Identification of key genes associated with idiopathic pulmonary fibrosis using bioinformatics analysis

Yu Chen, Ping Jiang

Department of Emergency, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China
Received November 6, 2015; Accepted January 3, 2016; Epub February 1, 2016; Published February 15, 2016

Abstract: Idiopathic pulmonary fibrosis (IPF) is a progressive and fatal interstitial lung disease. The goal of this study is to elucidate the molecular mechanism of IPF. GSE24206 was downloaded from Gene Expression Omnibus, which included 17 IPF and 6 control samples. The t-test was applied to identify differentially expressed genes (DEGs) between IPF and control samples. Pathway and functional enrichment analyses were used to investigate the functions involving these DEGs. According to the information of TRANSFAC, Tumor Associated Gene (TAG) and Tumor Suppressor Gene (TSGene) databases, the screened DEGs were further annotated. To comprehensively understand the interactions between proteins encoded by the DEGs, protein-protein interactions (PPIs) were predicted by STRING and PPI network was visualized by Cytoscape software. Additionally, module analysis for PPI network was performed using BioNet tool. Total 192 up-regulated and 28 down-regulated genes were identified. Both down-regulated PDGFRA and up-regulated CCND1 were TAGs. Pathway enrichment analysis indicated that PDGFRA were involved in all of the 8 pathways for the 28 down-regulated genes. Besides, LTBP3 and THY1 separately were involved in extracellular matrix organization and cell adhesion. After PPI network analysis, we discovered that the degree of COL1A2, TGFB1, COL1A1, COL3A1, ASPN, CD4, SDC1, CXCL12, COL5A1, and COMP were significantly higher. In conclusions, our results showed that the pathology of IPF involved multiple dysregulated genes, and our study would pave ways for further study of IPF.

Keywords: Idiopathic pulmonary fibrosis, differentially expressed genes, enrichment analysis, protein-protein interaction network, module analysis

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive and fatal interstitial lung disease, which is characterized by temporally heterogeneous lung architectural distortion, dense collagen and extracellular matrix (ECM) deposition in interstitium, alveolar collapse, and the presence of fibroblastic foci [1]. Nowadays, IPF affects about five million people worldwide, and its incidence is about 20 to 60 per 100,000 persons [2]. Besides, IPF occurs usually in middle-aged and older adults, and men are more susceptible to IPF [3]. Notably, the incidence of lung cancer seems to be increased in IPF patients compared with general population [4, 5]. Lung cancer may occur before, after, or at the time when IPF is diagnosed [6].

Though IPF cannot be cured, oxygen therapy, lung transplantation, and drugs have been used to help IPF patients. Nowadays, many drugs (like macitentan, sildenafil, warfarin, and bosentan) have been developed, but these drugs show little benefit [7]. Recently, in order to get novel therapeutic targets, the pathogenesis of IPF has been studied. Previous studies showed that deficiencies of surfactant protein C (SP-C, encoded by SFTPC) [8] and surfactant protein A2 (SP-A2, encoded by SFTPA2) [9] are associated with IPF. Seibold et al. identify a common variant in the putative promoter of mucin 5B (MUC5B), which presents in 38% of patients with IPF [10]. Disease-causing heterozygous mutations in two components of telomerase complex, telomerase reverse transcriptase (TERT) and telomerase RNA component (TERC), are also involved in IPF [11, 12]. Besides, many biological pathways which are linked to IPF have been identified. Epithelial-mesenchymal transition in alveolar epithelial cells (AECs) is hypothesized as a source of myo-
fibroblasts which serves as the primary collagen-producing cell [13]. ECM deposition, which is regulated by matrix metalloproteinases (MMPs) and their inhibitors, can be triggered by chronic inflammation and lead to the formation of a permanent fibrotic scar [14]. Despite extensive research, the pathogenesis of IPF still remains unclear.

In 2011, Meltzer et al. screened differentially expressed genes (DEGs) between upper and lower lobe samples using paired t-tests, and identified DEGs between IPF explants and IPF biopsies using unpaired Student’s t-tests [15]. Using the data deposited by Meltzer et al. [15], the DEGs between IPF and normal samples were screened, and their underlying functions were predicted by functional and pathway enrichment analyses. Besides, gene functional annotation analysis was performed. Additionally, protein-protein interaction network (PPI) network and module were constructed to investigate the interactions between these DEGs.

**Materials and methods**

Collection and preprocessing of mRNA expression profile data

The mRNA expression profile of GSE24206 deposited by Meltzer et al. [15] was downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database using the platform of Affymetrix Human Genome U133 Plus 2.0 Array. The dataset included 17 samples (GSM595421, GSM595422, GSM595423, GSM595424, GSM595425, GSM595426, GSM595427, GSM595428, GSM595429, GSM595432, GSM595434, GSM595435, GSM595437, GSM595439, GSM595441, GSM595443, GSM595445) from 11 IPF patients (6 patients contributed twin samples from upper and lower lobes, and 5 patients provided singleton samples) and 6 samples (GSM595407, GSM595411, GSM595414, GSM595416, GSM595417, GSM595419) from healthy donors (Healthy donors provided lung samples obtained from routine lung volume reduction of lung during lung transplantation). Combining with the probe annotation file of Affy [16] chip provided by Brain Array Lab, the original data were preprocessed using AFFY package in Bioconductor [17]. After Robust Multi-array Average (RMA) [18] background correction, quantile normalization and probe summarization, gene expression matrix of the samples were obtained.

**DEGs screening**

For the preprocessed data, t-test [19] was performed to identify DEGs between IPF and control samples. We defined FDR < 0.05 and |log2 fold change (FC)| ≥ 1 as the thresholds.

**Functional and pathway enrichment analysis**

To study the DEGs at a functional level, Gene Ontology (GO) functional [20] enrichment analy-
Mechanisms of IPF

Table 3. The enriched GO functions for the DEGs

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Gene counts</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up</td>
<td>GO: 0030198 Extracellular matrix organization</td>
<td>19</td>
<td>8.33E-10</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0043062 Extracellular structure organization</td>
<td>19</td>
<td>8.79E-10</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0007155 Cell adhesion</td>
<td>31</td>
<td>2.25E-08</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0005576 Extracellular region</td>
<td>71</td>
<td>0</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0044421 Extracellular region part</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0005615 Extracellular space</td>
<td>36</td>
<td>1.59E-13</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0005201 Extracellular matrix structural constituent</td>
<td>10</td>
<td>4.75E-09</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0008201 Heparin binding</td>
<td>10</td>
<td>5.48E-07</td>
</tr>
<tr>
<td>Up</td>
<td>GO: 0048407 Platelet-derived growth factor binding</td>
<td>4</td>
<td>2.60E-06</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0034754 Cellular hormone metabolic process</td>
<td>3</td>
<td>0.000451975</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0060325 Face morphogenesis</td>
<td>2</td>
<td>0.001132088</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0060323 Head morphogenesis</td>
<td>2</td>
<td>0.001442882</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0048008 Platelet-derived growth factor receptor signaling pathway</td>
<td>2</td>
<td>0.001699747</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0060324 Face development</td>
<td>2</td>
<td>0.001789859</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0010171 Body morphogenesis</td>
<td>2</td>
<td>0.00217258</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0004745 Retinol dehydrogenase activity</td>
<td>2</td>
<td>0.000243186</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0005001 Transmembrane receptor protein tyrosine Phosphatase activity</td>
<td>2</td>
<td>0.000353342</td>
</tr>
<tr>
<td>Down</td>
<td>GO: 0019198 Transmembrane receptor protein phosphatase activity</td>
<td>2</td>
<td>0.000353342</td>
</tr>
</tbody>
</table>

GO: Gene Ontology; DEGs: differentially expressed genes.

Table 4. The functional statistics of DEGs between IPF and control samples

<table>
<thead>
<tr>
<th>TF counts</th>
<th>TF genes</th>
<th>TAG counts</th>
<th>TAG (Onco-genes)</th>
<th>TGA (Tumor Suppressor Gene)</th>
<th>TAG (other)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down</td>
<td>1</td>
<td>NFI3</td>
<td>5</td>
<td>PDGFRA</td>
<td>PTPRG, HOPX</td>
</tr>
<tr>
<td>Up</td>
<td>3</td>
<td>SOX4, NR1H3, MEOX1</td>
<td>18</td>
<td>CD24, CCND1</td>
<td>THY1, STEAP3, SCGB3A1, SCARA3, PDLIM4, NBL1, NAPEPLD, IGFBP4, HTRA1, ENC1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>XAF1, TGFβ1, SSPN, MUC5B, LRRC17, FHL2</td>
</tr>
</tbody>
</table>

DEGs: differentially expressed genes; IPF: Idiopathic pulmonary fibrosis.

sis and Kyoto Encyclopedia of Gene and Genomes (KEGG) [21] pathway enrichment analysis were performed. The p-value < 0.01 was set as the cut-off criterion.

Functional annotation analysis

According to the information of transcription factors provided by TRANSFAC [22] database, the DEGs were further screened and annotated to obtain genes with transcriptional regulation function. Besides, all known oncogenes and tumor suppressor genes were extracted based on Tumor Associated Genes (TAG) database [23] and Tumor Suppressor Gene (TSGene) database [24].

PPI network and module construction

Containing known and predicted protein-protein interactions, STRING database [25] has been widely used to construct PPI network. Here, STRING database was used to search interactions of the proteins encoded by DEGs. The Cytoscape software [26] was used to visualize the PPI network. Subsequently, BioNettool [27] was employed for performing module analysis for PPI network, and FDR < 0.0001 was set as the criterion.

Results

DEGs screening

Using t-test, a total of 220 DEGs were screened in IPF samples compared with normal samples, including 192 up-regulated genes (corresponding to 521 transcripts) were, and 28 down-regulated genes (corresponding to 109 transcripts) (Table 1).

Functional and pathway enrichment analysis

Pathway enrichment analysis indicated that the 192 up-regulated genes were enriched in 5
Mechanisms of IPF

Among the up-regulated genes, 3 genes were transcription factors, and 18 genes (e.g. cyclin

Figure 1. PPI network of DEGs between IPF and control samples. Red and green nodes represent up-regulated and down-regulated genes, respectively. PPI: protein-protein interaction; DEGs: differentially expressed genes; IPF: idiopathic pulmonary fibrosis.

pathways, such as complement and coagulation cascades ($P = 0.000101308$), ECM-receptor interaction ($P = 0.000321922$) and staphylococcus aureus infection ($P = 0.00032338$). Meanwhile, the 28 down-regulated genes were enriched in 8 pathways, including constitutive PI3K/AKT signaling in cancer ($P = 0.006803308$) and PI3K events in ERBB4 signaling ($P = 0.008763174$). Notably, insulin receptor substrate 2 ($IRS2$) and platelet-derived growth factor alpha receptor ($PDGFRA$) were enriched in all of the 8 pathways for down-regulated genes (Table 2). Moreover, the 192 up-regulated genes were enriched in some GO functions, including extracellular matrix organization ($P = 8.33E-10$, which involved latent TGF-β binding protein-3, $LTBP3$) and cell adhesion ($P = 2.25E-08$, which involved thymus cell antigen 1, $THY1$). And the 28 down-regulated genes were also enriched in several GO functions, including cellular hormone metabolic process ($P = 0.000451975$) and face morphogenesis ($P = 0.001132088$) (Table 3).

Functional annotation analysis

Among the up-regulated genes, 3 genes were transcription factors, and 18 genes (e.g. cyclin
Mechanisms of IPF

Based on STRING database, PPI network was constructed (Figure 1), and the top 10 genes with degree ≥ 9 were α 2 type I collagen gene (COL1A2, degree = 17), transforming growth-factor-β1 (TGFB1, degree = 17), α 1 type I collagen gene (COL1A1, degree = 16), α 1 type III collagen gene (COL3A1, degree = 15), asporin (ASPN, degree = 14), CD4 (degree = 12), syndecan-1 (SDC1, degree = 10), stromal cell-derived factor 1 (CXCL12, degree = 10), α 1 type V collagen gene (COL5A1, degree = 9) and cartilage oligomeric matrix protein (COMP, degree = 9). The module involving 37 nodes was obtained from the PPI network, in which COL1A1 (degree = 11) had the highest degree (Figure 2). After KEGG pathway enrichment analysis, genes in this module were mainly involved in ECM-receptor interaction (P = 7.39E-06), protein digestion and absorption (P = 0.000137815) and axon guidance (P = 0.0-0.08992409) (Table 5). Moreover, through GO functional enrichment analysis, the genes in this module were mainly enriched in extracellular matrix organization (P = 3.12E-
Mechanisms of IPF

Table 5. KEGG pathway and GO functional enrichment analysis of DEGs in the identified module

<table>
<thead>
<tr>
<th>Enriched terms</th>
<th>Gene Counts</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG pathway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECM-receptor interaction</td>
<td>5</td>
<td>7.39E-06</td>
</tr>
<tr>
<td>Protein digestion and absorption</td>
<td>4</td>
<td>0.000137815</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>5</td>
<td>0.000445781</td>
</tr>
<tr>
<td>Amoebiasis</td>
<td>3</td>
<td>0.005217842</td>
</tr>
<tr>
<td>Axon guidance</td>
<td>3</td>
<td>0.008992409</td>
</tr>
<tr>
<td>GO function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracellular matrix organization</td>
<td>13</td>
<td>3.12E-13</td>
</tr>
<tr>
<td>Extracellular structure organization</td>
<td>13</td>
<td>3.25E-13</td>
</tr>
<tr>
<td>Multicellular organismal catabolic process</td>
<td>7</td>
<td>9.13E-10</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>15</td>
<td>3.80E-09</td>
</tr>
<tr>
<td>Biological adhesion</td>
<td>15</td>
<td>3.91E-09</td>
</tr>
</tbody>
</table>

KEGG: Kyoto Encyclopedia of Gene and Genomes; GO: Gene Ontology; DEGs: differentially expressed genes.

13, extracellular structure organization (P = 3.25E-13) and biological adhesion (P = 3.91E-09) (Table 5).

Discussion

IPF is a devastating form of interstitial lung disease [28]. However, there is no effective treatment. To understand the potential mechanism of IPF, bioinformatics might be an effective method. In this study, a total of 220 DEGs were identified in IPF samples in comparison to control samples.

Peroxisome proliferator-activated receptor γ (PPARγ) agonists can suppress TGF-β-induced myofibroblast differentiation and production of collagen protein, hence, PPARγ agonists have potential antifibrotic effects and may be used in therapy of fibrotic lung diseases [29]. The synthesis of type I collagen reacts both positively and negatively to stimulation generated by tissue injury and repair, and is accumulated in IPF patients [30, 31]. Via transcriptional activating COL1A2, connective tissue growth factor (CTGF) contributes to lung fibrosis and may serve as a promising target for treatment of fibrotic diseases [32]. ASPN can bind with collagen and calcium, and then induce collagen mineralization which is essential for ECM deposition [33]. Previous study shows that down-regulated CD28 in circulating CD4 T-cells are related to manifestations and progression of IPF [34]. Increased syndecan-1 (which is encoded by SDC1) have been detected in lung homogenates and lavage fluid of lungs in patients with IPF, and syndecan-1 ectodomain induces neutrophil chemotaxis, inhibits wound healing in alveolar epithelial, and promotes fibrogenesis [35]. In the bleomycin model, up-regulated CXCL12 is the major chemokine responsible for recruiting bone-marrow derived fibrocytes to lung [36]. COMP was overexpressed in serum of IPF patients and it may be a novel biomarker for disease activity and TGF-β1 activity [37, 38]. In the PPI network, COL1A2, TGFβ1, COL1A1, COL3A1, ASPN, CD4, SDC1, CXCL12, COL5A1, and COMP had higher degrees. Module analysis showed that COL1A1 had the highest degree in the identified module. These indicated that these genes might be key genes in IPF.

Functional enrichment indicated that LTBP3 was involved in extracellular matrix organization. The ECM protein LTBP3 have a dual function, which is required both for the secretion of small latent TGF-beta complex and binding latent TGF-beta to ECM microfibrils [39, 40]. As stated before, growth factors TGF-beta stimulates ECM production of fibroblast, myofibroblast differentiation, and resistance to apoptosis [41, 42]. THY1, which involved in cell adhesion, has been proposed as a “fibrosis suppressor” gene [43]. THY1 is present in normal lung fibroblasts [44], but absent in the fibroblasts of IPF patients because of methylation [43]. Thus, LTBP3 and THY1 might play an important role in IPF progression.

Additionally, gene functional annotation analysis showed that 5 down-regulated genes (e.g. PDGFR) and 18 up-regulated genes (e.g. CCND1) are TAGs. Schwartz et al. hypothesis that CCND1 plays an instrumental role in the pro-fibrogenic process, which was further validated by in situ growth factor overproduction and exaggerated extracellular matrix deposition [45]. Intedanib is a triple kinase inhibitor that blocks PDGFR, vascular endothelial growth factor receptor (VEGFR) and fibroblast growth factor receptor (FGFR) for the therapy of IPF.
and several types of cancer [46]. Pathway enrichment analysis indicated that PDGFRA were involved in all of the 8 pathways for the 28 down-regulated genes. These suggested that PDGFRA and CCND1 might be implicated in IPF.

Conclusions

In conclusion, to illustrate the pathological mechanism of IPF, the gene expression profile containing 23 samples was downloaded and analyzed. Total 220 DEGs were identified in IPF samples. Besides, several genes (COL1A2, TGFB1, COL1A1, COL3A1, ASPN, CD4, SDC1, CXCL12, COL5A1, COMP, LTBP3, THY1, CCND1 and PDGFRA) might play important roles in IPF. However, further experimental validation is still needed to prove this speculation.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Ping Jiang, Department of Emergency, Shanghai Pulmonary Hospital, Tongji University School of Medicine, 507 Zheng Min Road, Yangpu District, Shanghai 200433, P. R. China. Tel +86-21-65115006-2038; Fax: +86-21-65115006-2038; E-mail: pingjiangpjpj@163.com

References


Mechanisms of IPF


[38] Farina G, Widom RL, Benvenuto R, De Santis M, Tolusso B. Ferraccioli G COMP as a marker of active fibrosis in scleroderma: Immuno-
Mechanisms of IPF


