Weighted Gene Coexpression Network Analysis Identifies Key Genes and Pathways Associated with Idiopathic Pulmonary Fibrosis

Zheng Wang*  
Jie Zhu*  
Fengzhe Chen  
Lixian Ma

* Zheng Wang and Jie Zhu are co-first authors

Corresponding Authors:
Lixian Ma, e-mail: mlx_sdu@163.com, Fengzhe Chen, e-mail: 18560082133@163.com

Source of support:
Departmental sources

Background:
Idiopathic pulmonary fibrosis (IPF) is a life-threatening disease with an unknown etiology. Gene expression microarray data have provided some insights into the molecular mechanisms of IPF. This study aimed to identify key genes and significant signaling pathways involved in IPF using bioinformatics analysis.

Material/Methods:
Differentially expressed genes (DEGs) were identified using integrated analysis of gene expression data with a robust rank aggregation (RRA) method. The Connectivity Map (CMAP) was used to identify gene-expression signatures associated with IPF. Weighted gene coexpression network analysis (WGCNA) was used to explore the functional modules involved in the pathogenesis of IPF.

Results:
A total of 191 patients with IPF and 101 normal controls from six genome-wide expression datasets were included. CMAP predicted several small molecular agents as potential gene targets in IPF. Several functional modules were detected that showed the highest correlation with IPF, including an extracellular matrix (ECM) component, and a myeloid leukocyte migration and activation component involved in the immune response. Hub genes were identified in the key functional modules that might have a role in the progression of IPF.

Conclusions:
WGCNA was used to identify functional modules and hub genes involved in the pathogenesis of IPF.

MeSH Keywords:
Biomarkers, Pharmacological • Gene Regulatory Networks • Idiopathic Pulmonary Fibrosis

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/916828
Background

Idiopathic pulmonary fibrosis (IPF) is a devastating illness characterized by irreversible lung fibrosis [1]. Although the overall prevalence of IPF is not high, the incidence of the disease has recently increased. In Europe and North America, the annual incidence of IPF is estimated to be between 2.8 and 18 cases per 100,000 individuals [2]. The median age of IPF is about 65 years, and men have a higher incidence [3,4]. Wound healing results in fibrosis and is believed to be the basis for the pathogenesis of IPF and includes the stages of homeostasis, inflammation, cell migration, cell proliferation, and extracellular matrix (ECM) remodeling. IPF may be due to chronic injury of the alveolar epithelium that results in pulmonary fibrosis and structural lung remodeling [5]. However, the etiology of IPF remains unknown.

Currently, IPF is considered to be a result of the interaction between genetic and environmental risk factors [6], and aging might influence the susceptibility to lung fibrosis, as the incidence of IPF increases with age [7]. Also, genome-wide association studies have shown that some genes related to host defense and epithelial barrier function may also be involved in the pathogenesis of IPF [8]. Among these genes, a variant of the MUC5B promoter region was shown to be involved in the development of IPF [9]. In terms of environmental factors, cigarette smoking has been proposed to be the most common association with IPF [10]. Other inhaled environmental agents include exposure to metal dust or wood dust, sand, and spores from soil [4,11]. IPF remains a challenging disease to treat, and further studies are needed to improve the understanding of the underlying molecular mechanisms to identify gene targets for the development of novel therapies.

There are a large number of microarray gene expression datasets that are publicly available from the Gene Expression Omnibus (GEO) database. There is an increasing demand to integrate gene expression datasets to obtain more accurate results [12]. There have been no previous studies that have undertaken comprehensive bioinformatics analysis in IPF. Therefore, this study aimed to perform a comprehensive approach by using a robust rank aggregation (RRA) method to identify the differentially expressed genes (DEGs) from several datasets [13]. The connectivity map (CMAP) (https://portals.broadinstitute.org/cmap/) transcriptional expression database was chosen to identify gene-expression signatures associated with IPF, as CMAP represents a valuable tool for establishing the connections between genes, drugs, and diseases and contains over 7000 gene expression profiles reflecting 1309 bioactive compounds [14]. CMAP can be used to identify mechanisms of action of small molecules and identify novel therapeutic targets [15]. The signature of differentially expressed genes (DEGs) can be used to input into the CMAP. Also, weighted gene coexpression network analysis (WGCNA) is a powerful approach for exploring the complex relationships between gene expression profiles and phenotype [16]. Therefore, in the present study, WGCNA was used to build a gene coexpression network and screen important modules in the network, and to filter the hub genes in the essential modules.

Therefore, this study aimed to identify key genes and significant signaling pathways involved in IPF using bioinformatics analysis. DEGs from lung tissue microarray data were identified from patients with IPF and normal controls, and potential gene targets for the treatment of IPF were detected using the CMAP database. WGCNA was used to construct a coexpression network associated with IPF to identify significant modules and hub genes, that may be related to the pathogenesis of IPF.

Material and Methods

Datasets used

The gene expression datasets from idiopathic pulmonary fibrosis (IPF) were downloaded from the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/). The GEO represents the largest resource of public microarray data and is widely used to identify key genes in disease. In this study, there were several selection criteria for data selection that included: (a) gene expression datasets which contained gene chip microarrays; (b) studies comparing gene expressions between patients with IPF and normal control lung tissue; (c) sample size in each chip dataset contained at least ten samples; (d) raw data or processed data were available in these databases; (e) hypersensitive pneumonitis, cryptogenic organizing pneumonia (COP), or respiratory bronchiolitis-interstitial lung disease (RBILD) were not included in this study. Searches were excluded if they did not meet the inclusion criteria.

All patients with IPF included in this study met the diagnostic criteria for IPF based on the American Thoracic Society (ATS) and European Respiratory Society (ERS) consensus statement. Lung tissue samples from the normal control groups were obtained from patients without IPF and included lung cancer patients and lung transplant patients. Ethical approval for this study was not required because the data were downloaded directly from public datasets.

Integrated analysis of gene expression datasets

Microarray data preprocessing were executed using R Project software and Bioconductor packages [17]. The raw microarray data were converted to expression data using the robust multiarray average (RMA) algorithm based on R language [18]. Also, two preprocessed series matrix files were directly downloaded.
The WGCNA package was used to construct the coexpression network. GSE32537 which including 119 patients with IPF was analyzed by WGCNA. We calculated the correlation between module eigengenes and clinical features. Finally, the modules that were highly correlated with clinical function were chosen for further analysis.

**Identification and validation of hub genes in key modules**

Module membership (MM) represented the correlation of gene expression profile with the module eigengene. The top 30 genes with the highest connectivity in key modules were regarded as hub genes and were visualized with Cytoscape. The expression statuses of hub genes in key modules were identified for validation. Venn diagram was plotted using the online JVenn tool to overlap the hub genes in significant modules and DEGs obtained from the RRA analysis.

**Functional enrichment analysis**

Genes in interest modules were uploaded to the online bioinformatics database Metascape and underwent Gene Ontology (GO) biological process, cellular component, and molecular function analysis as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The number of enriched genes ≥3 and P<0.01 were regarded as the cutoff criteria. Only the top ten enriched pathways were extracted.

**Results**

**Microarray datasets for idiopathic pulmonary fibrosis (IPF)**

After selection according to the eligibility criteria, six datasets were incorporated into the study: GSE10667, GSE15197, GSE21369, GSE24206, GSE32537, GSE110147 (Table 2). The main characteristics of these publicly available databases including gender ratio, and age distribution, origin, number of samples, GSE number, platforms, and source types are listed in Table 2. The number of patients with IPF ranged from 8 to 119, and the number of normal controls ranged from 6 to 50. In total, 191 patients with IPF and 101 normal controls were enrolled in the study.

**Robust rank aggregation (RRA) analysis**

By using the RRA approach to integrate six microarray datasets, a total of 368 differentially expressed genes (DEGs) comprising 248 upregulated and 120 downregulated genes were obtained (Supplementary Table 1). The top 25 upregulated and randomly color-labeled. Genes that could not be clustered into any given module were assigned to the grey module. The co-expression networks were visualized by Cytoscape software version: 3.6.1 (http://www.cytoscape.org) [20]. We identified modules that were significantly associated with IPF by calculated the correlation between module eigengenes and clinical features.

**Table 1. Results of CMap analysis.**

| cmap name | Mean  | n  | p     |
|-----------|-------|----|-------|
| Pregnenolone | –0.524 | 4  | 0.00022 |
| Lycorine    | –0.425 | 5  | 0.00555 |
| Chloropyrazine | –0.347 | 4  | 0.00573 |
| Securinine  | –0.661 | 4  | 0.01215 |
| Rotenone    | –0.477 | 4  | 0.02043 |
| Indoprofen  | –0.326 | 4  | 0.02449 |
| Megestrol   | –0.348 | 4  | 0.03463 |
| Terazosin   | –0.328 | 4  | 0.03766 |

* The compounds tested in at least four experiments were ranked according to p value.

from the GEO repository. Probes were mapped to gene symbols. The average gene expression levels were calculated for genes represented by more than one probe. The differentially expressed genes (DEGs) between patients with IPF and normal controls in each dataset were screened by the Limma package in R [19]. Integrated analysis for the DEGs obtained from the six eligible datasets was performed using the R package (Robust Rank Aggregation), based on a robust rank aggregation (RRA) method. Also, |logFC| ≥1 and P<0.05 were used as the cutoff criteria for screening the DEGs.

**Screening of small molecules**

To identify candidate small molecules, we compared the DEGs with the 1,309 different compounds in the connectivity map (CMap) used to identify gene-expression signatures associated with IPF. The DEGs were input to CMap software for analysis, according to the website instructions. Small molecules with negative connectivity scores, representing the compounds that might reverse the input signature, were identified as therapeutic agents. The small molecules with average scores <–0.3 and P<0.05 were selected (Table 1).

**Weighted gene coexpression network analysis (WGCNA)**

The weighted gene coexpression network (WGCNA) analysis was used to identify hub genes and pathways in IPF. The modules were ranked according to p value. The top 30 genes with the highest connectivity in key modules were regarded as hub genes and were visualized with Cytoscape. The expression statuses of hub genes in key modules were identified for validation. Venn diagram was plotted using the online JVenn tool to overlap the hub genes in significant modules and DEGs obtained from the RRA analysis.
downregulated genes are shown in Figure 1. Among these DEGs, the roles of some genes in IPF have been validated in previous studies, including KRT5, BPIFB1, and AGER.

**Screening for small molecules**

The connectivity map (CMAP) database was used to search for small molecules with therapeutic potential in IPF. The small molecules with a high negative connectivity score are presented in Table 1. Among these potential therapeutic agents, pregnenolone and lycorine showed a smaller P-value and were potential therapeutic targets for IPF.

**Weighted gene coexpression network analysis (WGCNA)**

According to the order of the coefficient of variation, the 5,000 most variable genes were chosen for WGCNA. Hierarchical clustering analysis was performed, and the results are shown in Supplementary Figure 1. There were three outlier samples in the 169 samples that included GSM806290, GSM806408, GSM806411 when the threshold was set as 55. Outlier samples were removed from the cohort before further analysis. As shown in Figure 2A, power=5 was chosen as the soft-thresholding to ensure a scale-free network. We set the minimum module size to 100 genes and the minimum cut height for merging of modules at 0.3. The weighted gene coexpression network analysis (WGCNA) identified eight modules ranging in size from 228 to 1808.

There were 45 genes that did not belong to any module were labeled in grey, which were excluded from further analysis. These coexpression modules were constructed and were presented in different colors (Figure 2B). Interactions between the eight coexpression modules were then analyzed (Figure 2C). To explore the clinical significance of the module, correlations between module eigengenes and clinic traits were analyzed. As shown in Figure 3, black, blue, magenta, and pink modules were positively correlated with two clinical traits, namely disease status and the St George’s score for the severity of IPF. By contrast, yellow, brown, and red modules were found to be negatively associated with disease status and St George’s score for severity of IPF.

Also, we found some positive correlations between the black module and smoking pack years. Combined with Figure 4, we observed that these seven modules yielded two main clusters; one included four modules (black, blue, magenta and pink module) while the other included three modules (yellow, brown and red). Furthermore, two pairs of modules had higher adjacencies, and they were the black and magenta module, yellow and brown module respectively. Also, the module eigengene of the black and yellow module showed a higher correlation with disease status (Figure 3). The black module had the strongest positive correlation with IPF (r=0.79; P=4e-37) while the yellow module had the strongest negative correlation with IPF (r=-0.81; P=3e-40). We plot a scatterplot for the correlation between module membership and gene significance in the two key modules, respectively (Figure 5). The coexpression networks were visualized with Cytoscape and are shown in Supplementary Figure 2.

**Identification and validation of hub genes in the key modules**

Hub genes may determine the characteristics of a module and play significant roles in biological processes. Consequently, the top 30 genes with the highest degree of connectivity in the black and yellow module were taken as hub genes, including COL14A1, TSHZ2, IL1R2, and SLCO4A1. Figure 6A and 6B showed the top 30 genes with the highest degree of connectivity in the black and yellow module respectively. Also, the module eigengene of the black and yellow module showed a higher correlation with IPF.

| Dataset ID | GSE number | Samples | IPF age(years) | IPF Sex (M/F) | Source types | Platform | Authors |
|------------|------------|---------|----------------|---------------|--------------|----------|---------|
| 1          | GSE10667   | 23 IPF samples and 15 controls | 61.7±5.51 | 19/4 | Lung tissues | GPL4133 | Konishi K, Richards TJ, Kaminski N |
| 2          | GSE15197   | 8 IPF samples and 13 controls | 60±5 | 5/3 | Lung tissues | GPL6480 | Rajkumar R, Konishi K, Richards TJ et al. |
| 3          | GSE21369   | 11 IPF samples and 6 controls | 58.0±9.48 | 8/3 | Lung tissue | GPL570 | Cho JH, Gelinas R, Wang K et al. |
| 4          | GSE24206   | 8 IPF samples and 6 controls | 61.2±6.32 | 7/1 | Lung tissue | GPL570 | Meltzer EB, Barry WT, D’Amico TA et al. |
| 5          | GSE32537   | 119 IPF samples and 50 controls | 62.5±8.75 | 77/42 | Lung tissue | GPL6244 | Yang IV, Coldren CD, Leach SM et al. |
| 6          | GSE110147  | 22 IPF samples and 11 controls | 62±6 | 17/5 | Lung tissue | GPL6244 | Cecchini MJ, Hosein K, Howlett CJ et al. |

Table 2. Summary of those 6 genome-wide gene expression datasets involving IPF patients.

---

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)
Figure 1. The top 25 upregulated genes and top 25 down-regulated genes in idiopathic pulmonary fibrosis (IPF). Each column represents one dataset, and each row represents one gene. The numbers in each rectangle show the logarithmic fold-change of genes in each dataset. Red indicates increased gene expression, and green indicates decreased gene expression.
**Figure 2.** Plots in the weighted gene coexpression network analysis (WGCNA) using gene expression data from 119 patients with idiopathic pulmonary fibrosis (IPF) and 50 controls from GSE32537 datasets. (A) Network topology of different soft-thresholding powers. The left panel shows the influence of the soft-threshold power on the scale-free topology fit index. The right panel shows the influence of the soft-threshold power on the mean connectivity. (B) Cluster dendrogram of coexpression genes and functional modules in IPF. Eight coexpression modules were constructed and are shown in different colors. (C) A heatmap plot shows the gene network. Different colors of the horizontal axis and the vertical axis represent different modules. The light color indicates lower overlap, and the dark red indicates higher overlap.

**Figure 3.** Heatmap of the correlation between module eigengenes and clinical traits of idiopathic pulmonary fibrosis (IPF). The table is color-coded by correlation according to the color legend. Each cell contains the corresponding correlation and p-value.
We used DEGs generated from RRA analysis to validate the expression status of the top two hub genes in the black and yellow module. We overlapped the DEGs and genes in the black and yellow module by plotting a Venn diagram. These four hub genes were present in DEGs and significant modules, indicating their value as potential biomarkers for IPF. The results are presented in Figure 7.

**Functional enrichment analysis**

Functional enrichment analysis was performed for the genes in the constructed seven modules. The genes in the black module were mainly enriched in extracellular matrix (ECM) organization, skeletal development, and vasculature development (Figure 8A). Genes in the yellow module were enriched in myeloid leukocyte migration, leukocyte activation, and involved in the immune response and the inflammatory response (Figure 8B). Genes in the brown module were mainly enriched in blood vessel development, cell junction organization, and sterol biosynthetic. Genes in the blue module were mainly involved in cilia, motile cilia, and O-glycan processing. Multiple signaling pathways were found to be involved in other modules, including nuclear division in the magenta module, T cell activation in the pink module, and olfactory transduction in the red module. The main results in Gene Ontology (GO)
Figure 6. Network diagram of the top 30 genes in idiopathic pulmonary fibrosis (IPF). (A) The network of top 30 genes in the dark magenta module. (B) The network of top 30 genes in the yellow module. Node size: larger size indicates a higher degree of connectivity, and a smaller size indicates a lower degree of connectivity. Node color: Red indicates an upregulated gene. Green indicates a down-regulated gene.

Figure 7. The overlap of differentially expressed genes (DEGs) and hub genes shown as a Venn diagram. (A) Identification of common genes between differentially expressed genes (DEGs) and the black module overlapping them. The two hub genes in the black module were also DEGs obtained from the robust rank aggregation (RRA) analysis. (B) Identification of common genes between DEGs and the yellow module overlapping them. The two hub genes in the yellow module were also DEGs obtained from the RRA analysis.
Idiopathic pulmonary fibrosis (IPF) is a progressive and devastating disease that usually leads to death within five years after diagnosis [29]. The pathogenesis of the disease remains poorly understood. In the current study, we conducted an integrated analysis of gene expression data to explore the molecular pathogenesis of IPF. In this study, the robust rank aggregation (RRA) method, was used to scan significant differentially expressed genes (DEGs) from six independent gene expression datasets. A large number of DEGs were identified from which 248 DEGs were upregulated and 120 DEGs were down-regulated in IPF patient samples compared with control samples. IPF is the result of chronic inflammation and healing in the lung. This study identified some DEGs that were closely associated with inflammation and healing. It was previously reported that KRT5 was highly expressed in the alveolar regions of the lung in IPF [30]. Also, airway stem cells have been shown to express KRT5 in lung injury [31].

Figure 8. Functional and pathway enrichment analysis. (A) Functional and pathway enrichment analysis of the black module. (B) Functional and pathway enrichment analysis of the yellow module. Heatmap of top 20 clusters, colored by P-values. Each bar represents a cluster. The darker the color of the bar is, the smaller the P-value.

Discussion

Idiopathic pulmonary fibrosis (IPF) is a progressive and devastating disease that usually leads to death within five years after diagnosis [29]. The pathogenesis of the disease remains poorly understood. In the current study, we conducted an integrated analysis of gene expression data to explore the molecular
these previous studies support that KRT5 is associated with the process of healing in the alveolar epithelium. BPIFB1 has also been previously shown to be upregulated in respiratory disease, including in chronic obstructive pulmonary disease (COPD) where it has been associated with disease severity [32]. Also, an ulcer-associated cell lineage plays an important role in the healing process in inflammatory bowel disease, and BPIFB1 is one of the ulcer-associated cell lineage genes [33]. Therefore, it is possible that BPIFB1 plays a significant role in the healing process in IPF.

PLA2G1B, a secreted phospholipase, was among the most downregulated genes in IPF in the present study. There have been few previous studies on PLA2G1B in IPF, and so the aberrant expression of this gene should be validated by future studies. AGER, or advanced glycosylation end product-specific receptor, is one of the DEGs that were significantly down-regulated in the present study. Advanced glycosylation end product (AGE) has been previously shown to inhibit the wound-healing in diabetes by reducing the activity of epidermal stem cells [34]. However, blocking the expression of AGER may facilitate the healing process, which may explain the possible role for AEGR in IPF.

Some candidate molecules with therapeutic potential in IPF were explored based on the use of the connectivity map (CMAP) database to identify gene expression signatures. Pregnenolone is an inactive precursor of steroid hormones, and its potential functional effects have been previously studied [35]. Pregnenolone therapy has shown beneficial effects in schizophrenia [36], bipolar depression [37], and drug dependence [38]. Prospective studies are needed to investigate pregnenolone as a candidate therapeutic compound in IPF. Lycorine is an alkaloid derived from Hymenocallis littoralis, which has antibacterial, antiviral, and wound-healing properties [39]. The findings from a previous study suggested that lycorine might have therapeutic potential in patients with IPF [40]. This previous finding was consistent with the findings from the present study, and lycorine is a molecule that requires further study.

The method of WGCNA that was the basis of this study is an efficient approach to construct coexpressed modules and hub genes in several diseases [41]. In this study, we investigated the gene expression profile of GSE32537, including 119 patients with IPF and 50 normal controls to explore the molecular mechanism of IPF. Using WGCNA, the weighted coexpression network was constructed, and eight coexpression modules were identified. Seven of the modules showed a significant correlation with disease status, which were divided into two clusters, providing new insights into the pathogenesis of IPF. Among the identified modules, the black, blue, magenta, and pink modules were positively associated with disease status, whereas the yellow, brown, and red modules were negatively correlated with disease status. To determine the importance of these coexpression modules in the pathogenesis of IPF, enrichment analysis was performed using the Metascape online bioinformatics gene annotation and analysis database. The black module was found to be mainly enriched in collagen-containing extracellular matrix (ECM) and the ECM component. Increased accumulation of ECM is an important phase of the wound-healing process. ECM has been identified as an active contributor to the progression of fibrosis [42]. A therapeutic approach that targets the ECM may reverse fibrosis and restore lung function. The results of this study showed that the yellow module was mainly enriched in myeloid leukocyte migration and leukocyte activation involved in the immune response, and in cell migration. These results were consistent with previous studies that immune signaling pathways may be protective in IPF [43].

The black and yellow modules were found to be most significantly correlated with the disease status of IPF. Therefore, we screened the hub genes and validated the top two hub genes in the two modules. These four hub genes were all present in DEGs and significant modules, indicating their important roles in biological processes. COL14A1 is a member of the fibril-associated collagens with interrupted triple helices (FACIT) collagen family, with the main function of collagen binding and ECM components [44]. COL14A1 expression has been reported to be significantly upregulated in the pulmonary artery intima and media of patients with pulmonary arterial hypertension [45]. It is likely that COL14A1 plays a key role in the process of pulmonary interstitial fibrosis.

In this study, TSHZ2, a zinc-finger homeobox gene, was identified as a hub gene. Studies on the role of TSHZ2 are limited, and future studies are required to confirm its function in IPF. IL1R2 is a member of the interleukin-1 receptor family and acts as a negative regulator of the IL-1 system. The anti-inflammatory effect of IL1R2 has been confirmed by in vivo studies, and include chronic skin inflammation [46], and arthritis [47]. The combination of acute inflammation and healing due to persistent injury results in fibrosis [48]. In the present study, IL1R2 was both a hub gene and a significantly down-regulated gene in the IPF group, suggesting that the overactive wound healing was most likely associated with the functional deletion of IL1R2. SLCO4A1 is a member of solute carrier organic anion transporter family, which is involved in the transport of a variety of compounds [49]. There have been few studies on SLCO4A. However, SLCO4A1 has been shown to have a role in the proliferation of colorectal cancer cell and cell migration in vitro and to be a negative prognostic marker in vivo [50]. Further studies are required to determine the role of SLCO4A1 in IPF.

The key genes identified in this study were associated with the pathogenesis of fibrosis, including in IPF, and the findings
were supported by those from previously published studies. However, this study had several limitations. Only six datasets of microarray data for IPF were used. Few clinical and regional characteristics of the patients were available in each dataset, and there may have been relevant clinical differences that affected the findings. Also, the results obtained by bioinformatics analysis alone should be verified by \textit{in vitro} and clinical studies, and so further studies are needed to validate these preliminary findings.

\section*{Conclusions}

Through the comprehensive bioinformatics analysis of several gene expression datasets, we identified differentially expressed genes (DEGs) that might be related to the pathogenesis, diagnosis, and prognosis of idiopathic pulmonary fibrosis (IPF). Also, relevant small molecules were predicted using the connectivity map (CMAP) database to identify gene expression signatures associated with IPF, which might be potential candidate therapeutic compounds for IPF. We used weighted gene coexpression network analysis (WGCNA) to build a coexpression network associated with IPF and found two modules and four hub genes, which might identify novel insights into the molecular mechanisms underlying IPF. The findings of this study should be confirmed by future \textit{in vitro} and clinical studies.

\section*{Conflict of interest}

None.

\section*{Supplementary Files}

\subsection*{Supplementary Table 1. Genes statistically differentially expressed between IPF and normal lung tissue.}

| Name  | Pvalue  | adjPvalue | logFC   |
|-------|---------|-----------|---------|
| KRT5  | 8.16E-19| 2.28E-14  | 2.935163698 |
| BPIFB1| 1.66E-17| 4.63E-13  | 2.816100513 |
| MMP1  | 7.1E-14 | 4.77E-10  | 2.975792266 |
| MMP7  | 1.34E-13| 3.74E-09  | 2.778138388 |
| GPR87 | 7.23E-13| 2.02E-08  | 2.297678786 |
| ZBBX  | 2.63E-12| 7.32E-08  | 2.091789774 |
| COL1A1| 3.10E-12| 8.63E-08  | 2.004344368 |
| TMEM45A| 3.76E-12| 1.05E-07  | 1.945792282 |
| IL13RA2| 5.28E-12| 1.47E-07  | 1.927518882 |
| MSMB  | 5.52E-12| 1.54E-07  | 2.086340891 |
| CXCL13| 5.89E-12| 1.64E-07  | 1.873998487 |
| LC2A2 | 7.64E-12| 2.03E-07  | 1.753165017 |
| PROM1 | 9.11E-12| 2.54E-07  | 2.236427095 |
| KRT15 | 9.44E-12| 2.63E-07  | 2.084821277 |
| CHST9 | 2.38E-11| 6.65E-07  | 1.946570653 |
| CP    | 3.27E-11| 9.12E-07  | 2.161658822 |
| TMPRSS4| 5.71E-11| 1.59E-06  | 1.929477651 |
| SFRP2 | 1.05E-10| 2.93E-06  | 2.299349319 |
| SPP1  | 1.97E-10| 5.49E-06  | 2.416271247 |
| COL1A1| 2.32E-10| 6.48E-06  | 1.865800024 |
| CNTN3 | 2.83E-10| 7.88E-06  | 1.429494649 |
| COL14A1| 3.89E-10| 1.09E-05  | 1.676170261 |
| CASC1 | 5.95E-10| 1.66E-05  | 1.506540304 |
| DIO2  | 6.13E-10| 1.71E-05  | 1.458989173 |
| ARM3  | 6.62E-10| 1.85E-05  | 1.864748762 |
| SULF1 | 8.52E-10| 2.38E-05  | 1.407674611 |
| SERPINB3| 8.76E-10| 2.44E-05  | 1.822592533 |
| KLHL13| 8.85E-10| 2.47E-05  | 1.415638864 |
| CXCL14| 9.30E-10| 2.39E-05  | 1.831177737 |
| C20orf85| 9.97E-10| 2.78E-05  | 1.801814831 |
| SPAG17| 1.04E-09| 2.91E-05  | 1.646375125 |
| LRR17 | 1.08E-09| 2.05E-05  | 1.392596034 |
| SLC27A2| 1.14E-09| 3.19E-05  | 1.556826086 |
| POSTN | 1.17E-09| 3.27E-05  | 1.853027833 |
| SNTN  | 1.42E-09| 3.97E-05  | 1.875956956 |
| KRT17 | 1.46E-09| 4.07E-05  | 1.959920244 |
| CDH3  | 1.68E-09| 4.68E-05  | 1.930510747 |
| FANK1 | 1.95E-09| 5.43E-05  | 1.503994499 |
| MMP13 | 2.09E-09| 5.84E-05  | 1.757375715 |
| CCDC146| 2.34E-09| 6.53E-05  | 1.492462573 |
| RSPH1 | 3.52E-09| 9.82E-05  | 1.724507872 |
| FAM81B| 3.76E-09| 0.000104985| 1.776467268 |
| Name    | Pvalue   | adjPvalue | logFC   |
|---------|----------|-----------|---------|
| VTCN1   | 4.32E-09 | 0.000120533 | 1.683907215 |
| COMP    | 5.67E-09 | 0.000158194  | 2.319702189  |
| RGS22   | 5.86E-09 | 0.000163423  | 1.446072149  |
| MUC16   | 6.35E-09 | 0.000177097  | 1.571239322  |
| SPATA18 | 6.42E-09 | 0.000179133  | 1.394998539  |
| TSHZ2   | 7.05E-09 | 0.0001996678 | 1.118911148  |
| CAPSL   | 7.48E-09 | 0.000208869  | 1.638856212  |
| TTC25   | 7.63E-09 | 0.000212835  | 1.30957343   |
| COL1A2  | 8.27E-09 | 0.000230588  | 1.502829774  |
| SFRP4   | 9.37E-09 | 0.000261414  | 1.230284362  |
| DCLK1   | 9.88E-09 | 0.000275575  | 1.272063729  |
| CYP2F1  | 1.02E-08 | 0.000283711  | 1.223127058  |
| WDR78   | 1.22E-08 | 0.000338888  | 1.267174919  |
| PSD3    | 1.32E-08 | 0.000368616  | 1.140913494  |
| WDR63   | 1.49E-08 | 0.000414831  | 1.401986713  |
| LRP1    | 1.58E-08 | 0.000490297  | 1.318716297  |
| CFH     | 1.60E-08 | 0.000447383  | 1.387921751  |
| COL6A3  | 1.67E-08 | 0.000466154  | 1.120968356  |
| NELL2   | 2.06E-08 | 0.000573796  | 1.111579211  |
| CCDC113 | 2.21E-08 | 0.000616382  | 1.227081048  |
| COL5A2  | 2.50E-08 | 0.000793675  | 1.739200141  |
| ST6GALNAC1 | 3.03E-08 | 0.000843723  | 1.39092015  |
| RPS4Y1  | 3.22E-08 | 0.000894867  | 1.966604231  |
| SPAG6   | 3.51E-08 | 0.000979708  | 1.755246732  |
| TTC29   | 3.73E-08 | 0.001040815  | 1.400611802  |
| SCGB1A1 | 3.96E-08 | 0.001105576  | 1.472963895  |
| MMP10   | 4.06E-08 | 0.001132187  | 1.850435932  |
| SPATA17 | 4.39E-08 | 0.001225456  | 1.386846091  |
| SLN     | 4.48E-08 | 0.001248606  | 1.539423446  |
| S100A2  | 4.48E-08 | 0.001250099  | 2.029971633  |
| DSC1    | 4.53E-08 | 0.001262111  | 1.633156066  |
| TSPAN1  | 5.09E-08 | 0.001418504  | 1.501762841  |
| AGBL2   | 5.14E-08 | 0.001433455  | 1.265354573  |
| DYNLRB2 | 6.25E-08 | 0.001742309  | 1.426611109  |
| LTF     | 6.70E-08 | 0.001868646  | 1.362389488  |
| MYH11   | 6.96E-08 | 0.001941834  | 1.020881259  |
| Name          | Pvalue   | adjPvalue | logFC       |
|--------------|----------|-----------|-------------|
| C11orf88     | 6.13E-07 | 0.017103725 | 1.484628508 |
| COL5A2       | 6.29E-07 | 0.017538773 | 1.183016083 |
| EYA2         | 6.29E-07 | 0.017543232 | 1.126653922 |
| SERPIN12     | 6.29E-07 | 0.017543232 | 1.126653922 |
| MEOX1        | 6.39E-07 | 0.017834057 | 1.529655688 |
| CTHRC1       | 6.78E-07 | 0.018915173 | 1.529655688 |
| MUC4         | 6.92E-07 | 0.019312725 | 1.084574554 |
| MORN5        | 7.01E-07 | 0.01955195  | 1.494365253 |
| TNFRSF17     | 7.30E-07 | 0.021371412 | 1.013826148 |
| SCG5         | 7.31E-07 | 0.023481583 | 1.604114371 |
| SERPIN1D     | 8.50E-07 | 0.025033914 | 1.380147985 |
| STK33        | 8.98E-07 | 0.025094148 | 1.255674554 |
| ENKUR        | 9.00E-07 | 0.025094148 | 1.010033179 |
| BCHE         | 9.00E-07 | 0.025094148 | 1.010033179 |
| NEK10        | 9.52E-07 | 0.025094148 | 1.010033179 |
| CTHRC1       | 9.74E-07 | 0.025094148 | 1.010033179 |
| LRRQ1        | 9.93E-07 | 0.025094148 | 1.010033179 |
| ALDH3A1      | 1.03E-06 | 0.025094148 | 1.010033179 |
| UBXN10       | 1.05E-06 | 0.025094148 | 1.010033179 |
| GSTA1        | 1.08E-06 | 0.025094148 | 1.010033179 |
| CCL18        | 1.09E-06 | 0.025094148 | 1.010033179 |
| CTSE         | 1.25E-06 | 0.025094148 | 1.010033179 |
| COL10A1      | 1.37E-06 | 0.025094148 | 1.010033179 |
| CTSK         | 1.38E-06 | 0.025094148 | 1.010033179 |
| TRIM29       | 1.46E-06 | 0.025094148 | 1.010033179 |
| TRIM5        | 1.62E-06 | 0.025094148 | 1.010033179 |
| GOLM1        | 1.64E-06 | 0.025094148 | 1.010033179 |
| TEKT2        | 1.65E-06 | 0.025094148 | 1.010033179 |
| C11orf70     | 1.77E-06 | 0.025094148 | 1.010033179 |
| SOX2         | 1.89E-06 | 0.025094148 | 1.010033179 |
| AMPD1        | 2.03E-06 | 0.025094148 | 1.010033179 |
| CYP24A1      | 2.07E-06 | 0.025094148 | 1.010033179 |
| CD24         | 2.21E-06 | 0.025094148 | 1.010033179 |
| WDR66        | 2.23E-06 | 0.025094148 | 1.010033179 |
| ANKFN1       | 2.28E-06 | 0.025094148 | 1.010033179 |
| H56ST2       | 2.46E-06 | 0.025094148 | 1.010033179 |
| Name       | Pvalue | adjPvalue | logFC |
|------------|--------|-----------|-------|
| VWA3B      | 1.09E-05 | 0.304184254 | 1.013600302 |
| FAM216B    | 1.22E-05 | 0.339444516 | 1.003245966 |
| ZMYND10    | 1.25E-05 | 0.349552078 | 1.069228349 |
| DNM2       | 2.50E-06 | 0.960392420 | 1.205967939 |
| ATP12A     | 1.26E-05 | 0.352113661 | 1.241793927 |
| CDHR3      | 1.31E-05 | 0.366149094 | 1.003213287 |
| C1orf87    | 1.33E-05 | 0.370438911 | 1.240668092 |
| FCR5       | 1.46E-05 | 0.407936307 | 1.074522645 |
| IQUB       | 1.55E-05 | 0.432687514 | 1.273083757 |
| DLEC1      | 1.56E-05 | 0.435688123 | 1.127617749 |
| TP63       | 1.79E-05 | 0.498074147 | 1.022168715 |
| DN2K       | 1.87E-05 | 0.521842029 | 1.006616734 |
| TDO2       | 1.92E-05 | 0.536740731 | 1.017985205 |
| CCDC81     | 1.96E-05 | 0.547864087 | 1.111704424 |
| DTHD1      | 2.21E-05 | 0.616533406 | 1.220706070 |
| CCNA1      | 2.22E-05 | 0.618380293 | 1.109596078 |
| DNAH15     | 2.50E-05 | 0.986199130 | 1.003600727 |
| CCDC39     | 2.73E-05 | 0.762305285 | 1.064688193 |
| SIX1       | 2.80E-05 | 0.781215901 | 1.166844918 |
| TMEM232    | 2.92E-05 | 0.815239855 | 1.277488281 |
| DSG3       | 2.98E-05 | 0.830395002 | 1.004793382 |
| UCP1A6     | 3.08E-05 | 0.995790880 | 1.003913111 |
| DNAH3      | 3.37E-05 | 0.904597181 | 1.020961433 |
| AK7        | 3.47E-05 | 0.967454427 | 1.025283586 |
| CLIC6      | 3.98E-05 | 1.050256203 |       |
| SERPINB4   | 4.48E-05 | 1.286082732 |       |
| DNAH7      | 5.29E-05 | 1.078210310 |       |
| DNAH5      | 5.93E-05 | 1.025356953 |       |
| CCDC78     | 6.17E-05 | 1.001758647 |       |
| CAPN13     | 6.26E-05 | 1.075410711 |       |
| DNA1       | 7.23E-05 | 1.169928958 |       |
| RIBC2      | 7.97E-05 | 1.006702152 |       |
| KLK12      | 8.41E-05 | 1.124257565 |       |
| CRLF1      | 8.88E-05 | 1.115088178 |       |
| SRDS5A2    | 9.56E-05 | 1.075894376 |       |
| CLDN8      | 9.74E-05 | 1.130318712 |       |
| C1orf107   | 0.000109055 | 1 | 1.060098777 |
| STOX1      | 0.000109344 | 1 | 1.109895823 |
| C1orf81    | 0.000118127 | 1 | 1.119302483 |
| DDX3Y      | 0.000140361 | 1 | 1.049553425 |
| FAM183A    | 0.000153129 | 1 | 1.040469983 |
| CILP       | 0.000162266 | 1 | 1.173032082 |
| RSPH4A     | 0.000198839 | 1 | 1.091928915 |
| USP9Y      | 0.000203297 | 1 | 1.007151508 |
| EYA1       | 0.000317943 | 1 | 1.10128004 |
| MS4A8      | 0.000323399 | 1 | 1.005164192 |
| PTPLN      | 0.000360859 | 1 | 1.169300297 |
| GSTA5      | 0.000373902 | 1 | 1.000889231 |
| FHAD1      | 0.000393343 | 1 | 1.047998957 |
| IGF2       | 0.000465987 | 1 | 1.312853186 |
| RHOC       | 0.000475109 | 1 | 1.033813468 |
| LDLRA1     | 0.000530851 | 1 | 1.027099125 |
| PLA2G2A    | 0.000583188 | 1 | 1.096318802 |
| DNAAF1     | 0.000907879 | 1 | 1.062363628 |
| FAM92B     | 0.001034847 | 1 | 1.00072388 |
| GSTA2      | 0.001066812 | 1 | 1.005385962 |
| SPP1       | 0.001097211 | 1 | 1.101737321 |
| MMP11      | 0.001108755 | 1 | 1.016851831 |
| CWH43      | 0.001450876 | 1 | 1.036477759 |
| TNS4       | 0.001582539 | 1 | 1.171200038 |
| PLA2G1B    | 1.44E-14  | 4.00E-10 | -2.10678444 |
| AGER       | 8.34E-14  | 2.33E-09 | -2.062570306 |
| CA4        | 4.34E-13  | 1.21E-08 | -2.088124716 |
| SLC6A4     | 3.08E-12  | 8.60E-08 | -2.237860051 |
| STX11      | 6.55E-12  | 1.83E-07 | -1.325682912 |
| HSD17B6    | 8.40E-12  | 2.34E-07 | -1.514855943 |
| CPB2       | 1.25E-11  | 3.47E-07 | -1.76408969 |
| CLDN18     | 1.65E-11  | 4.59E-07 | -1.520279583 |
| CRTC1      | 3.07E-11  | 8.55E-07 | -1.598044156 |
| GKN2       | 4.35E-11  | 1.21E-06 | -1.339491525 |
| PPLP       | 5.22E-11  | 1.46E-06 | -1.354622717 |
| FAM107A    | 5.54E-11  | 1.55E-06 | -1.504206175 |
| Name    | Pvalue     | adjPvalue | logFC    |
|---------|------------|-----------|----------|
| PEBP4   | 6.23E-11   | 1.74E-06  | -1.712436843 |
| IL1R1   | 7.30E-11   | 2.04E-06  | -1.393300726 |
| IL6     | 1.78E-10   | 4.97E-06  | -2.045946662 |
| LRRN4   | 2.55E-10   | 7.11E-06  | -1.300700327 |
| FCN3    | 2.76E-10   | 7.69E-06  | -2.097486722 |
| RND1    | 2.79E-10   | 7.79E-06  | -1.466163154 |
| AGTR2   | 4.24E-10   | 1.18E-05  | -1.676640353 |
| STC1    | 5.83E-10   | 1.63E-05  | -1.497051803 |
| SFTA2   | 5.83E-10   | 1.63E-05  | -1.149782392 |
| MS4A15  | 6.77E-10   | 1.98E-05  | -1.392100693 |
| FCN3    | 7.30E-11   | 2.04E-06  | -1.393300726 |
| IL1RL1  | 7.30E-11   | 2.04E-06  | -1.393300726 |
| IL6     | 1.78E-10   | 4.97E-06  | -2.045946662 |
| LRRN4   | 2.55E-10   | 7.11E-06  | -1.300700327 |
| FCN3    | 2.76E-10   | 7.69E-06  | -2.097486722 |
| RND1    | 2.79E-10   | 7.79E-06  | -1.466163154 |
| AGTR2   | 4.24E-10   | 1.18E-05  | -1.676640353 |
| STC1    | 5.83E-10   | 1.63E-05  | -1.497051803 |
| SFTA2   | 5.83E-10   | 1.63E-05  | -1.149782392 |
| MS4A15  | 6.77E-10   | 1.98E-05  | -1.392100693 |
| FCN3    | 7.30E-11   | 2.04E-06  | -1.393300726 |

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]
| Name        | Pvalue   | adjPvalue | logFC       | Name        | Pvalue   | adjPvalue | logFC       |
|-------------|----------|-----------|-------------|-------------|----------|-----------|-------------|
| CCDC85A     | 2.39E-06 | 0.06661217 | -1.06773488 | PTX3        | 1.34E-05 | 0.374984707| -1.338994441|
| IL18RAP     | 2.44E-06 | 0.06816221 | -1.09242787 | F11         | 1.52E-05 | 0.423768653| -1.184438918|
| ERKF1I      | 2.88E-06 | 0.08029286 | -1.09242787 | HTR3C       | 1.66E-05 | 0.464166403| -1.178516229|
| ORM1        | 3.26E-06 | 0.09085371 | -1.085120865| HMGC52      | 1.70E-05 | 0.47451476 | -1.14093025 |
| CXCL3       | 3.26E-06 | 0.09101093 | -1.001088364| ARC         | 2.03E-05 | 0.567374491| -1.211855264|
| SPOCK2      | 3.43E-06 | 0.09577850 | -1.023342385| ORM2        | 2.07E-05 | 0.577788023| -1.019830926|
| S100A3      | 3.49E-06 | 0.09745283 | -1.061209678| FPR2        | 2.11E-05 | 0.587693475| -1.029517016|
| CEBPD       | 3.87E-06 | 0.10798090 | -1.093847526| GPX3        | 2.24E-05 | 0.623819328| -1.09177645 |
| S100A8      | 4.39E-06 | 0.112247366| -1.139030199| XIST        | 2.28E-05 | 0.635377729| -1.103246381|
| SLC01A2     | 5.14E-06 | 0.14335538 | -1.086207968| MT1E        | 2.36E-05 | 0.658399738| -1.006714478|
| ESM1        | 6.05E-06 | 0.16871941 | -1.078977525| NR4A2       | 2.54E-05 | 0.706999207| -1.027905066|
| CSRN1P      | 6.49E-06 | 0.18090092 | -1.136678398| RS1         | 2.64E-05 | 0.734963593| -1.020756087|
| FGF21       | 8.96E-06 | 0.24993493 | -1.148678049| FOSB        | 3.08E-05 | 0.857680858 | -1.206609882|
| BDNF        | 9.86E-06 | 0.25047609 | -1.07653749 | RBP2        | 3.38E-05 | 0.941426241| -1.007947503|
| ZFP36       | 1.01E-05 | 0.28112421 | -1.023145887| KLRF1       | 3.44E-05 | 0.96012864 | -1.045789521|
| DLL4        | 1.05E-05 | 0.29233894 | -1.011654806| IL13        | 4.17E-05 | 1          | -1.023145887|
| DEFA1B      | 1.16E-05 | 0.32233994 | -1.011654806| S100A9      | 8.75E-05 | 1          | -1.045992124|
| VNN2        | 1.17E-05 | 0.32556005 | -1.015560522| ADM         | 0.00020169| 1          | -1.022596555|

Supplementary Table 2. Pathway and process enrichment analysis of those functional coexpression modules in IPF.

| Modules | GO          | Category                  | Description                              | Count | %      | Log10(P) | Log10(q) |
|---------|-------------|---------------------------|------------------------------------------|-------|--------|----------|----------|
| Black   | GO: 0062023 | GO Cellular Components   | Collagen-containing extracellular matrix  | 55    | 16.57  | -42.21   | -37.86   |
|         | GO: 0044420 | GO Cellular Components   | Extracellular matrix component           | 14    | 4.22   | -14.54   | -10.96   |
|         | GO: 0005539 | GO Molecular Functions    | Glycosaminoglycan binding                 | 23    | 6.93   | -13.09   | -9.58    |
|         | GO: 0005509 | GO Molecular Functions    | Calcium ion binding                      | 39    | 11.75  | -12.87   | -9.47    |
|         | GO: 0001501 | GO Biological Processes   | Skeletal system development              | 33    | 9.94   | -12.8    | -9.45    |
|         | GO: 0001578 | GO Molecular Functions    | Integrin binding                         | 17    | 5.12   | -11.75   | -8.43    |
|         | GO: 0001944 | GO Biological Processes   | Vasculature development                  | 38    | 11.45  | -10.39   | -7.37    |
|         | GO: 0005604 | GO Cellular Components   | Basement membrane                        | 13    | 3.92   | -9.38    | -6.27    |
|         | GO: 0001503 | GO Biological Processes   | Ossification                              | 24    | 7.23   | -9.13    | -6.1     |
|         | GO: 0060485 | GO Biological Processes   | Mesenchyme development                   | 20    | 6.02   | -8.9     | -5.9     |
|                  | GO: 0097529 | GO: 0002366 | GO: 0006954 | GO: 0098542 | GO: 010942 | GO: 0072521 | GO: 0043410 | hsa05206 | GO: 1901342 | GO: 000497 | GO: 1904158 | GO: 0023024 | GO: 0002223 | GO: 0001889 | GO: 0045177 | GO: 0000280 | GO: 0071103 | GO: 0044770 | GO: 0006260 | GO: 005819 | GO: 1903046 | GO: 0006281 | GO: 0034508 | GO: 005815 | GO: 0030496 |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| **Description** | Myeloid leukocyte migration | Leukocyte activation involved in immune response | Inflammatory response | Defense response to other organism | Positive regulation of cell death | Specific granule | Positive regulation of MAPK cascade | MicroRNAs in cancer | Regulation of vasculature development | Acute-phase response | Axonemal central apparatus assembly | MHC class I protein complex binding | Stimulatory C-type lectin receptor signaling pathway | Liver development | Apical part of cell | Nuclear division | DNA conformation change | Cell cycle phase transition | DNA replication | Spindle | Meiotic cell cycle process | DNA repair | Centromere complex assembly | Microtubule organizing center | Midbody |
| **Count**       | 23          | 41          | 42          | 30          | 34          | 16          | 30          | 21        | 25          | 9           | 3           | 3           | 10          | 15          | 28          | 54          | 46          | 47          | 36          | 35          | 26          | 36          | 15          | 36          |
| **%**           | 6.04        | 10.76       | 11.02       | 7.87        | 8.92        | 4.2         | 7.87        | 5.51      | 6.56        | 2.36        | 0.36        | 0.36        | 1.21        | 1.56        | 3.38        | 24         | 20.44       | 20.89       | 16          | 15.56       | 11.56       | 16          | 16          |
| **Log10(P)**    | –12.85      | –12.25      | –11.26      | –8.34       | –8.28       | –8.25       | –8.09       | –7.84     | –7.67       | –7.35       | –4.12       | –4.39       | –4.15       | –4.22       | –3.95       | –45.01     | –42.92      | –30.44      | –30.31      | –25.8       | –22.27      | –19.92      | –13.36      |
| **Log10(q)**    | –8.61       | –8.42       | –7.86       | –5.44       | –5.41       | –5.41       | –5.26       | –5.04     | –4.89       | –4.62       | –1.43       | –1.43       | –1.26       | –1.43       | –1.14       | –40.65     | –39.27      | –27.28      | –27.19      | –22.84      | –19.43      | –17.18      | –10.96      |
| Modules       | GO            | Category                        | Description                              | Count | %      | Log10(P) | Log10(q) |
|---------------|---------------|---------------------------------|------------------------------------------|-------|--------|----------|----------|
| **Pink module** | GO: 0042110   | GO Biological Processes         | T cell activation                        | 56    | 19.24  | −39.21   | −34.85   |
|               | GO: 0098552   | GO Cellular Components          | Side of membrane                         | 48    | 16.84  | −27.35   | −22.72   |
|               | GO: 0050852   | GO Biological Processes         | T cell receptor signaling pathway        | 27    | 9.28   | −23.27   | −20.03   |
|               | GO: 0001816   | GO Biological Processes         | Cytokine production                      | 50    | 17.18  | −23.19   | −19.98   |
|               | GO: 0045058   | GO Biological Processes         | T cell selection                         | 16    | 5.5    | −18.68   | −15.81   |
|               | GO: 0019221   | GO Biological Processes         | Cytokine-mediated signaling pathway      | 43    | 14.78  | −17.21   | −14.39   |
|               | hsa04660      | KEGG Pathway                    | T cell receptor signaling pathway        | 18    | 6.19   | −15.43   | −12.71   |
|               | GO: 0042101   | GO Cellular Components          | T cell receptor complex                  | 10    | 3.44   | −15      | −12.29   |
|               | GO: 0046631   | GO Biological Processes         | alpha-beta T cell activation             | 19    | 6.53   | −14.5    | −11.84   |
|               | GO: 0031349   | GO Biological Processes         | Positive regulation of defense response  | 29    | 9.97   | −13.11   | −10.51   |
|               |               |                                 | Blood vessel development                 | 66    | 10.48  | −17.01   | −12.65   |
|               | GO: 0034330   | GO Biological Processes         | Cell junction organization               | 34    | 5.4    | −12.64   | −8.98    |
|               | GO: 0016126   | GO Biological Processes         | Sterol biosynthetic process              | 18    | 2.86   | −11.61   | −8.1     |
|               | GO: 0070848   | GO Biological Processes         | Response to growth factor                | 53    | 8.41   | −10.72   | −7.26    |
|               | GO: 0030155   | GO Biological Processes         | Regulation of cell adhesion              | 48    | 7.62   | −9.5     | −6.25    |
|               | GO: 000318    | GO Biological Processes         | Vascular process in circulatory system   | 21    | 3.33   | −8.83    | −5.7     |
|               | GO: 0005911   | GO Cellular Components          | Cell-cell junction                       | 34    | 5.6    | −7.03    | −4.72    |
|               | GO: 0070372   | GO Biological Processes         | regulation of ERK1 and ERK2 cascade      | 28    | 4.44   | −7.38    | −4.5     |
|               | GO: 0008285   | GO Biological Processes         | Negative regulation of cell proliferation| 46    | 7.3    | −6.92    | −4.18    |
|               | GO: 0007610   | GO Biological Processes         | Behavior                                 | 38    | 6.03   | −6.59    | −3.93    |
|               | GO: 0001568   | GO Biological Processes         | Blood vessel development                 | 66    | 10.48  | −17.01   | −12.65   |
|               | hsa04740      | KEGG Pathway                    | Olfactory transduction                   | 108   | 6.76   | −34.96   | −30.6    |
|               | GO: 0031424   | GO Biological Processes         | Keratinization                           | 58    | 3.63   | −19.08   | −15.72   |
|               | GO: 0030594   | GO Molecular Functions          | Neurotransmitter receptor activity        | 26    | 1.63   | −7.43    | −4.33    |
|               | GO: 0005179   | GO Molecular Functions          | Hormone activity                         | 26    | 1.63   | −7.12    | −4.04    |
|               | GO: 0005261   | GO Molecular Functions          | Cation channel activity                  | 48    | 3      | −7.04    | −3.98    |
|               | hsa04080      | KEGG Pathway                    | Neuroactive ligand-receptor interaction   | 41    | 2.57   | −5.92    | −2.96    |
|               | GO: 0007188   | GO Biological Processes         | Adenylate cyclase-modulating G protein-coupled receptor signaling pathway | 34    | 2.13   | −5.38    | −2.5     |
|               | GO: 0009566   | GO Biological Processes         | Fertilization                            | 29    | 1.81   | −5.22    | −2.35    |
|               | GO: 0005549   | GO Molecular Functions          | Odorant binding                          | 13    | 1.19   | −4.03    | −2.03    |
|               | GO: 0007210   | GO Biological Processes         | Serotonin receptor signaling pathway     | 11    | 0.69   | −4.51    | −1.82    |

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

This work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)
Supplementary Figure 1. Sample clustering to detect outliers. The red line is used to distinguish the outlier samples. The threshold is set as 55, and three outlier samples (GSM806290, GSM806408, GSM806411) were removed from the sample cohort.

Supplementary Figure 2. The visualization of the coexpression networks. Nodes represent genes, and node color is the same as the module color node.

References:
1. Richeldi L, Collard HR, Jones MG: Idiopathic pulmonary fibrosis. Lancet, 2017; 389: 1941–52
2. Hutchinson J, Fogarty A, Hubbard R, McKeever T: Global incidence and mortality of idiopathic pulmonary fibrosis: A systematic review. Eur Resp J, 2015; 46: 795–806
3. Raghu G, Weycker D, Edelberg J et al: Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med, 2006; 174: 810–16.

4. Raghu G, Collard HR, Egan JJ et al: An official ATS/ERS/IAAF statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med, 2011; 183: 788–824.

5. Betensley A, Sharif R, Karamichos D: A systematic review of the role of dysfunctional wound healing in the pathogenesis and treatment of idiopathic pulmonary fibrosis. J Clin Med, 2016; 6(1): pii: 62.

6. Kropski JA, Lawson WE, Young LR, Blackwell TS: Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. Dis Model Mech, 2013; 6: 9–17.

7. Gulati S, Thannickal VJ: The aging lung and idiopathic pulmonary fibrosis. Am J Med Sci, 2019; 357: 384–89.

8. Noth I, Zhang Y, Ma SF et al: Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: A genome-wide association study. Lancet Respir Med, 2013; 1: 309–17.

9. Seibold MA, Wise AL, Speer MC et al: A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med, 2011; 364: 1503–12.

10. Baumgartner KB, Samet JM, Stidley CA et al: Cigarette smoking: A risk factor for idiopathic pulmonary fibrosis. Am J Respir Crit Care Med, 1997; 155: 242–48.

11. Taskar VS, Coultas DB: Is idiopathic pulmonary fibrosis an environmental disease? Proc Am Thorac Soc, 2006; 3: 293–98.

12. Rung J, Braza M: Use of public genome-wide gene expression data. Nat Rev Genet, 2013; 14: 89–99.

13. Kolde R, Laur S, Adler P, Vilo J: Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics, 2012; 28: 573–80.

14. Lamb J, Crawford ED, Peck D et al: The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science, 2006; 313: 1929–35.

15. Song Y, Liu J, Huang S, Zhang L: Analysis of differentially expressed genes in placental tissues of preeclampsia patients using microarray combined with the Connectivity Map database. Placenta, 2013; 34: 1190–95.

16. Langfelder P, Horvath S: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics, 2008; 9: 559.

17. Gentleman RC, Carey VJ, Bates DM et al: Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol, 2004; 5(10): R80.

18. Izirrray RA, Hobbs B, Collin F et al: Exploration, normalization, and summaries of high-density oligonucleotide array probe level data. Biostatistics, 2004; 3: 249–64.

19. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in macroarray experiments. Stat Appl Genet Mol Biol, 2004; 3: Article3.

20. Shannon P, Markiel A, Ozier O et al: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 2003; 13: 2498–504.

21. Bardou P, Mariette J, Escudié F et al: jvenn: An interactive Venn diagram study. Lancet Respir Med, 2013; 1: 309–17.

22. Tripathi S, Verhamme FM et al: Association of innate defense proteins BPFFA1 and BPFFB1 with disease severity in COPD. Int J Chron Obstruct Pulm Dis, 2018; 13: 11–27.

23. Zuo W, Zhang T, Wu DZ et al: p36(C)kril5(+) distal airway stem cells are essential for lung regeneration. Nature, 2015; 517: 616–20.

24. De Smet EG, Vanhaecke M et al: Identification of novel molecular subtypes of idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score. Sci Rep, 2017; 7: 12569.

25. Ban MJ, Ji SH, Lee CK et al: Organic anion transporting polypeptide of the solute carrier organic anion transporter family member 4A1 (SLCO4A1) as a prognosis marker of colorectal cancer. J Gastroenterol, 2017; 52: 802–11.

26. Hagenbuch B, Meier PJ: Organic anion transporting polypeptides of the OATP/SLC21 family: Physiogenetic classification as OATP/SLCO superfamily, new nomenclature and molecular/functional properties. Pflugers Arch, 2004; 447(5): 653–65.

27. Wang Z et al.: WGCNA for key genes and pathways in PF