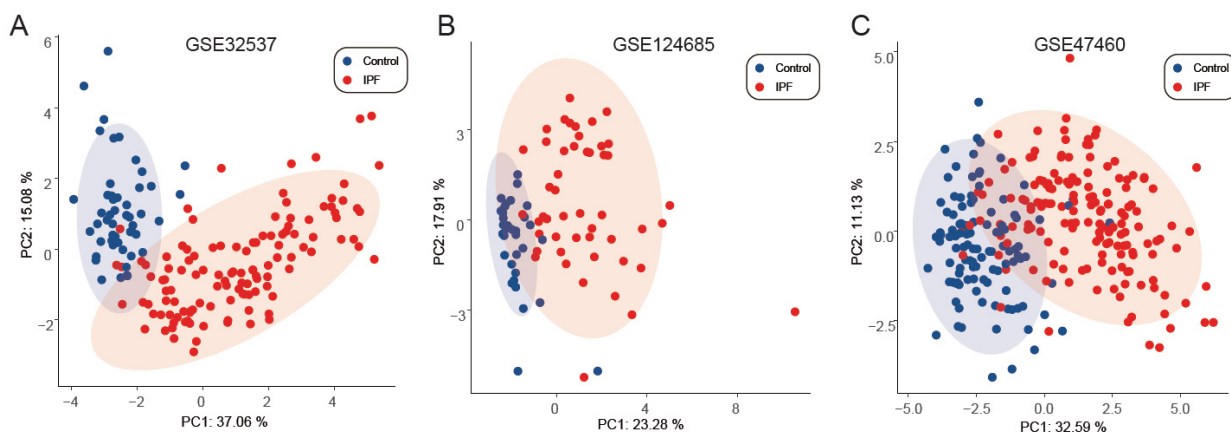
### 3.1. MMPs Expression Significantly Altered in IPF

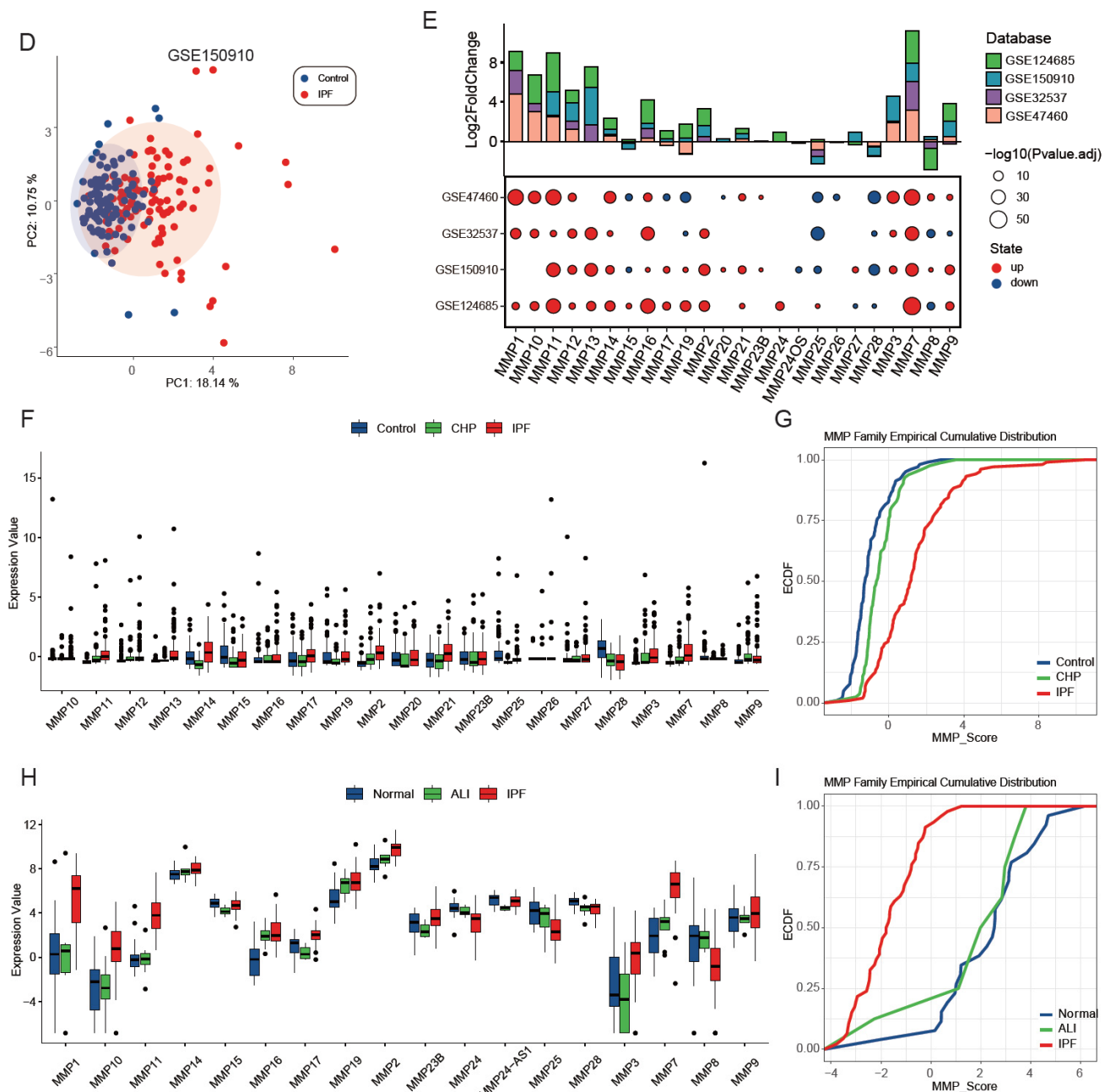
This study included four independent gene expression data (GSE32537, GSE47460, GSE150910, and GSE124695) from GEO database, including two sets of high-throughput sequencing data and two sets of microarray data, a total of 296 normal subjects and 431 IPF patients (Table S1). There were significant differences in MMPs expression between the healthy and IPF, while the MMPs expression profiles from the four datasets were sufficient to clearly distinguish between the normal and IPF in the PCA (Figure 1A–D). The majority of MMPs were obviously upregulated in IPF and the alterations of MMPs were much consistent across the microarray and high-throughput sequencing technologies. MMP1, MMP10, MMP11, MMP12, MMP13, MMP14, MMP16, MMP2, MMP21, MMP3 and MMP7 were consistently up-regulated in the four sets of data, while MMP28 was consistently down-regulated in the four sets of data. Some others, such as MMP15, MMP17, MMP19, MMP25, MMP8 and MMP9, were uncertain (Figure 1E).

To further confirm whether the changes in the MMPs are an important feature that distinguishes IPF disease from other interstitial lung diseases, we introduced gene expression data from chronic hypersensitivity pneumonia (CHP) and acute lung injury (ALI). Even though persistent chronic hypersensitivity pneumonia and repetitive acute lung injury are the main initiative event of IPF, surprisingly, the expression of MMPs didn't show significant change in CHP (Figure 1F) or ALI (Figure 1H) patients. Further, the empirical cumulative distribution curve analysis determined that the abnormal expression of MMPs is specifically present in IPF (Figure 1G,I). This suggests that the MMPs plays a key role in the pathogenesis of IPF.

### 3.2. MMPs Are Significantly Correlated with Clinical Features of IPF

To study the correlation between the expression of MMPs and the clinical features, further analysis was performed on lung function (diffusing capacity of the lungs for carbon monoxide, DLCO, forced vital capacity, FVC, etc.), susceptibility and alveolar surface density. Principal component analysis was performed to derive a single measure of meta lung function representing a combination of multiple pulmonary function parameters (Figure S1A). Next, to consider the potential confounding demographic variables, we performed linear multiple regression analysis to assess the correlated gene expression levels with meta lung function variables accounting for age and sex. Most of MMPs (MMP1, MMP7, MMP28, MMP11, MMP10, MMP14, MMP3, MMP12, MMP9, MMP15, MMP16) are closely related with meta lung function (Figure 2A). 13 of MMPs (MMP8, MMP9, MMP12, MMP16, MMP14, MMP3, MMP11, MMP10, MMP7, MMP1, MMP28, MMP15, MMP25) show a dramatically association with lung function and can be the potential diagnostic markers. MMP19, MMP26, MMP17 and MMP21 have no relationship with lung function (Figure 2B). MMP1 and MMP7 have a significant positive correlation with meta lung function which are currently the most promising diagnostic markers (Figure 2C).

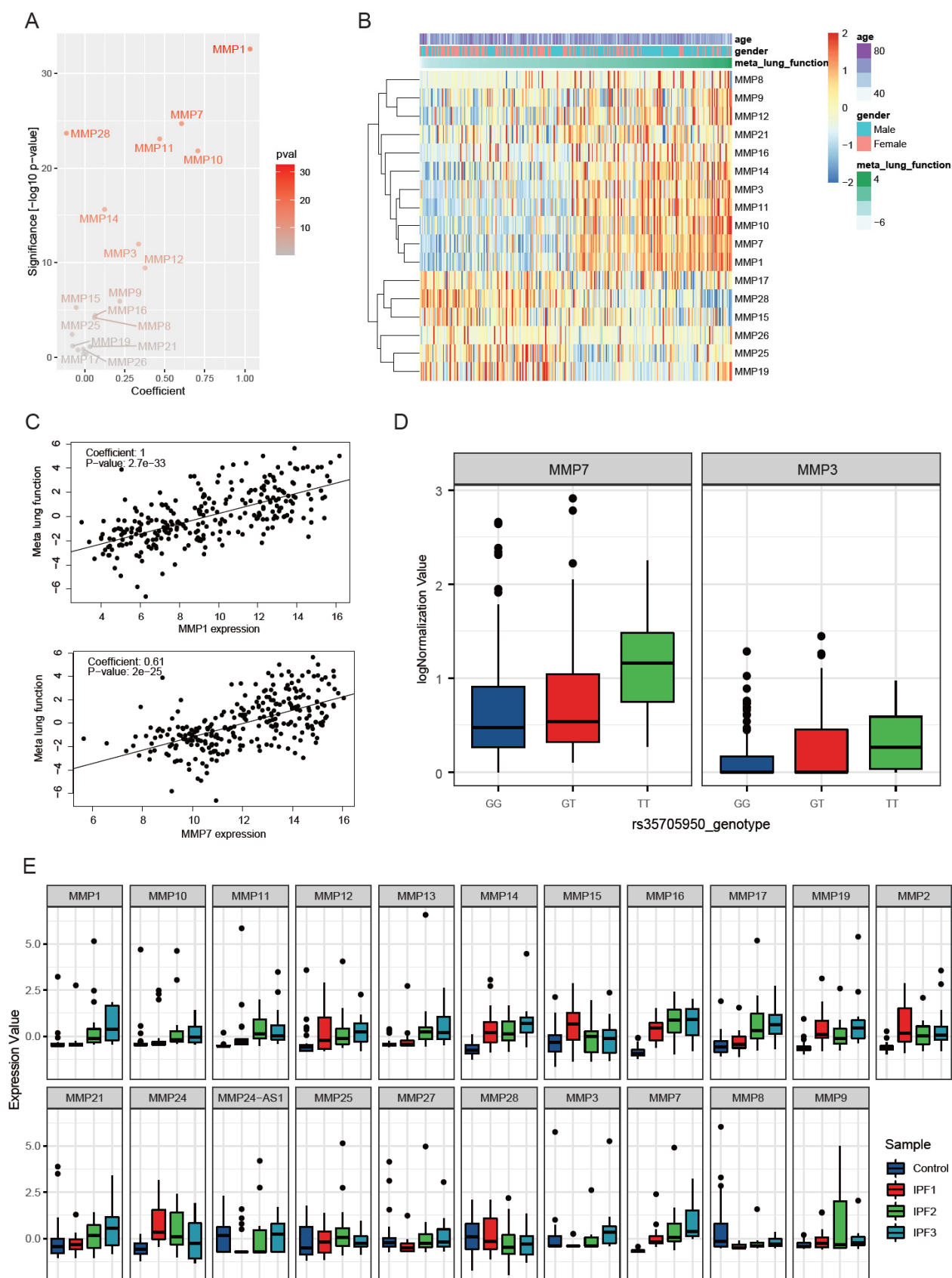




**Figure 1.** MMPs have unique abnormal changes in IPF. (A–D) The PCA plot shows gene expression profiles of MMPs which distinguish IPF patients from the healthy in dataset GSE32537 (A), GSE124685 (B), GSE47460 (C), GSE150910 (D) respectively. (E) The bar plot shows the Log2FoldChange in MMPs in each dataset (top), and the bubble plot shows the significance of the difference in MMPs in each dataset (bottom). (F) The boxplot shows the FPKM values of the MMPs among normal, CHP and IPF state. (G) Empirical cumulative density plot depicts the distribution of MMPs among normal, CHP and IPF state. (H) The boxplot shows the FPKM values of the MMPs among normal, ALI and IPF state. (I) Empirical cumulative density plot depicts the distribution of MMP family among normal, ALI and IPF state.

Currently homozygous (TT) for the minor allele of the MUC5B polymorphism (rs35705950) possesses higher susceptibility and quicker progress of the disease than heterozygous (GT) [6]. The association between MUC5B mutation and the expression of MMPs by expression quantitative trait locus (eQTL) analysis in all samples demonstrated MMP3 and MMP7 expression were closely related to MUC5B rs35705950 (Figure 2D), indicating that the susceptibility of pulmonary fibrosis may also be mediated by the MMPs. In other words, the MUC5B polymorphism (rs35705950) homozygous (TT) is susceptible to fibrosis due to its ability to up-regulate the expression of MMPs (Mainly MMP7). Besides, MMP2 was strongly associated with the MUC5B polymorphism homozygous (TT) as well in IPF patients.

The increase of alveolar surface density is often considered as the gold standard of pulmonary fibrosis. Further analysis of the relationship between MMPs and cellular lung density exhibited that most persistent upregulated MMPs in IPF were accompanied by the increase of alveolar surface density, and MMP1, MMP10, MMP11, MMP12, MMP13, MMP14, MMP16, MMP21, MMP3 and MMP7 increased significantly. There are also genes whose expression decreases as alveolar surface density increases, such as MMP24, and some MMPs were not associated with alveolar surface density, for example MMP25, MMP28, etc. (Figure 2E).



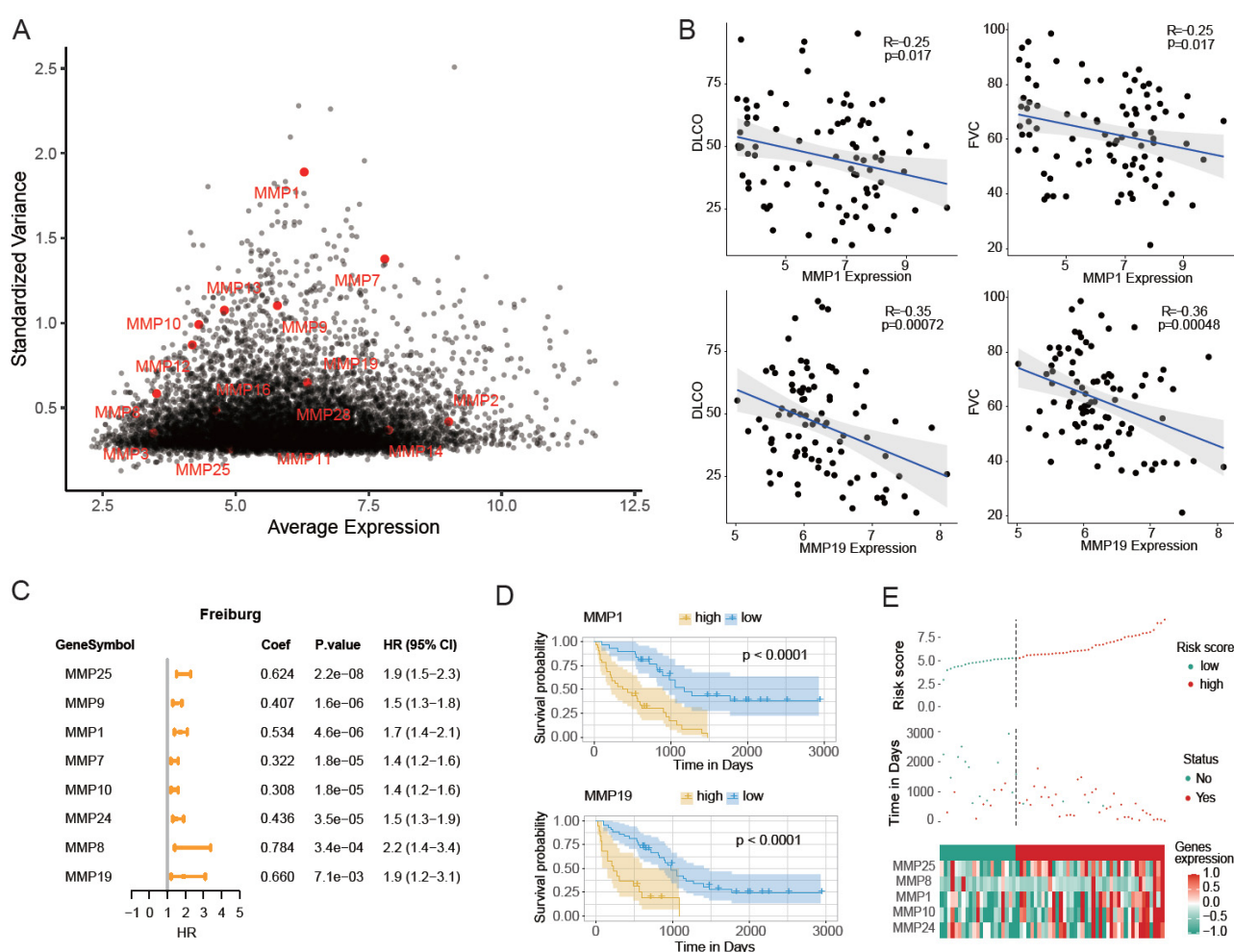
**Figure 2.** MMPs were associated with clinical features. (A) The scatter plot shows the multivariate regression coefficients (X-axis) and the  $-\log_{10}P$ -value (Y-axis) for MMPs with meta lung function. (B) The heatmap illustrates the expression of MMPs and distribution of age, gender, meta lung function in cohort. (C) The scatter plots show the positive correlation of MMP1 and MMP7 with meta lung function. (D) The boxplots show the expression of MMP3 and MMP7 among different MUC5B genotypes. (E) The boxplots show the correlation between MMPs and lung surface density.

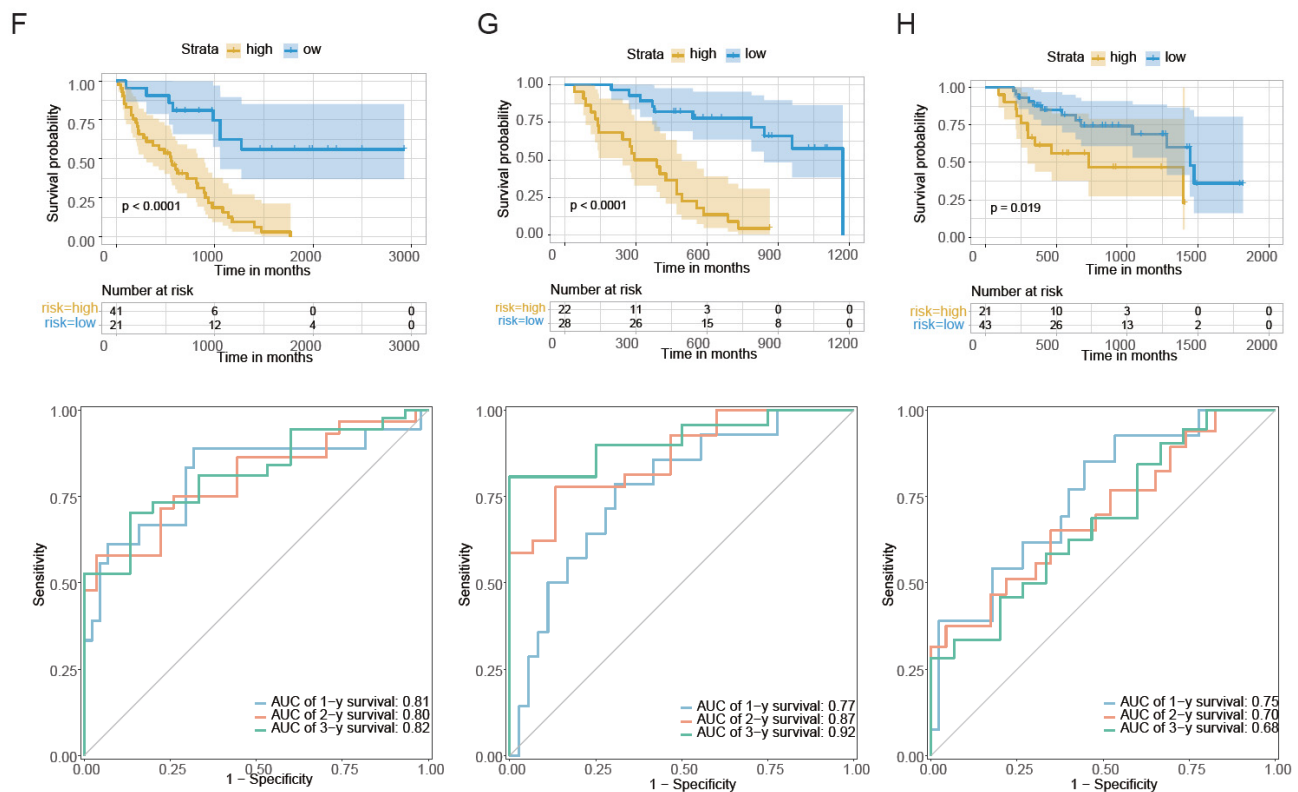


### 3.3. MMPs Predicted the Outcome of IPF

To explore whether the alteration of MMPs family in IPF could explain the prognosis of IPF, the erratic gene expression in MMPs within IPF was determined by standard deviation and average expression value (Figure 3A). The correlation between the expression of MMPs and clinical features only in IPF patients was performed. Among them, MMP19 had an incredible negative correlation with the DLCO and FVC only in IPF patients in comparison to the whole group (Figure 2B), suggesting that MMP19 expression was extremely critical to pulmonary fibrosis (Figure 3B).

The heterogeneity of MMPs expressions predicted the survival and distinguish high-risk from low-risk of IPF patients. BALF microarray data including a total of 176 IPF patients from three regions were performed. Univariate COX regression analysis showed that several MMPs (MMP25, MMP9, MMP1, MMP7, MMP10, MMP24, MMP8, MMP19) were identified as important risk prognostic factors, and most of them were poor prognostic factors (HR > 1) (Figure 3C). Survival analysis further confirmed that MMP1 and MMP19 were prognostic indicators (Figure 3D). A gene set was optimized to construct a multivariate COX regression model including MMP1, MMP10, MMP8, MMP24 and MMP25 (Figure S1C). The risk score obtained by this model can distinguish the survival status of patients (Figure 3E). Time ROC analysis found that the AUC value was 0.81/0.80/0.82 at 1/2/3 years in Freiburg cohort (Figure 3F), 0.77/0.87/0.92 at 1/2/3 years in SIENA cohort (Figure 3G) and 0.75/0.70/0.68 at 1/2/3 years in LEUVEN cohort (Figure 3H) respectively.

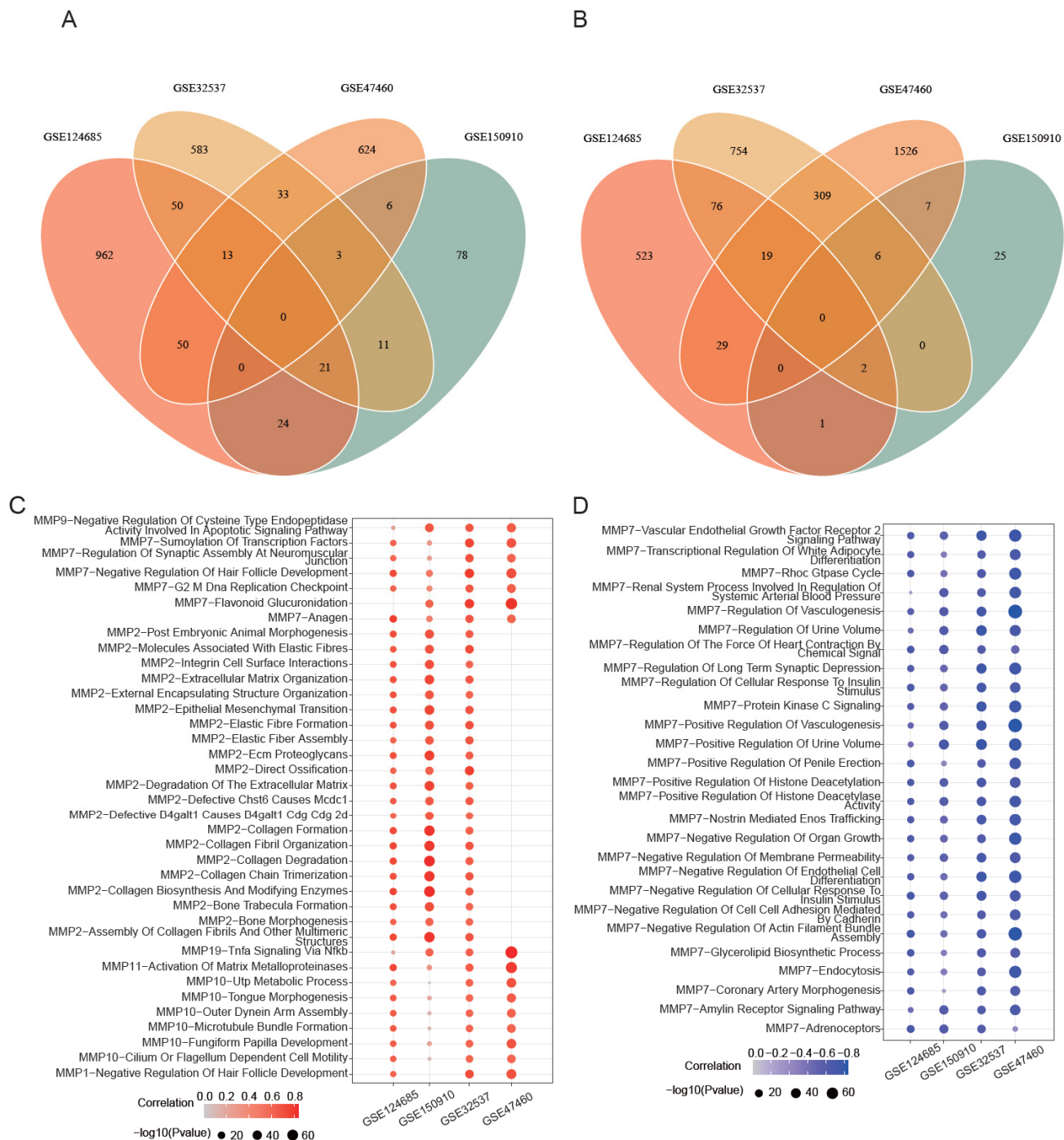




**Figure 3.** MMPs indicated the progressing of IPF pathogenesis. (A) The scatter plot shows the average expression (X-axis) and the standardized variance (Y-axis) for MMPs in IPF. (B) The scatter plot shows the expression (X-axis) and DLCO or FVC (Y-axis) for MMP1 and MMP19 in IPF. (C) The forest plot displays the univariate Cox regression analysis results of the risk score and clinical factors with OS. (D) Kaplan–Meier OS curve for high- and low-risk groups differentiated according to MMP1 or MMP19 expression in Freiburg cohort. The assessment of difference was achieved by log-rank test. (E) The association of high-risk score with MMPs and prognosis of IPF patients. The distributions of risk score (top), OS status (median) and MMPs expression (bottom) in the IPF. (F–H) Kaplan–Meier OS curve for high- and low-risk groups (top). MMPs-related risk prognostic model in Freiburg (F), SIENA (G), LEUVEN (H). Receiver operating characteristic curves (ROC curves) for the 1-, 2-, and 3-year OS periods (bottom). MMPs-related risk prognostic model in Freiburg (F), SIENA (G), LEUVEN (H).

### 3.4. MMPs Mediate a Variety of Biological Functional Pathways in IPF

MMPs play a role in many biological processes in tissue injury and repair, such as tissue remodeling and regeneration, wound repair, tissue defense, and immune responses. To further understand the biological pathways which MMPs were involved in, gene set variation analysis (GSVA) was performed to analyse all MMPs based on four datasets (GSE32537, GSE47460, GSE150910, and GSE124695). We set the Pearson correlation coefficient greater than 0.6 and the adjusted P value less than 0.05 in at least three datasets. A total of 37 positively correlated pathways (Figure 4A) and 27 negatively correlated pathways (Figure 4B) were obtained (Table S2). Most of MMPs were involved in the production of extracellular matrix, which was the most fundamental role of MMP family. MMP10 was closely related to cytoskeleton formation and cell migration, and MMP11 was involved in the activation of other matrix metalloproteinases (Figure 4C). MMP2 reflects the classic functions of matrix metalloproteins, and MMP7-related pathways involve a wide biological process such as DNA replication, SUMOylation, apoptosis and cell-cell adhesion etc. (Figure 4D). Most of enriched signal pathways showed that MMP19 is involved in the TNF signaling pathways (Figure 4C, Table S2). 23 positively correlated signaling pathways was identified by correlation analysis between MMPs and signaling pathways in IPF patients (Figure S1D). Notably, MMP7 is no longer associated with biological processes including DNA replication, SUMOylation, apoptosis and cell-cell adhesion, etc. in IPF samples (Figure S1D,E), which are significantly associated in both normal samples and IPF samples (Figure 4A,B). The significance of MMP7 tends to be lower in pathogenesis of IPF, which may be the reason why MMP7 is not finally applied to biomarker in clinical practice by ATS/ERS.



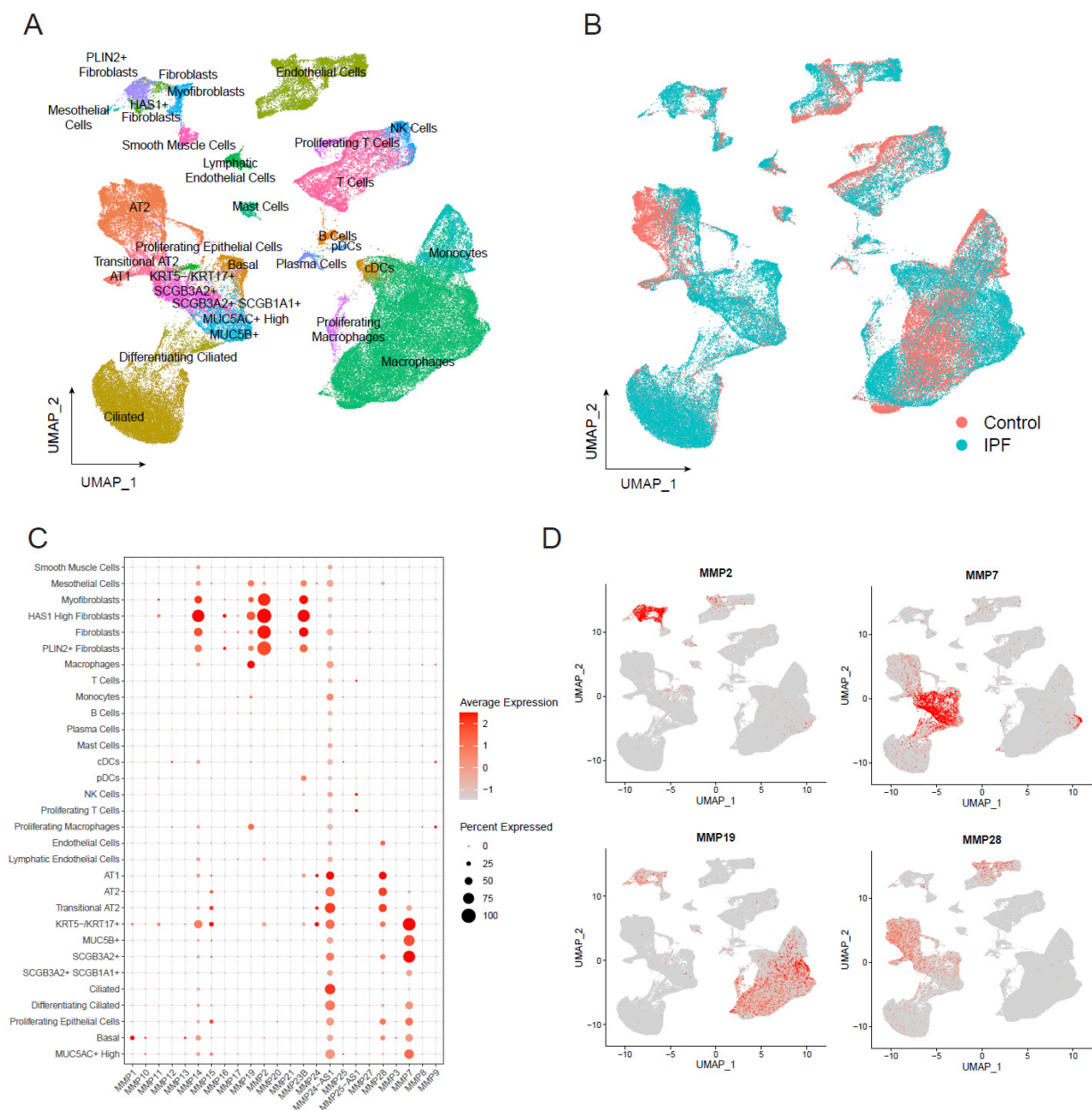
**Figure 4.** Involvement of MMPs signal pathways enrichment analysis. **(A)** The Venn plot shows the overlap of biological pathways positively correlated with MMPs in GSE32537, GSE124685, GSE47460, GSE150910 respectively. **(B)** The Venn plot shows the overlap of biological pathways negatively correlated with MMPs in GSE32537, GSE124685, GSE47460, GSE150910 respectively. **(C)** The bubble plot shows biological pathways positively correlated with MMPs in GSE32537, GSE124685, GSE47460, GSE150910 respectively. **(D)** The bubble plot shows biological pathways negatively correlated with MMPs in GSE32537, GSE124685, GSE47460, GSE150910 respectively.

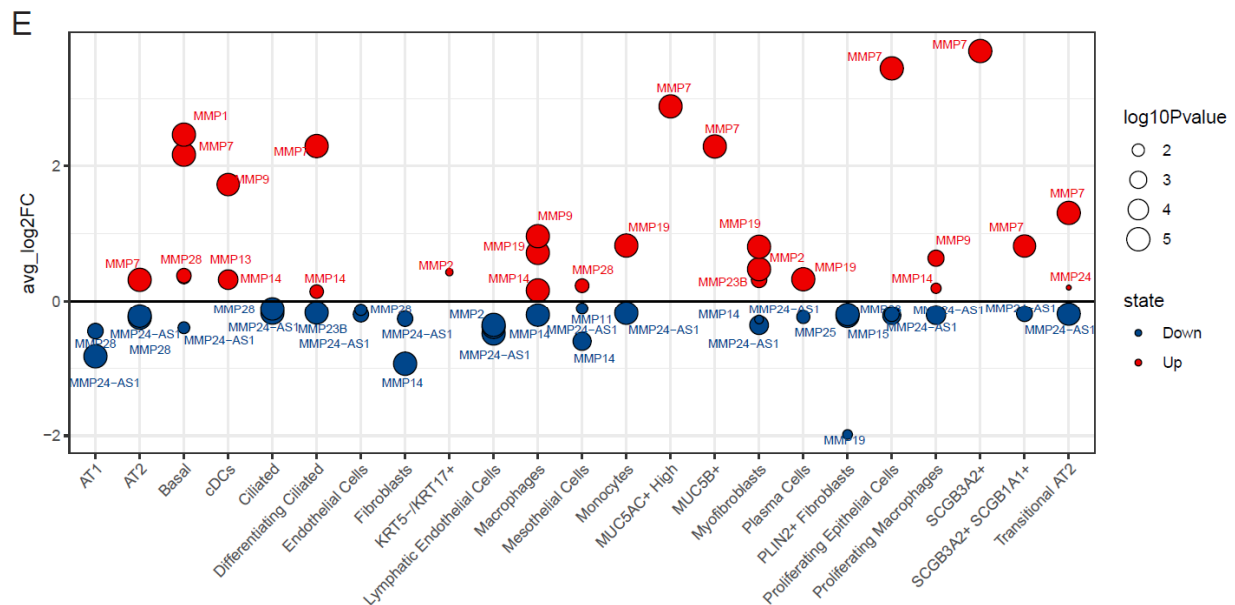
### 3.5. The Roles of MMPs at the Single Cell Resolution

Many previous studies have demonstrated the important role of the MMPs, but most of them have been limited to the tissue sample and therefore may not explain the corresponding cell type specificity. In this study, we obtained 31 major lung cell populations from GSE135893 (Figure 5A) and screened only the single cell sequencing of normal and IPF patients (Figure 5B). MMP2 tends to be highly expressed in mesenchymal cells (Figure 5C). As shown in Figure 5D, MMP7 is specifically expressed in basal cells and secretory cells, and MMP19 highly expressed in macrophages, in line with our previous prediction based on TNF signaling pathway, and MMP28 expressed in epithelial cells and endothelial cells, which may directly affect the formation of epithelial barrier and the remodeling of alveolar structure. Subsequently, we focused on the abnormal expression of MMPs in the disease state. MMP7 has been recognized to be upregulated in the lung tissue of IPF patient. The profiling of MMPs in single cell



demonstrated that MMP7 was indeed highly expressed in most cell types and MMP24 was much lower in all cell types of IPF lung tissues (Figure 5E), suggesting that more attention should be paid to the role of MMPs antisense RNA in the pathogenesis of IPF. The single cell sequencing data provide a rational interpretation of the unstable expression of MMP19 at the lung tissues, and MMP19 in myofibroblasts and macrophages of IPF patients is obviously upregulated, but the expression of MMP19 in PLIN2+ fibroblasts is downregulated.

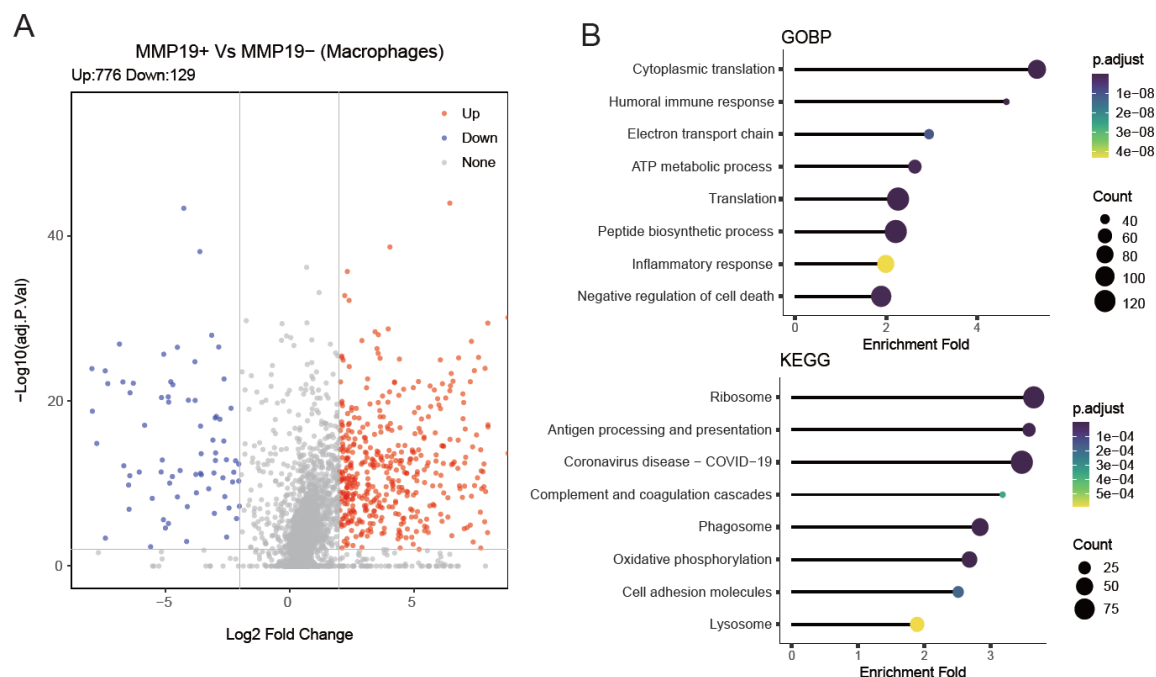


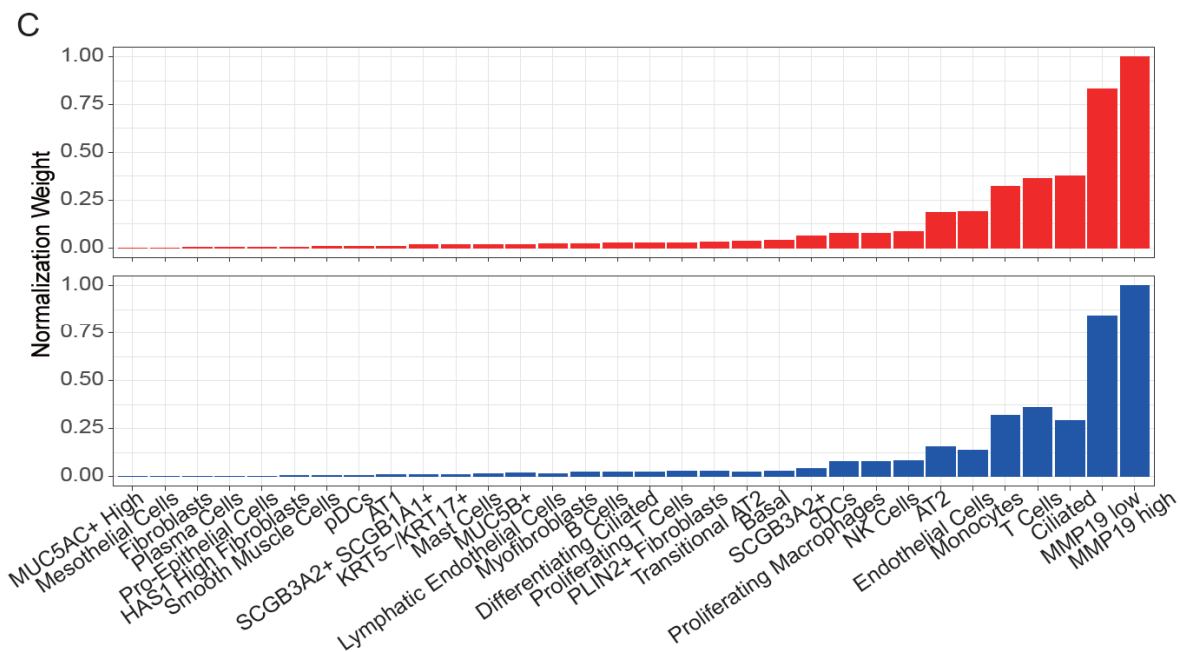


**Figure 5.** MMPs expression at the single-cell resolution. (A) Uniform manifold approximation and projection (UMAP) plot shows cell types in scRNA-seq. Each point represents a single cell, colored according to cell types. (B) Uniform manifold approximation and projection (UMAP) plot shows cell distribution grouped by diagnosis. (C) The bubble plot shows the expression of MMPs in each cell type. (D) UMAP plot shows the cell preference of MMPs expression. (E) The bubble plot shows changes in MMPs expression after disease in each cell type.

### 3.6. Immunomodulatory Effects of MMP19+ Macrophages

The upregulated expression of MMP19 in macrophages prompted us to further explore its role in the immunomodulatory. So far, no studies have systematically elucidated the immunomodulatory function of MMP family in IPF. MMP9 and MMP19 in immune cells are upregulated in IPF, and hence it may be the main form of immunomodulatory function of MMPs. We found that MMP9 was specifically expressed in some macrophages, which is consistent with previous studies, showing that MMP9 secreted by macrophages mediates epithelial mesenchymal transition [21], but the immunomodulatory role of MMP19 has not been reported. We firstly divided the macrophages into subgroups according to the alteration of MMP19. We have MMP19+ macrophages and MMP19− macrophage group gene expression spectrum difference analysis to get 76 raised genes ( $|\text{Log2Fold Change}| > 0.5$  and P value.  $\text{Adj} < 0.01$ ). Functional enrichment analysis revealed 17 primary pathways. The differences are mainly reflected in Chylomicron remodeling, long-chain fatty acid transport, phospholipase inhibitor activity and chemokine activity (Figure 6A). To further explore how transcription factors were involved in these ways, we integrated the four-transcription database (JASPAR, CHEA, ENCODE, Motif Map) and found that MMP19 was regulated by the 79-transcription factor, at the same time these transcription factors regulating the differentially expressed genes (Figure 6B). Eight transcription factors with significant regulatory effects were selected and all were found to be highly expressed in MMP19+ macrophages (Figure 6C).





**Figure 6.** MMP19<sup>+</sup> macrophage-mediated immune function. (A) The volcano plot shows the log2FoldChange (X-axis) between MMP19<sup>+</sup> Macrophages and MMP19<sup>-</sup> Macrophages and the -log10Pvalue (Y-axis). (B) The lollipop plots show pathway enriched by differentially expressed genes between MMP19<sup>+</sup> Macrophages and MMP19<sup>-</sup> Macrophages. (C) The bar plot shows the intensity of communication between MMP19<sup>+</sup> Macrophages or MMP19<sup>-</sup> Macrophages and other cells.

#### 4. Discussion

The expression and function of several matrix metalloproteinases have been explored within the context of IPF; but the scope of these papers is generally limited to the expression of a single or few matrix metalloproteinases [7] as compared with our research where the high-throughput sequencing technology for cDNA sequencing was used, almost all the transcripts of organs or tissues in a certain state can be obtained comprehensively and rapidly [9], excluding the possible experimental errors in QPCR, and the stable and comprehensive MMP gene expression profile can also be obtained. In this study, we achieved global observation of MMPs and obtained a robust and reliable MMPs gene expression profile. We also implemented the association analysis of MMPs with clinical features and biological pathways. At the same time, the specific expression and mechanism of MMPs was innovatively considered at the single cell resolution, which provided theoretical support for MMPs to clinical application.

We first discussed the possibility of MMPs as diagnostic or prognostic markers. MMP7 is one of the most studied MMPs, which is considered to be the most promising as a diagnostic marker [22]. In our study, the expression of MMP7 was found to have similar results with previous studies, showing highly significant differential expression between the normal and IPF. Moreover, MMP7 is strongly correlated with lung function and susceptibility and alveolar surface density. Unfortunately, the expression of MMP7 was irregular in IPF samples, and has no significant correlation with lung function (DLCO and FVC), suggesting that MMP7 is effective in distinguishing normal subjects from patients but not an indicator of disease severity in patients with IPF. MMP19 has a very different performance from MMP7. MMP19 shows a weak correlation with lung function in healthy samples and IPF samples, but it becomes extremely good in IPF samples and suggests that MMP19 may play an important role in the progression of IPF [23]. Using MMP expression profiles in BAL, a novel approach to prognostic classification of IPF was explored, hoping to better understand the different stages of disease progression in a less invasive manner. We further investigated the abnormal expression of MMP in BAL of IPF and determined that the expression of eight MMP genes could define the prognosis of the disease. Multigene prognostic models have previously been developed with machine learning methods but are not recommended as a major ATS recommendation because they do not fully explain model efficacy from the genes themselves [24]. The MMP prognostic model we obtained showed excellent results in terms of model interpretability and accuracy.

We think MMPs are key regulators of IPF disease occurrence and progression. MMPs were originally thought to modulate ECM proteins and subsequently affect fibrosis [2,25]. However, a growing number of studies have shown that MMPs can affect the activity of proteins other than the ECM, including inflammatory mediators, latent growth factors, antifibrotic growth factors, and dividing cell surface molecules and receptors [8]. The function of MMPs has been implicated in many aspects of IPF biology, including abnormal cell proliferation, epithelial mesenchymal transition, and immune system dysregulation. Alveolar surface density is a direct reflection of extracellular matrix accumulation. Our results showed that most MMPs were accompanied by the increase of alveolar surface density, and their expression increased gradually, but there were still some MMPs that were not positively correlated with alveolar surface density, such as MMP15, MMP19, MMP2, MMP24, MMP28, etc. This may also indicate that the regulation of fibrosis by MMPs is not only reflected in extracellular matrix accumulation. Furthermore, we tried to consider

the corresponding biological pathways related to MMPs, to explain the other functions of MMPs. It can be found that MMP7 is the star MMPs gene, and its related biological pathway is very complex, involving cell proliferation, autophagy, and other processes, which not only emphasizes the core and key role of MMP7 in IPF, but also makes it difficult for MMP7 to be used clinically. In the correlation analysis of biological pathways, the performance of MMP2 is also noteworthy. It can be found that most significant biological pathways are related to MMP2, but these pathways basically reflect the basic functions of MMPs, such as the formation and degradation of extracellular matrix. Subsequently, we tried to extract IPF samples separately for MMPs-pathway correlation analysis. We found that MMP2 remained strongly associated with extracellular matrix formation and degradation pathways, but MMP7 lost its correlation with many pathways. Combined with the results of quantitative trait loci analysis, we found that the expression level of MMP7 in IPF susceptible population was often higher than that in non-susceptible population. MUC5B Homozygous (TT) MMP2 is lower in IPF patients, indicating that the basic function of MMPs is blocked, leading to further exacerbation of fibrosis [26]. In addition, we observed a strong correlation between MMP19 and TNF signaling pathway. TNF is mainly produced by activated macrophages, NK cells and T lymphocytes, indicating that MMP19 plays an important role in immune regulation [27].

In previous studies, the need for cellularity-based correction of the RNA-seq dataset has rarely been considered, which in IPF is critical due to hugely variable and often low number of malignant cells present throughout IPF. This seemingly innocuous oversight obscures the effects of genes, like MMP7, which are predominantly expressed by conducting airway epithelial cells, MMP2, which are predominantly expressed by fibroblasts, MMP19, which are predominantly expressed by macrophages. We propose that MMP7 is mainly used to distinguish the normal from IPF, and is related to a variety of pathways reflecting epithelial function such as cell proliferation and division, gene expression modification, etc. MMP2 is mainly used to distinguish the degree of fibrosis, mainly related to the formation and degradation of extracellular matrix [28]. Combining the cell-specific expression profile, we can further infer that abnormal conducting airway epithelial cells are the key node of IPF, and fibroblasts are the main cause of IPF deterioration. This may be why the drugs we have developed to target fibroblasts are not enough to cure IPF, simply by slowing the progression of the disease. The main key to curing IPF may lie in targeting the epithelium, thereby reversing the epithelial lesions. Our previous studies confirmed that MMP19 promotes epithelial cell migration and wound healing. The up-regulation of MMP19 induced by lung injury may play a protective role in the development of fibrosis through the induction of PTGS2. According to the results of the current analysis, MMP19 may also play an important role in immune regulation. The final cell communication analysis found that MMP19+ macrophages can indeed secrete TNF factors to activate downstream receptors and further affect energy metabolism in AT2. MMP19+ macrophages have important implications for the disease development of IPF and should be considered when evaluating outcome expectations in patients. Pirfenidone, as one of the drugs approved by FDA for IPF, is believed to reduce the secretion of fibrotic proteins and cytokines, the biosynthesis and accumulation of extracellular matrix, and the synthesis and secretion of TNF- $\alpha$  in inflammatory cells.

There are still some limitations in our research remaining to be further solved. The clinical information of IPF is incomplete, and the study will analyze the clinical information of multiple cohorts separately, which will inevitably lead to bias. Therefore, the best plan is to use the tissue, serum, alveolar lavage fluid, clinical information, and other data of the same cohort.

## 5. Conclusions

In conclusion, our analysis shows that MMPs has great application potential in the diagnosis, treatment, and prognosis of IPF. We believe that MMP7 can be used as a diagnostic marker and MMP19 is an excellent prognostic marker. In addition, MMPs also play an important role in the occurrence and progression of IPF. The dysregulation of MMP7-related signaling pathway in airway epithelial cells may be the core cause of IPF, and the dysregulation of MMP2-related signaling pathway in fibroblasts may be the main cause of IPF deterioration. In addition, MMP19 plays an important role in immune regulation mediated by the TNF superfamily.

## Supplementary Materials

Summarize the supplementary information with the caption names in this section. The following supporting information can be found at: <https://www.sciopublish.com/index/journals/article/fibrosis/32/id/43>.

## Author Contributions

All authors contributed to the study conception and design. Data retrieval and integration were performed by Z.L. (Zhongzheng Li), H.Z., S.W. Bulk RNA sequencing data were analysis by Z.L. (Zhongzheng Li), P.Y., Z.L. (Zhenwei Li) and Y.L. single-cell RNA sequencing data were analysis by Z.L. (Zhongzheng Li), H.Y., J.W. and Z.Z. The first draft of the manuscript was written by L.W., J.Y., M.Z., R.W. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Ethics Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

The bulk RNA sequencing data and single-cell RNA sequencing data used in this study have been deposited in Gene Expression Omnibus (GSE32537, GSE47460, GSE150910, GSE124685, GSE135893 and GSE70867) that are publicly accessible at [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo).

## Funding

The work was supported by a Startup package (GY), Henan Normal University, the 111 Project “State Innovation Base for Pulmonary Fibrosis”, the Ministry of Science and Technology, PR China, 2019YFE0119500, and Henan Province Science and Technology Project, 212102310894.

## Declaration of Competing Interest

The authors declare no conflict of interest.

## References

- Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet* **2017**, 389, 1941–1952.
- Barratt SL, Creamer A, Hayton C, Chaudhuri N. Idiopathic Pulmonary Fibrosis (IPF): An Overview. *J. Clin. Med.* **2018**, 7, 201.
- Sgalla G, Iovene B, Calvello M, Ori M, Varone F, Richeldi L. Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir. Res.* **2018**, 19, 32.
- Craig VJ, Zhang L, Hagood JS, Owen CA. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **2015**, 53, 585–600.
- Dancer RC, Wood AM, Thickett DR. Metalloproteinases in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **2011**, 38, 1461–1467.
- Cui N, Hu M, Khalil RA. Biochemical and Biological Attributes of Matrix Metalloproteinases. *Prog. Mol. Biol. Transl. Sci.* **2017**, 147, 1–73.
- Yue L, Shi Y, Su X, Ouyang L, Wang G, Ye T. Matrix metalloproteinases inhibitors in idiopathic pulmonary fibrosis: Medicinal chemistry perspectives. *Eur. J. Med. Chem.* **2021**, 224, 113714.
- Mahalanobish S, Saha S, Dutta S, Sil PC. Matrix metalloproteinase: An upcoming therapeutic approach for idiopathic pulmonary fibrosis. *Pharmacol. Res.* **2020**, 152, 104591.
- Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S, et al. RNA sequencing: new technologies and applications in cancer research. *J. Hematol. Oncol.* **2020**, 13, 166.
- Konigsberg IR, Borie R, Walts AD, Cardwell J, Rojas M, Metzger F, et al. Molecular Signatures of Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* **2021**, 65, 430–441.
- Yin Q, Strong MJ, Zhuang Y, Flemington EK, Kaminski N, De Andrade JA, et al. Assessment of viral RNA in idiopathic pulmonary fibrosis using RNA-seq. *BMC Pulm. Med.* **2020**, 20, 81.
- Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, et al. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* **2013**, 68, 1114–21.
- Yu G, Tzouveleakis A, Wang R, Herazo-Maya JD, Ibarra GH, Srivastava A, et al. Thyroid hormone inhibits lung fibrosis in mice by improving epithelial mitochondrial function. *Nat. Med.* **2018**, 24, 39–49.
- Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, et al. Chronic Hypersensitivity Pneumonitis, an Interstitial Lung Disease with Distinct Molecular Signatures. *Am. J. Respir. Cell Mol. Biol.* **2020**, 202, 1430–1444.
- McDonough JE, Ahangari F, Li Q, Jain S, Verleden SE, Herazo-Maya J, et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* **2019**, 4, e131597.
- Sivakumar P, Thompson JR, Ammar R, Porteous M, Mccoubrey C, Cantu E 3rd, Ravi K, et al. RNA sequencing of transplant-stage idiopathic pulmonary fibrosis lung reveals unique pathway regulation. *ERJ Open Res.* **2019**, 5, 00117-2019.
- Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **2020**, 6, eaba1972.
- Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. BAL Cell Gene Expression Is Indicative of Outcome and Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* **2019**, 199, 622–630.
- Iacono G, Massoni-Badosa R, Heyn H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* **2019**, 20, 110.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of



- biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504.
21. Peng L, Wen L, Shi QF, Gao F, Huang B, Meng J, et al. Scutellarin ameliorates pulmonary fibrosis through inhibiting NF- $\kappa$ B/NLRP3-mediated epithelial-mesenchymal transition and inflammation. *Cell Death Dis.* **2020**, *11*, 978.
  22. Rosas IO, Richards TJ, Konishi K, Zhang Y, Gibson K, Lokshin AE, et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med.* **2008**, *5*, e93.
  23. Yu G, Kovkarova-Naumovski E, Jara P, Parwani A, Kass D, Ruiz V, et al. Matrix metalloproteinase-19 is a key regulator of lung fibrosis in mice and humans. *Am. J. Respir. Crit. Care Med.* **2012**, *186*, 752–762.
  24. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* **2018**, *198*, e44–e68.
  25. Drakopanagiotakis F, Wujak L, Wygrecka M, Markart P. Biomarkers in idiopathic pulmonary fibrosis. *Matrix Biol.* **2018**, *68–69*, 404–421.
  26. Rozenberg D, Sitzer N, Porter S, Weiss A, Colman R, Reid WD, et al. Idiopathic Pulmonary Fibrosis: A Review of Disease, Pharmacological, and Nonpharmacological Strategies With a Focus on Symptoms, Function, and Health-Related Quality of Life. *J. Pain Symptom Manage* **2020**, *59*, 1362–1378.
  27. Collins MK, Shotland AM, Wade MF, Atif SM, Richards DK, Torres-Llompарт M, et al. A role for TNF- $\alpha$  in alveolar macrophage damage-associated molecular pattern release. *JCI Insight* **2020**, *5*, e134356.
  28. Bormann T, Maus R, Stolper J, Tort Tarrés M, Brandenberger C, Wedekind D, et al. Role of matrix metalloprotease-2 and MMP-9 in experimental lung fibrosis in mice. *Respir Res.* **2022**, *23*, 180.