Identification of key modules and hub genes associated with lung function in idiopathic pulmonary fibrosis

Yuechong Xia, Cheng Lei, Danhui Yang and Hong Luo

Department of Respiratory Medicine, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China
Research Unit of Respiratory Disease, Central South University, Changsha, Hunan, China
Hunan Diagnosis and Treatment Center of Respiratory Disease, Changsha, Hunan, China

ABSTRACT

Background. Idiopathic pulmonary fibrosis (IPF) is a chronic and progressive interstitial lung disease, characterized by a decline in lung function. To date, the pathophysiologic mechanisms associated with lung dysfunction remain unclear, and no effective therapy has been identified to improve lung function.

Methods. In the present study, we used weighted gene co-expression network analysis (WGCNA) to identify key modules and hub genes associated with lung function in IPF. Three datasets, containing clinical information, were downloaded from Gene Expression Omnibus. WGCNA was performed on the GSE32537 dataset. Differentially expressed gene s (DEGs) between IPF patients and healthy controls were also identified to filter hub genes. The relationship between hub genes and lung function was then validated using the GSE47460 and GSE24206 datasets.

Results. The red module, containing 267 genes, was positively correlated with the St. George's Respiratory Questionnaire score ($r = 0.37$, $p < 0.001$) and negatively correlated with the percent predicted forced vital capacity (FVC% predicted) ($r = -0.46$, $p < 0.001$) and the percent predicted diffusion capacity of the lung for carbon monoxide (Dlco% predicted) ($r = -0.42$, $p < 0.001$). Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis suggested that the genes in the red module were primarily involved in inflammation and immune pathways. Based on Module Membership and Gene Significance, 32 candidate hub genes were selected in the red module to construct a protein-protein interaction network . Based on the identified DEGs and the degree of connectivity in the network, we identified three hub genes, including interleukin 6 (*IL6*), suppressor of cytokine signaling-3 (*SOCS3*), and serpin family E member 1 (*SERPINE1*). In the GSE47460 dataset, Spearman correlation coefficients between Dlco% predicted and expression levels of *IL6*, *SERPINE1*, *SOCS3* were –0.32, –0.41, and –0.46, respectively. Spearman correlation coefficients between FVC% predicted and expression levels of *IL6*, *SERPINE1*, *SOCS3* were –0.29, –0.33, and –0.27, respectively. In the GSE24206 dataset, all three hub genes were upregulated in patients with advanced IPF.

Conclusion. We identified three hub genes that negatively correlated with the lung function of IPF patients. Our results provide insights into the pathogenesis underlying the progressive disruption of lung function, and the identified hub genes may serve as biomarkers and potential therapeutictargets for the treatment of IPF patients.

Corresponding author
Hong Luo, luohonghuxi@csu.edu.cn

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic and progressive interstitial lung disease of unknown etiology, characterized by fibrosis or structural deformations, honeycomb lung, plaque pulmonary parenchymal fibrosis, and fibroblast foci (*Raghu et al., 2011*; *Richeldi, Collard & Jones, 2017*). The median survival time after diagnosis is 2–3 years (*Kim, Perlman & Tomic, 2015*; *Martinez et al., 2017*; *Richeldi, Collard & Jones, 2017*). To date, only pirfenidone and nintedanib have been approved by the United States Food and Drug Administration to treat patients with IPF. However, these treatments merely slow the progression of IPF, without improving lung function (*Costabel et al., 2019*; *Costabel et al., 2017*; *Richeldi et al., 2014*; *Richeldi et al., 2020*). Studies have confirmed that age, genetics, environmental factors, maladaptive repair processes, and the immune system are involved in the etiology of IPF (*Martinez et al., 2017*; *Meng et al., 2020*; *Richeldi, Collard & Jones, 2017*). However, the pathophysiologic mechanisms that underly IPF are complex and remain incompletely understood (*Lederer & Martinez, 2018*; *Richeldi, Collard & Jones, 2017*).

Transcriptomics studies of patients with IPF have demonstrated that transcriptional changes are involved in the pathophysiologic mechanisms of these diseases (*Yang et al., 2007*; *Zuo et al., 2002*). Genes that are differentially expressed in different groups are almost always associated with a particular disease phenotype (*Huang et al., 2015*; *Konishi et al., 2009*; *Todd et al., 2019*). Yang and colleagues (*2013*) analyzed the transcriptional profiles of lung tissue, collected from IPF patients and non-diseased controls, and found that the elevated expression of cilium genes was associated with more extensive microscopic honeycombing. *Boon et al. (2009)* studied the lung expression profiles of six patients with relatively stable IPF and six patients with progressive IPF, and found that genes associated with cell proliferation, migration, and cell morphology were highly expressed in the progressive IPF group relative to the stable IPF group.

The Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo) is a public database that provides a large quantity of gene expression datasets. These datasets can be downloaded freely and reused, to reveal the molecular pathogenesis of diseases. In this study, to identify key modules and hub genes associated with lung function in IPF, we downloaded datasets containing information regarding the clinical characteristics of lung function from GEO and performed weighted gene co-expression network analysis (WGCNA) on one dataset. Our study provides insights into the pathogenesis of progressive lung function decline in IPF, and the identified hub genes may represent therapeutic targets for the treatment of IPF patients.
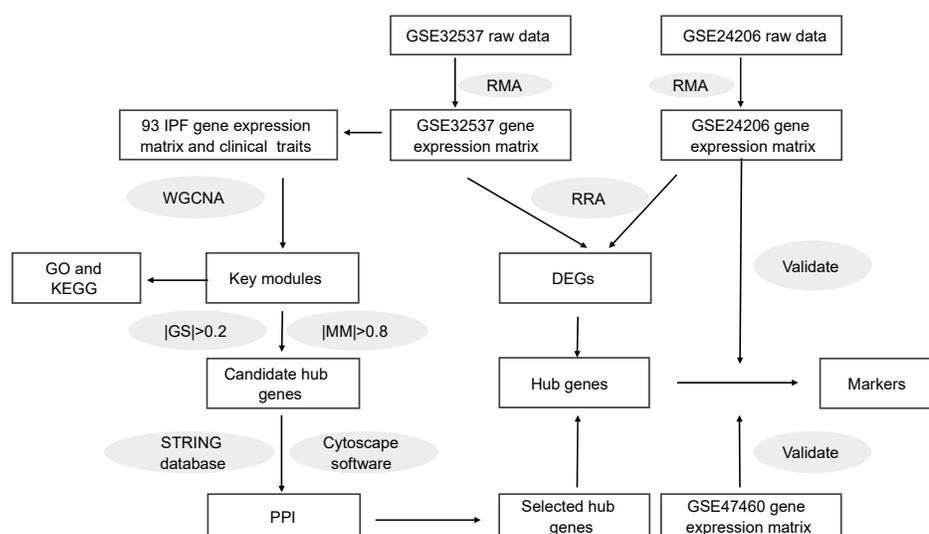
Figure 1 Workflow of this study.

MATERIALS & METHODS

Microarray data

Figure 1 shows the workflow of our study. On the GEO home page, "IPF" was used as the search term. We selected datasets according to the following criteria: (1) the gene expression profile was measured using microarray chip technology; (2) the samples for the study were lung tissues from healthy donors or patients with IPF; (3) the dataset provided raw data or a gene expression matrix; and (4) the dataset contained information regarding the clinical characteristics of lung function. Finally, we selected two datasets, GSE32537 and GSE47460. Although the GSE24206 dataset did not contain clinical characteristics, the patients with IPF were divided into two groups in this dataset. Lung samples that were obtained at the time of biopsy were considered to represent early IPF, whereas samples obtained at the time of explant were considered to represent advanced IPF (*Meltzer et al., 2011*). Therefore, the GSE24206 dataset was also included. Table S1 shows the details of the three datasets.

Data preprocessing

We downloaded the raw data for GSE32537 and GSE24206. The raw data was obtained in CEL format. Data quality control was performed before data analysis. We used the affyPLM package (, http://www.bioconductor.org/packages/release/bioc/html/affyPLM.html, v.1.60.0) to assess the array quality, by calculating relative log expression (RLE) and normalized unscaled error (NUSE). We used the affy package (http://bioconductor.org/packages/release/bioc/html/affy.html, v.1.62.0) to normalize the array data by using the robust multi-array average (RMA) method. Then, we annotated the probes using platform annotation files, and genes that were represented by more than one probe were used to calculate the average gene expression levels. The gene expression matrix file for
Table 1 Demographic data for subjects used in this study.

Characters	GSE32537 (93 IPF)	GSE47460 (86 IPF)
Age (years)	62.81 ± 8.32	63.65 ± 7.89
SGRQ score	46.81 ± 20.90	–
FEV1% predicted	–	68.20 ± 16.88
FVC% predicted	62.76 ± 15.86	61.38 ± 15.35
Dlco% predicted	46.76 ± 20.22	47.28 ± 18.71
Sex (%)		
Male	62(67)	61(71)
Female	31(33)	25(29)
Smoking history (%)		
Non-smoke	32(34)	30(35)
Former	61(66)	56(65)

Notes.
Data are presented as mean ± SD or $n(\%)$.

GSE47460, which had been normalized using a cyclic loess approach in pairwise fashion, was downloaded. We extracted IPF samples with complete data, including lung function [percent predicted forced vital capacity (FVC% predicted) and percent predicted diffusion capacity of the lung for carbon monoxide (Dlco% predicted)], and healthy donor samples from the GSE32537 and GSE47460 datasets. Table 1 shows the clinical characteristics of IPF patients in both datasets.

Construction of co-expression network with WGCNA

Co-expression network analysis was performed using the R package "WGCNA" (https://cran.r-project.org/web/packages/WGCNA/index.html, v.1.69) (*Langfelder & Horvath, 2008*). First, we extracted the gene expression profile data from IPF patients in the GSE32537 dataset and selected the top 25% of variable genes. We constructed scale-free co-expression networks using these genes. Second, we used Pearson correlation matrices to calculate a correlation matrix among these genes. Third, we transformed the correlation matrix into a weighted adjacency matrix, through a power function. To construct scale-free networks, we chose the soft threshold power value, using the following criteria (*Zhang & Horvath, 2005*): (1) the generated Scale free Topology Model Fit $R^2 > 0.90$; (2) the mean connectivity in the network should be as large as possible; and (3) the slope of the linear fitting model is around −1. Finally, we performed automatic network construction and module detection, using the following major parameters: power = 4, networkType = unsigned, maxBlockSize = 4,000, minModuleSize = 30, and mergeCutHeight = 0.25.

Identification of clinically significant modules

Module eigengene represents the first principal component of a given module and the gene expression profiles in this module. When a sample trait is incorporated into the co-expression network, the correlation coefficient between the module eigengene and this sample trait can be calculated. The eigengene significance is defined as the correlation coefficient. Based on the eigengene significance, we were able to identify key modules.

Gene ontology and KEGG enrichment analysis

To further understand the functions of the genes in the key module, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on the genes in the key module, using the R package "clusterProfiler" (http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html,v.12.0) (*Yu et al., 2012*). We selected GO terms including biological process (BP), cellular component (CC), and molecular function (MF). We regarded a *p*-value <0.01 as the cut-off criterion.

Identification of candidate hub genes in key modules

Module membership (MM) represents the intramodular connectivity of any gene in a given module. A higher absolute value of MM indicates that a gene has a higher negative or positive correlation with the module eigengenes (MEs). Gene significance (GS) is used to incorporate external information into the co-expression network. A higher absolute value of GS indicates the increased biological significance of a gene for a given clinical trait. Candidate hub genes in key modules were selected based on |MM|>0.8 and |GS|>0.2.

Identification of DEGs

Differentially expressed genes (DEGs) between IPF lung tissues and healthy lung tissues were analyzed by the R package "limma" (http://www.bioconductor.org/packages/release/bioc/html/limma.html, v.3.42.2). DEGs were defined by |log$_2$ Fold Change|>0.5 and adjusted *p*-value <0.05. During this process, the adjusted *p*-value, which is referred to as the false discovery rate (FDR), was calculated using the Benjamini–Hochberg correction method. Subsequently, the lists obtained from the differential expression analysis of each dataset were integrated, using the R package "RobustRankAggreg (RRA)" (https://cran.r-project.org/web/packages/RobustRankAggreg/index.html, v.1.1) (*Bardou et al., 2014*).

Construction of a protein–protein interaction network

We uploaded candidate hub genes into the STRING database (https://string-db.org, v.11.0) to construct a protein-protein interaction (PPI) network, and we visualized the interaction relationships among the candidate hub genes using Cytoscape software (https://cytoscape.org, v.3.7.0). Finally, we identified hub genes according to the DEGs and the degree of connectivity within the network.

Validation of hub genes

We validated the hub genes using the GSE47460 and GSE24206 datasets. The differential expression of hub genes between healthy lung tissue and IPF lung tissue in the GSE47460 dataset were calculated and visualized using the R package "ggpubr" (https://cran.r-project.org/web/packages/ggpubr/index.html, v.0.2.5). Then, we validated the correlations between lung function and expression levels of hub genes, using the R package "ggstatsplot" (https://cran.r-project.org/web/packages/ggstatsplot/index.html, v.0.3.1). We also validated the differential expression of hub genes between early IPF and advanced IPF using the GSE24206 dataset.
Statistical analysis

Continuous variables were compared between two groups by applying the Student's t-test or a non-parametric Wilcoxon rank-sum test, as appropriate. Associations between the expression levels of genes and lung function were determined by Spearman correlation coefficient. All statistical analyses were performed in R (v.3.6.1), and $p < 0.05$ was regarded as significant.

RESULTS

Weighted gene co-expression network analysis (WGCNA)

We selected the top 25% of variant genes identified in the GSE32537 dataset. A total of 93 IPF samples, containing 4,705 genes, were used for WGCNA. Hierarchical clustering analysis was performed, and when the threshold was set to 60, GSM806234 and GSM806335 were considered to be outliers (Fig. S1). Outlier samples were removed prior to further analyses. When the soft threshold power value was set to 4, the co-expression network exhibited an approximate scale-free topology (Figs. 2A–2C). WGCNA identified 14 modules, containing between 65 and 1,672 genes (Fig. 2D).

The red module had the strongest positive correlation with the St. George's Respiratory Questionnaire (SGRQ) score ($r = 0.37$, $p < 0.001$) and was negatively correlated with FVC% predicted ($r = -0.46$, $p < 0.001$) and Dlco% predicted ($r = -0.42$, $p < 0.001$, Fig. 2E). The red module was identified as the key module. We randomly selected 400 genes, to visualize the relationship among the modules, using a heatmap plot showing topological overlap (Fig. 3A). The heatmap suggests a high degree of independence among the modules and genes at the tip of the module branches have high intramodular connectivity with the rest of the genes in this module. Then, the correlations between MM and GS for FVC% predicted, Dlco% predicted, and SGRQ score in the red module were shown in Figs. 3B–3D. GS showed a significant correlation with the MM in the red module, which indicated that the hub genes identified in the red module tended to be highly associated with lung function.

GO and KEGG enrichment analysis

GO and KEGG pathway enrichment analyses were performed on the genes in the red module. GO enrichment results demonstrated that the red module genes were significantly associated with inflammation and immune responses, such as the response to lipopolysaccharide (LPS), leukocyte differentiation, cell chemotaxis, and the cellular response to molecules of bacterial origin. The KEGG pathway enrichment results indicated that genes in the red module were primarily enriched in the tumor necrosis factor (TNF) signaling pathway, the interleukin (IL)-17 signaling pathway, the Janus kinase (JAK)-signal transducer and activator of transcription (STAT) signaling pathway, and cytokine-cytokine receptor interactions. GO and KEGG terms were ranked in ascending order, based on p-values. Table 2 shows the top 10 categories associated with BP, CC, and MF. Table 3 shows the top 10 categories identified in the KEGG analysis.
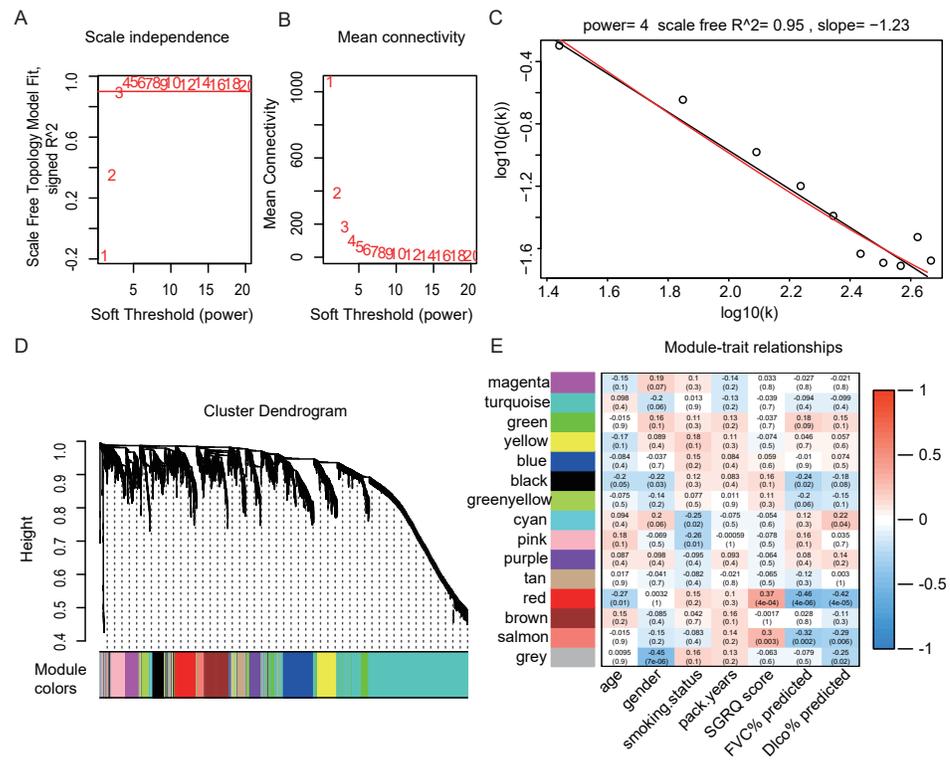
Figure 2 **Construct gene co-expression network and identify key module in 93 IPF patients from GSE32537 dataset.** (A–C) The process of selecting soft threshold. When we set $R^2 > 0.9$, $\beta = 4$ is chosen, log-log plot of network connectivity distribution is almost a straight line, which represents that the network is approximately scale-free topology. (D) Cluster dendrogram and module color. branch represents gene cluster by average linkage hierarchical clustering and each color under cluster represents one co-expression gene module by the Dynamic Tree Cut. (E) Heatmap of correlation between module eigengenes and clinical traits. red color indicates positive correlation and green indicates negative correlation. In each cell, the up number represents correlation coefficients and the bottom number represents P value. The red module has the strongest negative correlation with lung function.

Candidate hub genes in the red module

We selected candidate hub genes in the red module, based on the criterion: |MM| > 0.8 and |GS| > 0.2. Then we examined the intersection of the three gene lists (Fig. 4A). Finally, 32 genes were identified as candidate hub genes. Table 4 shows the GS and MM values for the 32 candidate hub genes in the red module.

Identification of DEGs

Using the thresholds |$\log_2$ FoldChange| >0.5 and adjusted p-values < 0.05, we identified 1,347 upregulated and 1,023 downregulated genes in the GSE32537 dataset and 3,285 upregulated and 596 downregulated genes in the GSE24206 dataset. RRA was performed to integrate the DEGs identified in the GES32537 and GSE24206 datasets. Finally, 247 upregulated and 78 downregulated genes were identified (Table S2). The top 20 upregulated and the top 20 downregulated genes are shown in Fig. 4B.
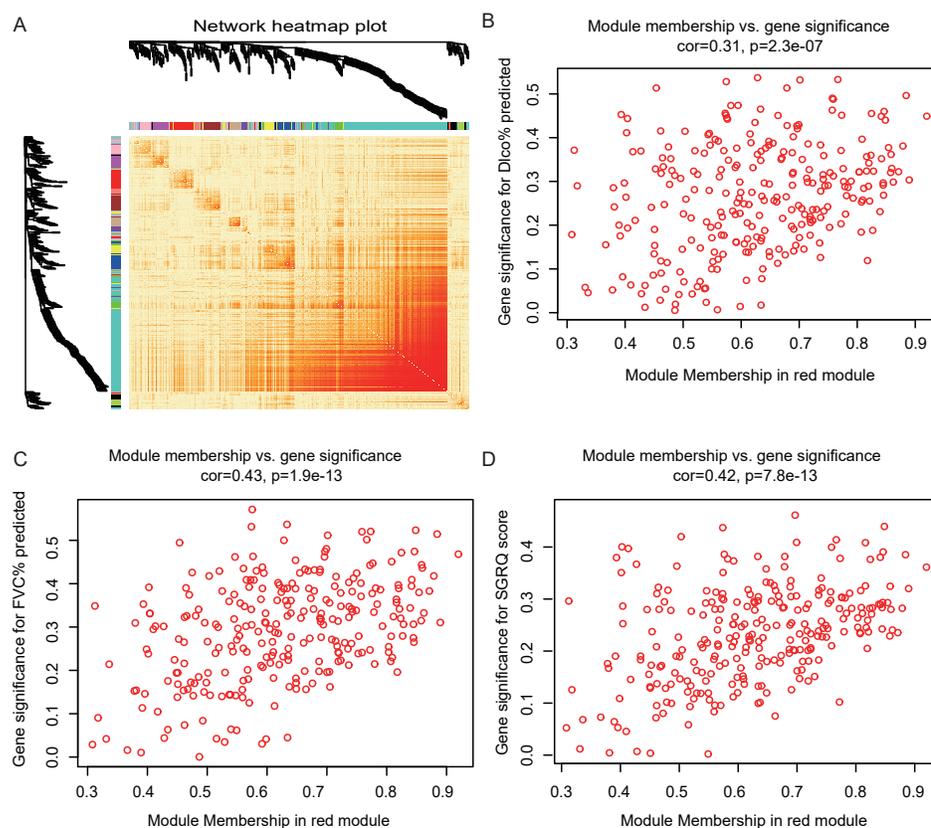
Figure 3 Module analysis. (A) Network heatmap plot of randomly selecting 400 genes. Branch represents gene cluster and each color under cluster represents one co-expression gene module. Light red indicates low overlap and darker red indicates higher overlap. The heatmap indicated the high independence between each module and genes at the tip of the module branches have high intramodular connectivity with the rest of the genes in this module. (B–D): scatter plot of gene significance (GS) for Dlco% predicted, FVC% predicted and SGRQ score versus module membership (MM) in the red module. GS is significantly correlated with MM, which indicated that the hub genes in the red module tended to be highly associated with lung function.

Construction of the PPI network and identification of hub genes

We uploaded 32 candidate hub genes into the STRING database to construct a PPI network (Fig. S2). Cytoscape software was used to visualize the interaction relationships among these hub nodes. The nodes with no connections were removed. The final network contained 20 nodes and 35 edges (Fig. 4C). The nodes with the top 5 degree of connectivity were interleukin-6 (*IL6*), MYC proto-oncogene (*MYC*), serpin family E member 1 (*SERPINE1*), thrombospondin-1 (*THBS1*), suppressor of cytokine signaling 3 (*SOCS3*), and CCAAT enhancer-binding protein delta (*CEBPD*).

The differentially expression genes *IL6*, *THBS1*, *SERPINE1*, and *SOCS3* were also included in the previously described nodes with great degree of connectivity in the PPI network (Fig. 4C). Therefore, we identified *IL6*, *SERPINE1*, *THBS1*, and *SOCS3* as hub genes, based on differential expression and connectivity.

Table 2 GO enrichment analysis in red module.

Category	ID	Description	P-value	Count
BP	GO:0002237	response to molecule of bacterial origin	4.95E−21	34
BP	GO:0032496	response to lipopolysaccharide	1.16E−20	33
BP	GO:0002521	leukocyte differentiation	8.93E−16	35
BP	GO:0060326	cell chemotaxis	1.29E−14	25
BP	GO:0050727	regulation of inflammatory response	2.41E−14	33
BP	GO:0048511	rhythmic process	2.32E−13	25
BP	GO:0071222	cellular response to lipopolysaccharide	4.61E−13	20
BP	GO:0071219	cellular response to molecule of bacterial origin	9.66E−13	20
BP	GO:0071216	cellular response to biotic stimulus	1.01E−12	21
BP	GO:0001819	positive regulation of cytokine production	4.26E−12	29
CC	GO:0101003	ficolin-1-rich granule membrane	1.39E−05	7
CC	GO:0070820	tertiary granule	5.88E−05	10
MF	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	3.56E−11	26
MF	GO:0005125	cytokine activity	2.30E−07	13
MF	GO:0000978	RNA polymerase II proximal promoter sequence-specific DNA binding	2.61E−06	20
MF	GO:0070888	E-box binding	4.68E−06	7
MF	GO:0000987	proximal promoter sequence-specific DNA binding	5.15E−06	20
MF	GO:0030545	receptor regulator activity	7.72E−06	19
MF	GO:0050786	RAGE receptor binding	1.10E−05	4
MF	GO:0048018	receptor ligand activity	1.11E−05	18
MF	GO:0001227	DNA-binding transcription repressor activity, RNA polymerase II-specific	1.51E−05	13
MF	GO:0000980	RNA polymerase II distal enhancer sequence-specific DNA binding	2.74E−05	8

Table 3 KEGG enrichment analysis in red module.

ID	Description	p-value	Count
hsa04668	TNF signaling pathway	9.26E−16	21
hsa04657	IL-17 signaling pathway	1.11E−12	17
hsa04625	C-type lectin receptor signaling pathway	7.24E−10	15
hsa04380	Osteoclast differentiation	1.36E−08	15
hsa05144	Malaria	2.19E−08	10
hsa04060	Cytokine-cytokine receptor interaction	2.44E−08	22
hsa04933	AGE-RAGE signaling pathway in diabetic complications	3.17E−07	12
hsa04064	NF-kappa B signaling pathway	4.88E−07	12
hsa04630	JAK-STAT signaling pathway	1.89E−06	14
hsa04061	Viral protein interaction with cytokine and cytokine receptor	2.41E−06	11
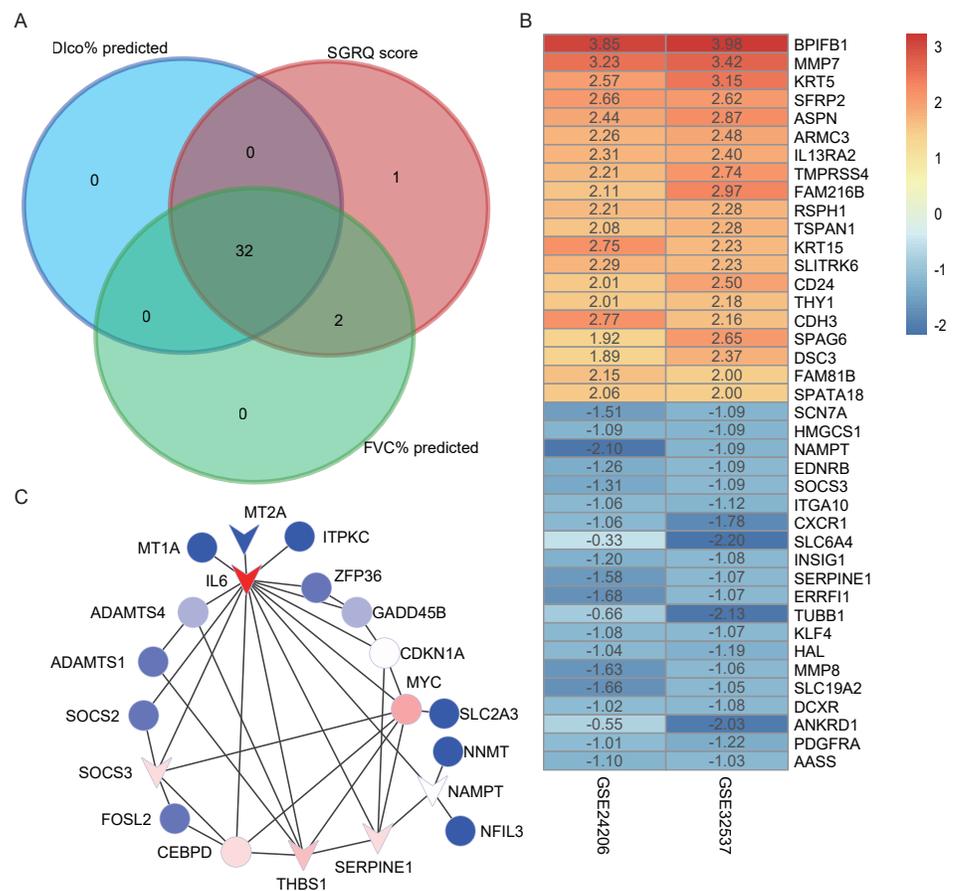
Figure 4 **Construct PPI network and identify hub genes.** (A) The intersection of candidate hub genes for clinical traits: Dlco% predicted, FVC% predicted and SGRQ score. (B) The top 20 upregulated genes and top 20 downregulated genes in IPF compared with healthy donors. The numbers in each rectangle show the logarithmic fold-change of genes in each dataset. Red represents upregulated gene and blue represents downregulated genes. (C) The PPI network visualized by Cytoscape software. We removed the 11 nodes which had no connections with others. The network contains 20 nodes and 35 edges. The triangle represented downregulation in IPF. Degree of connectivity was showed by different colors; the darker shade of red represents the higher degree of connectivity and, conversely, the darker shade of blue color indicates the lower degree of connectivity.

Validation of hub genes using additional GEO datasets

We compared the expression levels of hub genes between IPF and healthy lung tissues in the GSE47460 dataset (Figs. 5A–5C) and found that *IL6*, *SERPINE1*, and *SOCS3* were significantly downregulated in IPF patients compared with healthy controls. However, *THBS1* showed no significantly different expression ($p = 0.75$), which was excluded from further analyses (Fig. S3). In the GSE24206 dataset, the expression of the final three hub genes were lower in the early IPF group than in the advanced IPF group, although the expression of *SOCS3* did not differ significantly between the two groups ($p = 0.074$, Figs. 5D–5F). The Spearman correlation coefficients between Dlco% predicted and *IL6*, *SERPINE1*, *SOCS* were calculated as -0.32, -0.41, and -0.46, respectively (Figs. 6A–6C).

Table 4 GS and MM of 32 candidate hub genes in red module.

| Gene | SGRQ score | | FVC% predicted | | Dlco% predicted | | | |
	GS	*P*-value	GS	*P*-value	GS	*P*-value	MM	*P*-value
NAMPT	0.36	4.34E−04	−0.47	2.90E−06	−0.45	8.02E−06	0.92	6.27E−38
GADD45B	0.32	1.98E−03	−0.31	2.75E−03	−0.30	3.47E−03	0.89	4.89E−32
THBS1	0.39	1.62E−04	−0.51	1.83E−07	−0.50	5.65E−07	0.88	3.46E−31
FOSL2	0.28	6.70E−03	−0.42	3.77E−05	−0.38	1.96E−04	0.88	2.49E−30
MYC	0.24	2.47E−02	−0.38	1.79E−04	−0.32	1.85E−03	0.87	2.99E−29
ITPKC	0.24	2.17E−02	−0.33	1.28E−03	−0.30	4.06E−03	0.86	5.16E−28
MT2A	0.28	7.17E−03	−0.44	1.55E−05	−0.33	1.41E−03	0.86	1.12E−27
NNMT	0.32	1.70E−03	−0.44	1.07E−05	−0.36	4.28E−04	0.86	1.59E−27
IL6	0.23	2.90E−02	−0.31	2.38E−03	−0.32	1.97E−03	0.86	1.96E−27
ERRFI1	0.29	4.82E−03	−0.38	2.04E−04	−0.38	2.27E−04	0.85	5.48E−27
SERPINE1	0.44	1.31E−05	−0.52	1.03E−07	−0.44	1.27E−05	0.85	2.54E−26
NFIL3	0.39	1.30E−04	−0.47	2.04E−06	−0.46	4.48E−06	0.85	4.49E−26
ITPRIP	0.29	4.77E−03	−0.27	1.11E−02	−0.27	1.01E−02	0.85	5.69E−26
PPP1R15B	0.24	2.06E−02	−0.31	3.14E−03	−0.29	5.05E−03	0.84	6.93E−26
SOCS3	0.28	6.45E−03	−0.27	8.37E−03	−0.30	4.31E−03	0.84	7.13E−26
SLC19A2	0.29	5.37E−03	−0.41	5.63E−05	−0.35	6.25E−04	0.84	1.44E−25
ADAMTS4	0.38	2.01E−04	−0.39	1.09E−04	−0.45	8.18E−06	0.84	1.61E−25
ZFP36	0.29	5.28E−03	−0.34	8.46E−04	−0.23	2.76E−02	0.84	2.86E−25
ADAMTS1	0.28	6.54E−03	−0.31	3.16E−03	−0.32	1.76E−03	0.83	4.31E−24
PHLDA1	0.38	2.37E−04	−0.47	2.38E−06	−0.46	5.78E−06	0.82	9.88E−24
SLC2A3	0.28	6.77E−03	−0.32	1.99E−03	−0.36	4.59E−04	0.82	1.25E−23
MT1M	0.24	2.38E−02	−0.34	9.14E−04	−0.26	1.47E−02	0.82	2.04E−23
MT1JP	0.26	1.31E−02	−0.39	1.38E−04	−0.34	9.10E−04	0.82	2.75E−23
MT1A	0.20	5.13E−02	−0.40	9.79E−05	−0.29	5.49E−03	0.82	3.22E−23
C11orf96	0.41	5.88E−05	−0.36	5.21E−04	−0.36	5.22E−04	0.82	6.40E−23
CDKN1A	0.33	1.65E−03	−0.36	3.92E−04	−0.32	1.74E−03	0.82	6.53E−23
CEBPD	0.27	9.00E−03	−0.32	1.72E−03	−0.28	6.34E−03	0.81	1.88E−22
APOLD1	0.28	6.42E−03	−0.23	2.85E−02	−0.26	1.37E−02	0.81	2.85E−22
SLCO4A1	0.32	2.26E−03	−0.45	7.30E−06	−0.40	1.00E−04	0.81	3.52E−22
RNF122	0.23	2.58E−02	−0.28	7.49E−03	−0.25	1.56E−02	0.81	4.16E−22
PELI1	0.25	1.89E−02	−0.25	1.49E−02	−0.32	1.87E−03	0.80	9.76E−22
SOCS2	0.26	1.15E−02	−0.24	2.09E−02	−0.32	1.89E−03	0.80	9.89E−22

The Spearman correlation coefficients between FVC% predicted and *IL6*, *SERPINE1*, *SOCS3* were calculated as −0.29, −0.33, and −0.27, respectively (Figs. 6D–6F).

DISCUSSION

In the present study, we found a significant inverse correlation between the red module and lung function by performing WGCNA on an IPF dataset. We identified three hub genes, including *IL6*, *SERPINE1,* and *SOCS3*. The expression levels of *IL6*, *SOCS3*, and *SERPINE1* were negatively correlated with lung function, and the advanced IPF patients
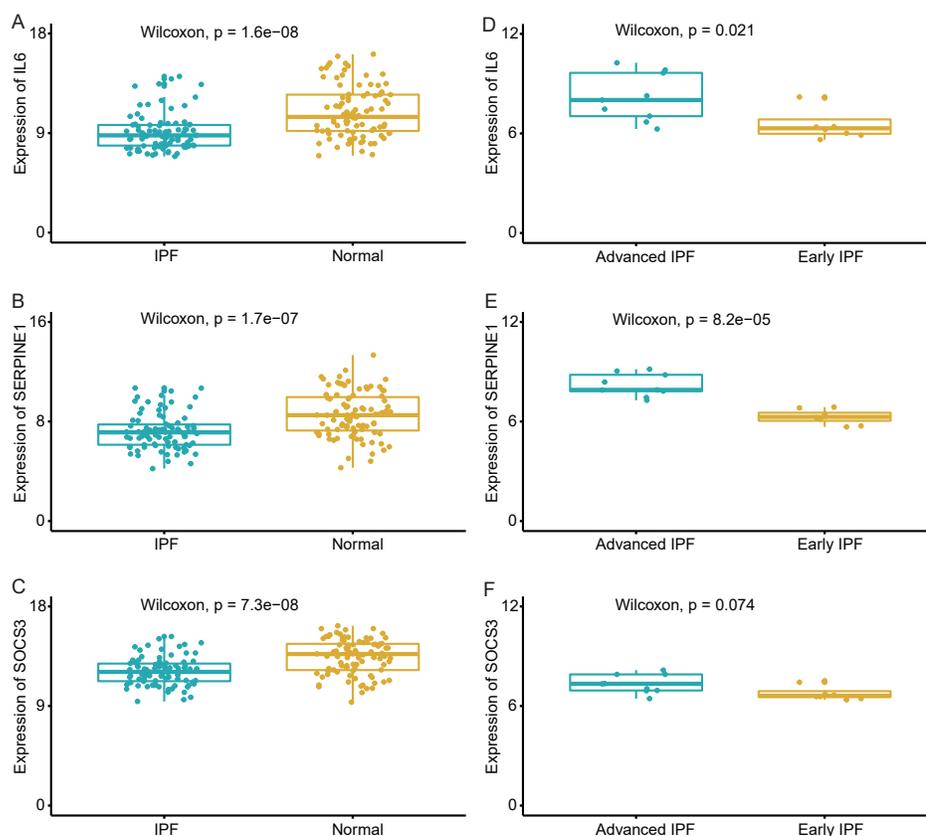
Figure 5 **Validate the expression of hub genes in different groups.** (A–C) In GSE47460 dataset, IL6, SERPINE1 and SOCS3 were significantly downregulated in IPF when compared with normal lung tissues. (D–E) In GSE24206 dataset, IL6 and SERPINE1 were significantly overexpressed in advanced IPF when compared with early IPF, (F) while the expression of SOCS3 was no statistical difference between the two groups ($p = 0.074$).

had higher expression levels of these genes than early IPF patients in the validated datasets. The most important characteristic of patients with IPF is a decline in lung function, and declines of FVC and Dlco can predict mortality risk (*Frankel & Schwarz, 2009*; *Nathan et al., 2011*). However, the pathophysiologic mechanisms of lung dysfunction remain unclear. Our results may provide insights into the pathogenesis underlying the progression of lung function.

WGCNA is a bioinformatic algorithm and has been used to identify candidate biomarkers and therapeutic targets for many diseases, especially in cancer and neuroscience research (*Giulietti et al., 2017*; *Li et al., 2020*; *Niemira et al., 2019*; *Rangaraju et al., 2018*; *Spiers et al., 2015*; *Zeleznik et al., 2020*). We can identify clusters (modules) of highly correlated genes using WGCNA. WGCNA can systematically study the interconnectedness among all genes and convert gene expression data into a weighted co-expression network, which represents its most important advantage (*Zhang & Horvath, 2005*; *Zhao et al., 2010*). Based on the module significance, we can incorporate external clinical information into the network
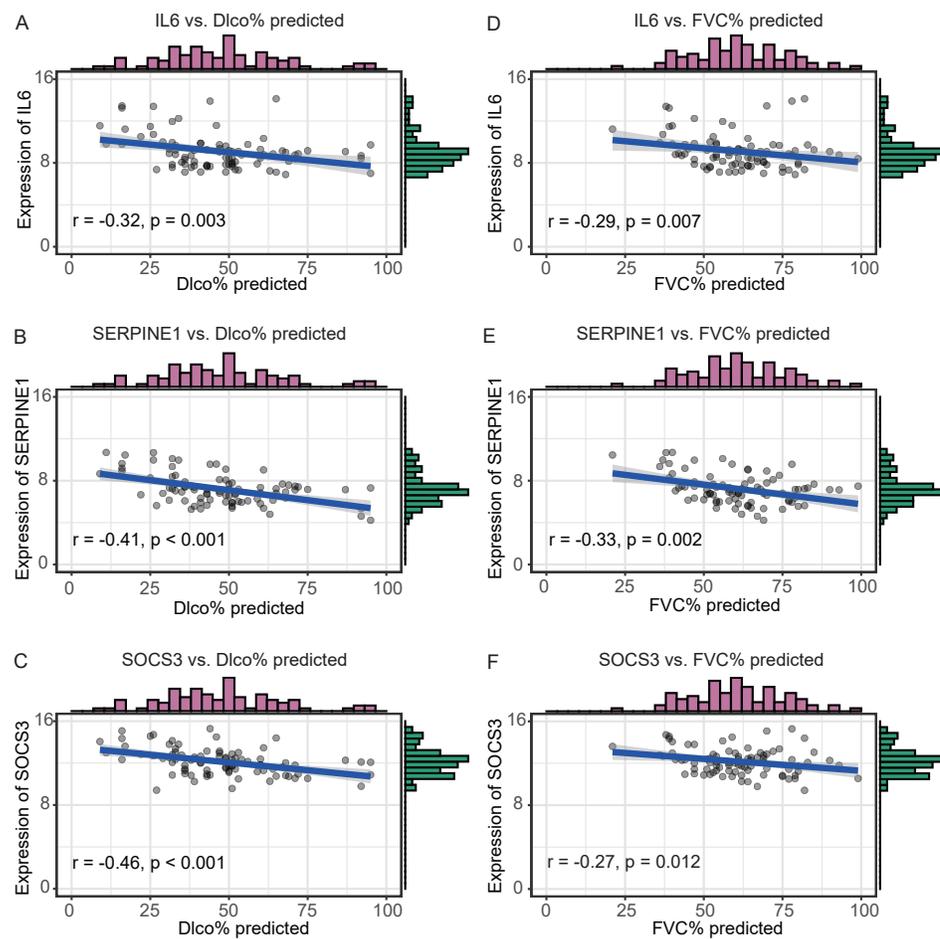
Figure 6 **Validate the negative relationship between hub genes and lung function in GSE47460 dataset.**
(A–C) The relationship between expression levels of IL6, SERPINE1, SOCS3 and the Dlco% predicted of patients with IPF. Spearman correlation coefficients between Dlco% predicted and expression levels of IL6, SERPINE1, SOCS3 were −0.32 ($p = 0.003$), −0.41 ($p < 0.001$), and −0.46 ($p < 0.001$), respectively. (D–F) The relationship between expression levels of IL6, SERPINE1, SOCS3 and the FVC% predicted of patients with IPF. Spearman correlation coefficients between FVC% predicted and expression levels of IL6, SERPINE1, SOCS3 were −0.29 ($p = 0.007$), −0.33 ($p = 0.002$), and −0.27 ($p = 0.012$), respectively.

and identify key modules and hub genes, which are believed to play core roles in the pathogenesis of the disease.

The GSE32537 and GSE47460 datasets contain the transcriptomic profiles and clinical characteristics of the included subjects. Studies examining IPF have been previously performed using these datasets. Yang and colleagues (*2013*) analyzed the GSE32537 dataset and found that the high expression of cilium-associated genes was associated with increased microscopic honeycombing. McDonough and colleagues selected DEGs to construct co-expression networks by performing WGCNA on GSE47460 dataset and identified regulatory factors that were associated with co-expression networks in IPF (*McDonough et al., 2019*). In the present study, we selected the top 25% variant genes to

perform WGCNA on patients with IPF in the GSE32537 dataset. We found that the red module had the strongest negative correlations with FVC% predicted and Dlco% predicted. We identified *IL6*, *SERPINE1,* and *SOCS3* as hub genes in the red module. We also validated the relationships between expression levels of these hub genes and lung function using independent datasets. To our knowledge, this study is the first to identify and validate genes that are negatively associated with lung function, based on transcriptomic files combined with the WGCNA approach.

In this study, the red module had the strongest negative correlation with lung function. The enrichment analysis of genes in the red module showed that they were primarily associated with inflammatory and immune responses, which indicated that the inflammatory and immune pathways are involved in the pathophysiologic mechanisms of lung dysfunction. Although multicenter trials of anti-inflammatory drugs for IPF treatment have failed (*Farrand et al., 2020*; *King et al., 2009*; *Raghu et al., 2004*; *Raghu et al., 2017*; *Raghu et al., 2008*), the immune system continues to be regarded as playing an important role in the development of fibrosis (*Heukels et al., 2019*; *Wynn, 2011*). Furthermore, many studies have also confirmed that changes in immune activity or the proportions of immune cell populations may be associated with declines in lung function (*Adegunsoye et al., 2016*; *Gilani et al., 2010*; *Xue et al., 2013*).

Interestingly, all three hub genes were downregulated in the IPF group compared with their levels in the healthy group. We speculated that hub genes had different molecular functions under different conditions. IL6 is a multifunctional cytokine, belonging to the IL-6 family of cytokines. Various cells, including alveolar macrophages, lung fibroblasts, and fibrocytes, can express and secret IL6 (*Shahar et al., 1996*). Takizawa and colleagues observed that IL6 concentrations were significantly higher in bronchoalveolar lung fluid (BALF) from IPF patients than in BALF from healthy controls (*Takizawa et al., 1997*). However, in this study, when compared with healthy lung tissues, the expression of *IL6* was lower in IPF tissues. A previous study demonstrated that IL6 promoted the proliferation of IPF lung fibroblasts but inhibited the proliferation of normal lung fibroblasts (*Moodley et al., 2003b*). In another study, in lung fibroblasts derived from IPF patients, IL6 contributed to resistance against Fas-induced apoptosis by increasing the expression of the anti-apoptotic protein BCL-2, whereas normal lung fibroblasts became more sensitive to Fas-induced apoptosis, which was mediated by the increased expression of the pro-apoptotic protein Bax when exposed to IL6 (*Moodley et al., 2003a*). In the present study, the expression of *IL6* was negatively associated with lung function, which also indicated IL6 promoted the progression of IPF.

Suppressor of cytokine signaling-3 (SOCS3) is a well-known regulatory cornerstone of intracellular signaling. SOCS3 not only acts as a feedback inhibitor of the JAK/STAT signaling pathway but can also regulate many cytokines, growth factors, and hormones associated with many cellular processes (*Mahony et al., 2016*). Whether SOCS3 acts to protect against or promote disease progression depends on the cells and pathological processes in which it is expressed, especially in innate and adaptive immunity (*Kubo, Hanada & Yoshimura, 2003*; *Yasukawa et al., 2003*). A previous study demonstrated that silencing *Socs3* in a rat diastolic heart failure model was able to significantly diminish

myocardial fibrosis and the inflammatory response (*Gao et al., 2019*). SOCS3 function has also been studied in lung diseases (*Gao & Ward, 2007*). Studies have revealed that SOCS3 acts as a pro-inflammatory molecule, by suppressing the IL-6-gp130 signaling pathway, and mice lacking *Socs3* in macrophages and neutrophils were resistant to LPS-induced shock (*Yasukawa et al., 2003*). Aboulhoda studied age-dependent *SOCS3* expression and myocardial fibrosis, and found that *SOCS3* activity was correlated with myocardial fibrosis (*Aboulhoda, 2017*). IPF is an aging-related disease, but the role played by *SOCS3* in pulmonary fibrosis has not been well-studied. The present study revealed a negative correlation between *SOCS3* expression and lung function in IPF patients, but the detailed mechanisms require further study.

Serpin Family E Member 1 (SERPINE1), also known as plasminogen activator inhibitor-1 (PAI-1), is the primary inhibitor of plasminogen activators, such as tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA), and acts as a major regulator of the fibrinolytic system. Impaired fibrinolytic activity is a common characteristic of acute and chronic inflammatory lung diseases, especially pulmonary fibrosis (*Marudamuthu et al., 2015*). Eitzman studied transgenic mice that either overexpressed or were completely deficient in murine *Serpine1*, and found that higher levels of *Serpine1* expression can increase collagen accumulation following inflammatory lung injury (*Eitzman et al., 1996*). Osterholzer and colleagues studied type-II alveoli epithelial cells in a lung injury model and found results consistent with those reported by previous studies (*Osterholzer et al., 2012*). Senoo and colleagues directly suppressed the expression of *Serpine1* in mice, through the intrapulmonary administration of *Serpine1*-siRNA, to reduce pulmonary fibrosis. They found that the suppression of epithelial-to-mesenchymal transformation may be involved in IPF (*Senoo et al., 2010*). The present study confirmed a negative correlation between the expression level of *SERPINE1* and IPF.

We identified *IL6*, *SOCS3*, and *SERPINE1* as IPF hub genes that were negatively associated with lung function. These hub genes may serve as therapeutic targets for IPF treatment. The downregulation of *SERPINE1* has been shown to attenuate pulmonary fibrosis (*Senoo et al., 2010*), indicating the reliability of the our results. However, the present study also has some limitations. First, the study is based on bioinformatics analysis, and the results remain to be verified by further research. Second, datasets that met the inclusion criteria were rare, which may decrease the statistical effectiveness. Finally, the difference of GeneChips between the datasets may also affect the reliability of the results.

CONCLUSIONS

In summary, we performed WGCNA on an IPF dataset. Among 14 modules, the red module was identified as a key module because it displayed the strongest correlation with lung function. Genes in the red module were primarily enriched in inflammatory and immune pathways. *IL6*, *SOCS3,* and *SERPINE1* were identified as hub genes from the red module. We also found that *IL6*, *SOCS3,* and *SERPINE1* were negatively associated with lung function in IPF patients. These results may suggest that further study is warranted to investigate the roles played by hub genes in IPF progression. Based on this research, the

proteins encoded by these hub genes may serve as biomarkers for IPF severity and may represent therapeutic targets for IPF.

Funding

This study was supported by the National Natural Science Foundation of China (No. 81770002), the Science and Technology Program of Changsha, China (No. kq1901120), the National Key Clinical Specialty Construction Projects of China, Hunan Provincial Innovation Foundation for Postgraduate, and the Fundamental Research Funds for the Central Universities of Central South University (No. 2020zzts281). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 81770002.
Science and Technology Program of Changsha, China: kq1901120.
National Key Clinical Specialty Construction Projects of China.
Hunan Provincial Innovation Foundation for Postgraduate.
The Fundamental Research Funds for the Central Universities of Central South University: 2020zzts281.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yuechong Xia conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Cheng Lei analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Danhui Yang analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Hong Luo conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

Data are available at NCBI GEO: GSE32537, GSE24206, GSE47460 and the raw data and results of WGCNA are available in the Supplementary Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.9848#supplemental-information.

REFERENCES

Aboulhoda BE. 2017. Age-related remodeling of the JAK/STAT/SOCS signaling pathway and associated myocardial changes: from histological to molecular level. *Annals of anatomy = Anatomischer Anzeiger: official organ of the Anatomische Gesellschaft* **214**:21–30 DOI 10.1016/j.aanat.2017.07.003.

Adegunsoye A, Hrusch CL, Bonham CA, Jaffery MR, Blaine KM, Sullivan M, Churpek MM, Strek ME, Noth I, Sperling AI. 2016. Skewed Lung CCR4 to CCR6 CD4 T Cell Ratio in Idiopathic Pulmonary Fibrosis Is Associated with Pulmonary Function. *Frontiers in Immunology* **7**:516 DOI 10.3389/fimmu.2016.00516.

Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**:293 DOI 10.1186/1471-2105-15-293.

Boon K, Bailey NW, Yang J, Steel MP, Groshong S, Kervitsky D, Brown KK, Schwarz MI, Schwartz DA. 2009. Molecular phenotypes distinguish patients with relatively stable from progressive idiopathic pulmonary fibrosis (IPF). *PLOS ONE* **4**:e5134 DOI 10.1371/journal.pone.0005134.

Costabel U, Albera C, Glassberg MK, Lancaster LH, Wuyts WA, Petzinger U, Gilberg F, Kirchgaessler K-U, Noble PW. 2019. Effect of pirfenidone in patients with more advanced idiopathic pulmonary fibrosis. *Respiratory Research* **20**:55 DOI 10.1186/s12931-019-1021-2.

Costabel U, Albera C, Lancaster LH, Lin C-Y, Hormel P, Hulter HN, Noble PW. 2017. An open-label study of the long-term safety of pirfenidone in patients with idiopathic pulmonary fibrosis (RECAP). *Respiration; International Review of Thoracic Diseases* **94**:408–415 DOI 10.1159/000479976.

Eitzman DT, McCoy RD, Zheng X, Fay WP, Shen T, Ginsburg D, Simon RH. 1996. Bleomycin-induced pulmonary fibrosis in transgenic mice that either lack or overexpress the murine plasminogen activator inhibitor-1 gene. *The Journal of Clinical Investigation* **97**:232–237 DOI 10.1172/JCI118396.

Farrand E, Vittinghoff E, Ley B, Butte AJ, Collard HR. 2020. Corticosteroid use is not associated with improved outcomes in acute exacerbation of IPF. *Respirology* **25**:629–635 DOI 10.1111/resp.13753.

Frankel SK, Schwarz MI. 2009. Update in idiopathic pulmonary fibrosis. *Current Opinion in Pulmonary Medicine* **15**:463–469 DOI 10.1097/MCP.0b013e32832ea4b3.

Gao H, Ward PA. 2007. STAT3 and suppressor of cytokine signaling 3: potential targets in lung inflammatory responses. *Expert Opinion on Therapeutic Targets* **11**:869–880 DOI 10.1517/14728222.11.7.869.

Gao J, Guo Y, Chen Y, Zhou J, Liu Y, Su P. 2019. Adeno-associated virus 9-mediated RNA interference targeting SOCS3 alleviates diastolic heart failure in rats. *Gene* **697**:11–18 DOI 10.1016/j.gene.2019.01.044.

Gilani SR, Vuga LJ, Lindell KO, Gibson KF, Xue J, Kaminski N, Valentine VG, Lindsay EK, George MP, Steele C, Duncan SR. 2010. CD28 down-regulation on circulating CD4 T-cells is associated with poor prognoses of patients with idiopathic pulmonary fibrosis. *PLOS ONE* **5**:e8959 DOI 10.1371/journal.pone.0008959.

Giulietti M, Occhipinti G, Principato G, Piva F. 2017. Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. *Cellular Oncology* **40**:181–192 DOI 10.1007/s13402-017-0315-y.

Heukels P, Moor CC, Von der Thüsen JH, Wijsenbeek MS, Kool M. 2019. Inflammation and immunity in IPF pathogenesis and treatment. *Respiratory Medicine* **147**:79–91 DOI 10.1016/j.rmed.2018.12.015.

Huang Y, Ma S-F, Vij R, Oldham JM, Herazo-Maya J, Broderick SM, Strek ME, White SR, Hogarth DK, Sandbo NK, Lussier YA, Gibson KF, Kaminski N, Garcia JGN, Noth I. 2015. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulmonary Medicine* **15**:147 DOI 10.1186/s12890-015-0142-8.

Kim HJ, Perlman D, Tomic R. 2015. Natural history of idiopathic pulmonary fibrosis. *Respiratory Medicine* **109**:661–670 DOI 10.1016/j.rmed.2015.02.002.

King TE, Albera C, Bradford WZ, Costabel U, Hormel P, Lancaster L, Noble PW, Sahn SA, Szwarcberg J, Thomeer M, Valeyre D, Du Bois RM. 2009. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (IN-SPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet* **374**:222–228 DOI 10.1016/S0140-6736(09)60551-1.

Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW, Kim DS, Kaminski N. 2009. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine* **180**:167–175 DOI 10.1164/rccm.200810-1596OC.

Kubo M, Hanada T, Yoshimura A. 2003. Suppressors of cytokine signaling and immunity. *Nature Immunology* **4**:1169–1176 DOI 10.1038/ni1012.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559 DOI 10.1186/1471-2105-9-559.

Lederer DJ, Martinez FJ. 2018. Idiopathic pulmonary fibrosis. *The New England Journal of Medicine* **378**:1811–1823 DOI 10.1056/NEJMra1705751.

Li W, Wang L, Wu Y, Yuan Z, Zhou J. 2020. Weighted gene co-expression network analysis to identify key modules and hub genes associated with atrial fibrillation. *International Journal of Molecular Medicine* **45**:401–416 DOI 10.3892/ijmm.2019.4416.

Mahony R, Ahmed S, Diskin C, Stevenson NJ. 2016. SOCS3 revisited: a broad regulator of disease, now ready for therapeutic use? *Cellular and Molecular Life Sciences: CMLS* **73**:3323–3336 DOI 10.1007/s00018-016-2234-x.

Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, Swigris JJ, Taniguchi H, Wells AU. 2017. Idiopathic pulmonary fibrosis. *Nature Reviews Disease Primers* **3**:17074 DOI 10.1038/nrdp.2017.74.

Marudamuthu AS, Bhandary YP, Shetty SK, Fu J, Sathish V, Prakash Y, Shetty S. 2015. Role of the urokinase-fibrinolytic system in epithelial-mesenchymal transition during lung injury. *The American Journal of Pathology* **185**:55–68 DOI 10.1016/j.ajpath.2014.08.027.

McDonough JE, Kaminski N, Thienpont B, Hogg JC, Vanaudenaerde BM, Wuyts WA. 2019. Gene correlation network analysis to identify regulatory factors in idiopathic pulmonary fibrosis. *Thorax* **74**:132–140 DOI 10.1136/thoraxjnl-2018-211929.

Meltzer EB, Barry WT, D'Amico TA, Davis RD, Lin SS, Onaitis MW, Morrison LD, Sporn TA, Steele MP, Noble PW. 2011. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. *Bmc Medical Genomics* **4**:70 DOI 10.1186/1755-8794-4-70.

Meng X, Wang H, Song X, Clifton AC, Xiao J. 2020. The potential role of senescence in limiting fibrosis caused by aging. *Journal of Cellular Physiology* **235**:4046–4059 DOI 10.1002/jcp.29313.

Moodley YP, Misso NLA, Scaffidi AK, Fogel-Petrovic M, McAnulty RJ, Laurent GJ, Thompson PJ, Knight DA. 2003a. Inverse effects of interleukin-6 on apoptosis of fibroblasts from pulmonary fibrosis and normal lungs. *American Journal of Respiratory Cell and Molecular Biology* **29**:490–498 DOI 10.1165/rcmb.2002-0262OC.

Moodley YP, Scaffidi AK, Misso NL, Keerthisingam C, McAnulty RJ, Laurent GJ, Mutsaers SE, Thompson PJ, Knight DA. 2003b. Fibroblasts isolated from normal lungs and those with idiopathic pulmonary fibrosis differ in interleukin-6/gp130-mediated cell signaling and proliferation. *The American Journal of Pathology* **163**:345–354 DOI 10.1016/S0002-9440(10)63658-9.

Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, Sheridan MJ, du Bois RM. 2011. Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. *Chest* **140**:221–229 DOI 10.1378/chest.10-2572.

Niemira M, Collin F, Szalkowska A, Bielska A, Chwialkowska K, Reszec J, Niklinski J, Kwasniewski M, Kretowski A. 2019. Molecular signature of subtypes of non-small-cell lung cancer by large-scale transcriptional profiling: identification of key modules and genes by weighted gene co-expression network analysis (WGCNA). *Cancers* **12(1)**:37 DOI 10.3390/cancers12010037.

Osterholzer JJ, Christensen PJ, Lama V, Horowitz JC, Hattori N, Subbotina N, Cunningham A, Lin Y, Murdock BJ, Morey RE, Olszewski MA, Lawrence DA, Simon RH, Sisson TH. 2012. PAI-1 promotes the accumulation of exudate macrophages and worsens pulmonary fibrosis following type II alveolar epithelial cell injury. *The Journal of Pathology* **228**:170–180 DOI 10.1002/path.3992.

Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, King TE. 2004. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *The New England Journal of Medicine* **350**:125–133 DOI 10.1056/NEJMoa030511.

Raghu G, Brown KK, Collard HR, Cottin V, Gibson KF, Kaner RJ, Lederer DJ, Martinez FJ, Noble PW, Song JW, Wells AU, Whelan TPM, Wuyts W, Moreau E, Patterson SD, Smith V, Bayly S, Chien JW, Gong Q, Zhang JJ, O'Riordan TG. 2017. Efficacy of simtuzumab versus placebo in patients with idiopathic pulmonary fibrosis: a randomised, double-blind, controlled, phase 2 trial. *The Lancet Respiratory Medicine* **5**:22–32 DOI 10.1016/S2213-2600(16)30421-0.

Raghu G, Brown KK, Costabel U, Cottin V, du Bois RM, Lasky JA, Thomeer M, Utz JP, Khandker RK, McDermott L, Fatenejad S. 2008. Treatment of idiopathic pulmonary fibrosis with etanercept: an exploratory, placebo-controlled trial. *American Journal of Respiratory and Critical Care Medicine* **178**:948–955 DOI 10.1164/rccm.200709-1446OC.

Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier J-F, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TE, Kondoh Y, Myers J, Müller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, Griss BS, Protzko SL, Schünemann HJ. 2011. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine* **183**:788–824 DOI 10.1164/rccm.2009-040GL.

Rangaraju S, Dammer EB, Raza SA, Rathakrishnan P, Xiao H, Gao T, Duong DM, Pennington MW, Lah JJ, Seyfried NT, Levey AI. 2018. Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. *Molecular Neurodegeneration* **13**:24 DOI 10.1186/s13024-018-0254-8.

Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, Cottin V, Flaherty KR, Hansell DM, Inoue Y, Kim DS, Kolb M, Nicholson AG, Noble PW, Selman M, Taniguchi H, Brun M, Le Maulf F, Girard M, Stowasser S, Schlenker-Herceg R, Disse B, Collard HR. 2014. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *The New England Journal of Medicine* **370**:2071–2082 DOI 10.1056/NEJMoa1402584.

Richeldi L, Collard HR, Jones MG. 2017. Idiopathic pulmonary fibrosis. *Lancet* **389**:1941–1952 DOI 10.1016/S0140-6736(17)30866-8.

Richeldi L, Kolb M, Jouneau S, Wuyts WA, Schinzel B, Stowasser S, Quaresma M, Raghu G. 2020. Efficacy and safety of nintedanib in patients with advanced idiopathic pulmonary fibrosis. *BMC Pulmonary Medicine* **20**:3 DOI 10.1186/s12890-019-1030-4.

Senoo T, Hattori N, Tanimoto T, Furonaka M, Ishikawa N, Fujitaka K, Haruta Y, Murai H, Yokoyama A, Kohno N. 2010. Suppression of plasminogen activator inhibitor-1 by RNA interference attenuates pulmonary fibrosis. *Thorax* **65**:334–340 DOI 10.1136/thx.2009.119974.

Shahar I, Fireman E, Topilsky M, Grief J, Kivity S, Spirer Z, Ben Efraim S. 1996. Effect of IL-6 on alveolar fibroblast proliferation in interstitial lung diseases. *Clinical Immunology and Immunopathology* **79**:244–251 DOI 10.1006/clin.1996.0075.

Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O'Donovan MC, Bray NJ, Mill J. 2015. Methylomic trajectories across human fetal brain development. *Genome Research* **25**:338–352 DOI 10.1101/gr.180273.114.

Takizawa H, Satoh M, Okazaki H, Matsuzaki G, Suzuki N, Ishii A, Suko M, Okudaira H, Morita Y, Ito K. 1997. Increased IL-6 and IL-8 in bronchoalveolar lavage fluids (BALF) from patients with sarcoidosis: correlation with the

clinical parameters. *Clinical and Experimental Immunology* **107**:175–181
DOI 10.1046/j.1365-2249.1997.d01-905.x.

Todd JL, Neely ML, Overton R, Durham K, Gulati M, Huang H, Roman J, Newby LK, Flaherty KR, Vinisko R, Liu Y, Roy J, Schmid R, Strobel B, Hesslinger C, Leonard TB, Noth I, Belperio JA, Palmer SM. 2019. Peripheral blood proteomic profiling of idiopathic pulmonary fibrosis biomarkers in the multicentre IPF-PRO Registry. *Respiratory Research* **20**:227 DOI 10.1186/s12931-019-1190-z.

Wynn TA. 2011. Integrating mechanisms of pulmonary fibrosis. *The Journal of Experimental Medicine* **208**:1339–1350 DOI 10.1084/jem.20110551.

Xue J, Kass DJ, Bon J, Vuga L, Tan J, Csizmadia E, Otterbein L, Soejima M, Levesque MC, Gibson KF, Kaminski N, Pilewski JM, Donahoe M, Sciurba FC, Duncan SR. 2013. Plasma B lymphocyte stimulator and B cell differentiation in idiopathic pulmonary fibrosis patients. *Journal of Immunology* **191**:2089–2095 DOI 10.4049/jimmunol.1203476.

Yang IV, Burch LH, Steele MP, Savov JD, Hollingsworth JW, McElvania-Tekippe E, Berman KG, Speer MC, Sporn TA, Brown KK, Schwarz MI, Schwartz DA. 2007. Gene expression profiling of familial and sporadic interstitial pneumonia. *American Journal of Respiratory and Critical Care Medicine* **175**:45–54 DOI 10.1164/rccm.200601-062OC.

Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, Rosen R, Neidermyer AJ, McKean DF, Groshong SD, Cool C, Cosgrove GP, Lynch DA, Brown KK, Schwarz MI, Fingerlin TE, Schwartz DA. 2013. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* **68**:1114–1121 DOI 10.1136/thoraxjnl-2012-202943.

Yasukawa H, Ohishi M, Mori H, Murakami M, Chinen T, Aki D, Hanada T, Takeda K, Akira S, Hoshijima M, Hirano T, Chien KR, Yoshimura A. 2003. IL-6 induces an anti-inflammatory response in the absence of SOCS3 in macrophages. *Nature Immunology* **4**:551–556 DOI 10.1038/ni938.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology* **16**:284–287 DOI 10.1089/omi.2011.0118.

Zeleznik OA, Eliassen AH, Kraft P, Poole EM, Rosner BA, Jeanfavre S, Deik AA, Bullock K, Hitchcock DS, Avila-Pacheco J, Clish CB, Tworoger SS. 2020. A prospective analysis of circulating plasma metabolites associated with ovarian cancer risk. *Cancer Research* **80**:1357–1367 DOI 10.1158/0008-5472.CAN-19-2567.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4**: Article17.

Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. 2010. Weighted gene coexpression network analysis: state of the art. *Journal of Biopharmaceutical Statistics* **20**:281–300 DOI 10.1080/10543400903572753.

Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, Lollini L, Morris D, Kim Y, DeLustro B, Sheppard D, Pardo A, Selman M, Heller RA. 2002. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and

humans. *Proceedings of the National Academy of Sciences of the United States of America* **99**:6292–6297 DOI 10.1073/pnas.092134099.