Bioinformatics Analysis Identifies Potential Key Genes of Peripheral Blood Mononuclear Cell in Idiopathic Pulmonary Fibrosis

Lijun Liu¹, Daxia Cai², Yi Wu¹,*

¹Department of Respiratory Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, China
²Institute of Hematology, Jinan University, Guangzhou, China

Email address:
twuyi@jnu.edu.cn (Yi Wu)
*Corresponding author

To cite this article:
Lijun Liu, Daxia Cai, Yi Wu. Bioinformatics Analysis Identifies Potential Key Genes of Peripheral Blood Mononuclear Cell in Idiopathic Pulmonary Fibrosis. Computational Biology and Bioinformatics. Vol. 8, No. 2, 2020, pp. 77-89. doi: 10.11648/j.cbb.20200802.17

Received: November 4, 2020; Accepted: November 17, 2020; Published: November 27, 2020

Abstract: Idiopathic pulmonary fibrosis (IPF) is a chronic progressive fibrotic interstitial pneumonia with progressive worsening of dyspnea and lung function. The etiology of IPF is unknown, and the pathogenesis remains unclear. Our study aimed to investigate the key genes of the peripheral blood mononuclear cell in IPF by bioinformatics analysis. Our study used the online Gene Expression Omnibus (GEO) microarray expression profiling dataset GSE28042 to identify differentially expressed genes (DEGs) between IPF patients and healthy controls. We performed the Gene Ontology (GO) and pathway enrichment analyses of genes for annotation, visualization, and integrated discovery. The STRING database constructed Protein-protein interaction (PPI) network analysis, and hub genes were identified by the CytoHubba plugin. Moreover, we used the receiver operating characteristic (ROC) curve to assess the diagnostic value of the hub genes. In total, 28 upregulated and 44 downregulated genes were identified in the differential expression analysis. The protein-protein interaction network (PPI) was established with 69 nodes and 68 edges. The top 10 hub genes were JUN, FOS, STAT3, SOCS3, JUNB, DUSP1, IL4, FCER1A, MS4A2, and CPA3. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched for the important module containing hub genes contained Fc epsilon RI signaling pathway, TNF signaling pathway, Jak-STAT signaling pathway, and MAPK signaling pathway. Additionally, the identified hub genes show high functional similarity and diagnostic value in IPF. Our study used bioinformatics analysis to provide new insight into the mechanisms underlying IPF. However, more experiments are needed to explore the relationships between the top 10 hub genes and IPF in the future.

Keywords: Bioinformatic Analysis, Genes, Peripheral Blood Mononuclear Cell, Idiopathic Pulmonary Fibrosis

1. Introduction

Idiopathic pulmonary fibrosis (IPF) is defined as a specific form of chronic, progressive fibrosing interstitial pneumonia and its cause is unknown [1]. Its clinical features are unexplained exertional dyspnea, chronic dry cough, or Velcro rale on examination [2]. The histopathology of this disease is interstitial fibrosis with spatial heterogeneity and patchy involvement of lung parenchyma, and microscopic honeycombing [3]. The incidence of IPF is high and its incidence rates increased over time in most countries. Its incidence ranges from 0.2 per 100000 per year to 93.7 per 100000 per year based on estimates from Europe and North America [4]. Besides, median survival from the time of diagnosis is only 2.5 to 3.5 years [5]. Currently, the risk factors of IPF are environmental, genetic, epigenetic alterations, and aging [6]. However, the pathogenesis is still not exact.

So far, understanding of the pathogenesis of IPF includes epithelial cell dysfunction caused by genetic susceptibility, defined profibrotic processes caused by TGF-beta activation, and progressive pulmonary fibrosis [7, 8]. However, recent studies have shown that epithelial cell dysfunction is still a central cause of IPF [7]. Currently, the diagnosis of IPF requires exclusion of other known causes of interstitial disease (ILD), high-resolution, and surgical lung biopsy [1]. However, the accuracy of the diagnosis of IPF requires experienced
physicians, which has certain limitations. IPF patients are usually in the terminal stages of the disease when diagnosed and there is no particularly effective treatment at this point. Therefore, early diagnosis and early treatment were most important.

Bioinformatics analyses can enable researchers to search online biological databases to explore the pathogenesis and molecular diagnosis, such as juvenile dermatomyositis [9], major depressive disorder [10], Tuberculosis [11]. Here, we use the peripheral blood mononuclear cell (PBMC) microarray dataset GSE28042 created by Herazo-Maya et al. to investigated the differentially expressed genes (DEGs) between IPF patients and healthy controls to explore the key genes. Our findings will provide new insights into the clinical diagnosis and treatment of IPF.

2. Materials and Methods

2.1. Microarray Data

The microarray data of GSE28042 was downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. The dataset was based on the GPL6480 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name Version). The data type was expression profiling by the array and the species selected was Homo sapiens. The peripheral blood mononuclear cell samples (PBMC) included 75 patients with the diagnosis of IPF (IPF group) and 19 healthy controls (control group). The clinical details of GSE28042 were listed (Table 1). The annotation file for GPL6480 was also downloaded from the GEO.

2.2. Analysis of Differentially Expressed Genes (DEGs)

We used the online analysis tool GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) to screen the DEGs between the IPF group and the control group. The GEO2R can allow us to compare two or more groups of samples to identify genes that are differentially expressed across experimental conditions. The results are presented as a table of genes ordered by significance. To exclude gender differences, we divided the IPF group and the control group into two groups (Male IPF group, Female IPF group, Male control group, Female control group), respectively, depending on the gender. P-values and adjusted P-values (adj. p) were calculated using t-tests. Genes with log2 fold change (FC) >1 and adj. p <0.05 were identified as DEGs. A Venn diagram of DEGs was drawn using the online tool Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/). The heatmap for the DEGs was created using R software (version 4.0.2).

2.3. Gene Ontology (GO) and Pathway Enrichment Analysis of DEGs

GO, a bioinformatics tool aims to establish a vocabulary that defines and describes the functions of genes and proteins for a variety of species. GO is divided into three parts: Molecular Function (MF), Biological Process (BP), and Cellular Component (CC) [12]. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database that systematically analyzes the metabolic pathways of gene products in cells and their functions [13]. KEGG integrates data on genomes, chemical molecules, and biochemical systems, including metabolic pathways, drugs, diseases, genes, and genomes [13]. The Database for Annotation, Visualization, and Integrated Discovery (DAVID; http://david.ncifcrf.gov; version 6.8) is a free online biological information database and it provides a comprehensive set of functional annotation tools for researchers to understand biological meaning behind a large list of genes. We performed GO and KEGG pathway enrichment analyses using the DAVID online database to analyze the function of DEGs. P-value <0.05 was set as the cut-off criteria.

2.4. Protein-protein Interaction (PPI) Network Analysis and Module Analysis

A significant number of proteins do not function alone. Proteins interact with each other to form complexes that then do their work. We used the Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/; version 11.0) online database to systematically predict and construct a PPI network of all DEGs. A combined score >0.4 of PPI pairs was regarded as a significant interaction. Cytoscape (version 3.8.0), a bioinformatic software available online, can be used to construct and visualize the network of PPI. MCODE (version 1.5.1), a plugin of Cytoscape software, can construct functional modules by clustering in a large network of PPI. CytobHubba, mainly used for exploring PPI network hub genes, is a Cytoscape plugin. We selected the genes with the highest ranking by the maximum correlation criterion (MCC). The selected genes are represented by redder color.

2.5. Go and Pathway Enrichment Analysis of Hub Genes

To analyze the function of hub genes, biological analyses were performed using the DAVID online database. P-value <0.05 was set as the cut-off criteria.

2.6. Hub Genes Diagnostic Efficacy Evaluation

The receiver operating characteristic (ROC) curve can be used to assess the diagnostic accuracy. We use the “PROC” package of the R software to plot the ROC curve, calculate the area under the curve (AUC) and evaluate the diagnostic capability of hub genes to distinguish IPF patients and healthy controls.

2.7. Statistical Analysis

All statistical analyses were performed as the
means ± standard deviation. The R software (version 4.0.2) was used to analyze the data. A P-value < 0.05 was considered statistically significant.

3. Results

3.1. Differentially Expressed Genes

We downloaded the microarray expression dataset GSE28042 from the GEO database and analyzed the DEGs between IPF patients and healthy controls using the GEO2R tool. In total, 107 upregulated and 244 downregulated genes were identified between male IPF patients and male healthy controls. Besides, 54 upregulated and 212 downregulated genes were identified between female IPF patients and female healthy controls. The intersection of these two datasets identified 28 upregulated and 44 downregulated genes (Table 2). The Venn diagram and heatmap for the DEGs are presented in Figure 1.

| Table 2. Differentially expressed genes of IPF |
|-----------------------------------------------|
| Gene Symbol | ID          | adj. P. Val | P. Value |
|             |             | male        | female   |
|             |             | male        | female   |
| Upregulated genes | | | |
| API1S | A_23_P157404 | 0.000121 | 0.0297 |
| API1S | A_24_P63950 | 0.000674 | 0.0262 |
| AF5B1 | A_24_P9883 | 0.000151 | 0.0037 |
| APOE6 | A_24_P941167 | 0.00146 | 0.0367 |
| COQ10B | A_32_P1192 | 0.00116 | 0.0290 |
| DUSP1 | A_23_P110712 | 0.0000000888 | 0.0179 |
| EMP1 | A_24_P921446 | 0.000652 | 0.0225 |
| FAM20A | A_23_P108254 | 0.0000741 | 0.0244 |
| FAM20A | A_24_P352952 | 0.0000433 | 0.023 |
| FCAR | A_24_P348265 | 0.0000142 | 0.0132 |
| FOS | A_23_P106194 | 0.0000000612 | 0.000409 |
| GSTTT2 | A_23_P109427 | 0.00684 | 0.0173 |
| JUN | A_23_P201538 | 0.0000000514 | 0.00507 |
| JUNB | A_24_P241815 | 0.0000117 | 0.0139 |
| LMNB1 | A_23_P258493 | 0.00227 | 0.0305 |
| MBD2 | A_23_P119201 | 0.000408 | 0.0471 |
| PDAP1 | A_23_P151198 | 0.00216 | 0.0382 |
| PDK4 | A_23_P257807 | 0.00119 | 0.0128 |
| PDK4 | A_24_P243749 | 0.00132 | 0.00476 |
| PEAK1 | A_24_P933801 | 0.000244 | 0.0081 |
| PER1 | A_24_P93916 | 0.0000554 | 0.0126 |
| RHOB | A_23_P51136 | 0.0000000749 | 0.00363 |
| SERPINB2 | A_23_P153185 | 0.000491 | 0.0159 |
| SLED1 | A_24_P927716 | 0.0001 | 0.0346 |
| SOCS3 | A_23_P207058 | 0.000000918 | 0.0111 |
| SOCS3 | A_23_P351069 | 0.0000018 | 0.00738 |
| SOD2 | A_24_P935819 | 0.0000179 | 0.0226 |
| SRGN | A_23_P915269 | 0.000000042 | 0.00287 |
| STAC3 | A_23_P10947 | 0.000000318 | 0.0081 |
| STAT3 | A_24_P923962 | 0.0001 | 0.0108 |
| TEX14 | A_32_P126079 | 0.000395 | 0.0181 |
| TMEM107 | A_23_P118791 | 0.0000914 | 0.0225 |

Downregulated genes | | | |
| ABC13 | A_23_P397480 | 0.0366 | 0.0257 |
| C10orf82 | A_23_P1286 | 0.000338 | 0.0291 |
| C1orf186 | A_23_P95640 | 0.00000435 | 0.0294 |
| CACNG6 | A_23_P501933 | 0.000000956 | 0.045 |
| CCD85A | A_23_P349566 | 0.0115 | 0.0111 |
| CCR3 | A_23_P250302 | 0.000000207 | 0.0108 |
| CCR3 | A_23_P367473 | 0.00000879 | 0.0465 |
| CD24 | A_23_P85250 | 0.0102 | 0.0244 |
| CLC | A_23_P101683 | 0.0000306 | 0.0376 |
| CPA3 | A_23_P18017 | 0.0000000888 | 0.0195 |
| ENPP3 | A_23_P404536 | 0.0000018 | 0.0181 |
| EPB42 | A_23_P140675 | 0.0319 | 0.0195 |
| FCER1A | A_23_P103765 | 0.000395 | 0.0382 |
| FCRLA | A_24_P276576 | 0.00319 | 0.0402 |
| GPR37 | A_23_P145995 | 0.0336 | 0.0167 |

adj. P. Val | P. Value | male | female |
|-------------|----------|------|--------|
| 0.00593 | 0.000317 |
| 0.000497 | 0.000505 |
| 0.000131 | 0.000453 |
| 0.000041 | 0.000142 |
| 0.000312 | 0.000312 |
| 0.000013 | 0.0000197 |
| 0.000179 | 0.000142 |
| 0.000196 | 0.000196 |
| 0.000883 | 0.000883 |
| 0.00000307 | 0.00000307 |
| 0.0000156 | 0.0000391 |
| 0.000000011 | 0.000000011 |
| 0.00000575 | 0.00000575 |
| 0.0000906 | 0.0000906 |
| 0.0000119 | 0.0000119 |
| 0.0000000387 | 0.0000000387 |
| 0.0000431 | 0.0000431 |
| 0.000000000259 | 0.000000000259 |
| 0.000000000348 | 0.000000000348 |
| 0.000000000721 | 0.000000000721 |
| 0.000000000694 | 0.000000000694 |
| 0.00000000017 | 0.00000000017 |
| 0.00000000051 | 0.00000000051 |
| 0.000000000259 | 0.000000000259 |
| 0.000000000348 | 0.000000000348 |
| 0.000000000711 | 0.000000000711 |
| 0.000000000787 | 0.000000000787 |
| 0.000000000824 | 0.000000000824 |
Table 2. Continued.

| Gene Symbol | ID        | log FC  | adj. P Val | P Value  |
|-------------|-----------|---------|------------|----------|
|             |           | male    | female     | male      | female   |

Uregulated genes

API1S1 A_23_P157404 1.1140727 1.2453211 adaport related protein complex 1 sigma 1 subunit
API1S1 A_24_P63950 1.0989288 1.2347777 adaport related protein complex 1 sigma 1 subunit
AP5B1 A_24_P9883 1.6985578 1.5936472 adaport related protein complex 5 beta 1 subunit
APOL6 A_24_P941167 1.5528797 1.6845138 apolipoprotein L6
COQ10B A_32_P1192 1.1098125 1.4745875 coenzyme Q10B
DUSP1 A_23_P110712 2.2050261 1.7767065 dual specificity phosphatase 1
EMP1 A_24_P921446 1.293242 1.2217363 epithelial membrane protein 1
FAM20A A_32_P108254 1.9548785 1.5243216 family with sequence similarity 20 member A
FAM20A A_24_P352952 2.080183 1.6580721 family with sequence similarity 20 member A
FCAR A_24_P348265 1.5712738 1.1219079 Fc fragment of IgA receptor
FOS A_23_P106194 2.8185577 2.8510469 Fos proto-oncogene, AP-1 transcription factor subunit
GSTT2 A_23_P109427 1.076046 1.2821832 glutathione S-transferase theta 2 (gene/pseudogene)
JUN A_23_P201538 2.3210056 2.2680829 Jun proto-oncogene, AP-1 transcription factor subunit
JUNB A_24_P241815 1.4505145 1.4068335 JunB proto-oncogene, AP-1 transcription factor subunit
LMNB1 A_23_P258403 1.2516638 1.6254543 lamin B1
MBD2 A_24_P119201 1.6597378 1.6676309 methyl-CpG binding domain protein 2
PDAP1 A_23_P151198 1.1131674 1.3884275 PDGFA associated protein 1
PDK4 A_23_P257087 1.089457 1.6686671 pyruvate dehydrogenase kinase 4
PDK4 A_24_P243749 1.0776004 1.6681819 pyruvate dehydrogenase kinase 4
PEAK1 A_24_P933801 1.1617162 1.2324933 pseudopodium enriched atypical kinase 1
PER1 A_24_P93916 2.5128018 1.6856721 period circadian clock 1
RHOB A_23_P51136 1.8707196 1.4680517 ras homolog family member B
SERPINB2 A_23_P153185 1.075777 1.0828608 serpin family B member 2
| Gene Symbol | ID   | log FC  | Gene. title                                      |
|------------|------|---------|------------------------------------------------|
| SLED1      | A24_P021766 | 2.2508577 | proteoglycan 3 pseudogene                        |
| SOCS3      | A23_P07058 | 2.1650652 | suppressor of cytokine signaling 3               |
| SOCS3      | A23_P051069 | 1.9930166 | suppressor of cytokine signaling 3               |
| SOD2       | A24_P035819 | 2.1525682 | superoxide dismutase 2, mitochondrial            |
| SRO              | A23_P152549 | 2.1742154 | serglycin                                       |
| STAC3      | A23_P10947 | 1.1170688 | SH and cysteine rich domain 3                    |
| STAT3      | A23_P253962 | 1.0796424 | signal transducer and activator of transcription 3 |
| TRIM64     | A32_P126079 | 1.2880819 | testis expressed 14, intercellular bridge forming factor |
| TRIM51     | A32_P118791 | 1.4412015 | transmembrane protein 107                       |

### Downregulated genes

| Gene Symbol | ID   | log FC  | Gene. title                                      |
|------------|------|---------|------------------------------------------------|
| ABCC13     | A23_P397480 | -1.951228 | ATP binding cassette subfamily C member 13 (pseudogene) |
| C10orf82   | A23_P1286 | -1.867092 | chromosome 10 open reading frame 82              |
| C1orf186   | A23_P05640 | -1.620385 | chromosome 1 open reading frame 186             |
| CACNG6     | A23_P051933 | -2.489393 | calcium voltage-gated channel auxiliary subunit gamma 6 |
| CACNG6     | A23_P051933 | -2.489393 | calcium voltage-gated channel auxiliary subunit gamma 6 |
| CCDC85A    | A23_P349566 | -1.066576 | coiled-coil domain containing 85A                |
| CCR3       | A23_P250302 | -2.382942 | C-C motif chemokine receptor 3                   |
| CCR3       | A23_P367473 | -1.410334 | C-C motif chemokine receptor 3                   |
| CD24       | A23_P082520 | -1.011483 | CD24 molecule                                    |
| CLC        | A23_P101683 | -2.57649 | Charcot-Leyden crystal galectin                  |
| CPA3       | A23_P18017  | -2.455389 | carboxypeptidase A3                              |
| ENPP3      | A23_P404536 | -1.88905 | ectonucleotide pyrophosphatase/phosphodiesterase 3 |
| EPB42      | A23_P140675 | -1.985456 | erythrocyte membrane protein band 4.2            |
| FCER1A     | A23_P037065 | -1.361214 | Fc fragment of IgE receptor 1A                   |
| FCRLA      | A23_P276576 | -1.056556 | Fc receptor like A                               |
| GPR37      | A23_P145905 | -1.190649 | G protein-coupled receptor 37                    |
| HBA1       | A23_P037856 | -2.165281 | hemoglobin subunit alpha 1                       |
| HBA2       | A23_P142305 | -2.755512 | hemoglobin subunit alpha 2                       |
| HBD        | A23_P075190 | -2.138927 | hemoglobin subunit delta                        |
| HBQ1       | A23_P49254  | -2.358415 | hemoglobin subunit theta                         |
| HDC        | A23_P117662 | -2.252604 | histidine decarboxylase                          |
| HRH4       | A23_P386310 | -1.84512 | histamine receptor H4                            |
| IL4        | A23_P213706 | -2.436502 | interleukin 4                                   |
| ITGBB      | A23_P075477 | -2.315214 | integrin subunit beta 8                         |
| LTK        | A23_P14853 | -1.084265 | leukocyte receptor tyrosine kinase               |
| MME        | A23_P213706 | -2.436502 | interleukin 4                                   |
| MMP26D     | A23_P52888 | -1.656583 | metallophosphoesterase domain containing 2       |
| MS4A2      | A23_P1904  | -2.626586 | membrane spanning 4-domains A2                   |
| PAGE2      | A23_P070927 | -1.139398 | PAG family member 2                             |
| PAGE2B     | A23_P09683 | -1.185484 | PAG family member 2B                             |
| PAGE5      | A23_P22744  | -1.058192 | PAG family member 5                             |
| RAB3J      | A23_P180920 | -1.021236 | RAB3A interacting protein                        |
| RBM20      | A24_P005320 | -1.116389 | RNA binding motif protein 20                     |
| RNFI12     | A23_P092525 | -2.692754 | ring finger protein 182                         |
| SLC45A3    | A23_P208345 | -1.514775 | solute carrier family 45-member 3               |
| SLC4A11    | A23_P093111 | -2.103964 | solute carrier family 4 member 10               |
| SLC4A10    | A23_P314786 | -1.750496 | solute carrier family 4 member 10               |
| THSD7A     | A23_P400324 | -1.690056 | thrombospondin type 1 domain containing 7A       |
| TRIM49     | A23_P1575  | -2.460168 | tripartite motif containing 49                   |
| TRIM51     | A23_P15083 | -2.767962 | tripartite motif-containing 51                   |
| TRIM53AP   | A23_P16535 | -2.039508 | tripartite motif containing 53A, pseudogene     |
| TRIM64     | A23_P12972 | -2.396233 | tripartite motif containing 64                   |
| TRPM6      | A23_P216712 | -2.766311 | transient receptor potential cation channel subfamily M member 6 |
| UGT2B11    | A23_P212968 | -1.109849 | UDP glucuronosyltransferase family 2 member B1    |
| UGT2B7     | A23_P136671 | -2.182992 | UDP glucuronosyltransferase family 2 member B7    |
| VWCE       | A23_P05966 | -1.113653 | von Willebrand factor C and EGF domains          |
| ZNF91      | A23_P209746 | -1.06124 | zinc finger protein 91                          |
3.2. Go and Pathway Enrichment Analysis of DEGs

To analyze the functions and mechanisms of DEGs, the functional and pathway enrichment analyses of upregulated and downregulated DEGs were performed using the DAVID 6.8 online tool. In our study, a total of 51 GO terms and 8 pathways of DEGs (P-value <0.05), including 27 BPs, 12 CCs, and 12 MFs, were obtained, and the top five of each item are shown in Table 3. GO analysis results showed that changes in BPs of upregulated DEGs were significantly enriched in response to cAMP, positive regulation of cell differentiation, cellular response to calcium ion, response to drug, and regulation of cell cycle. Downregulated DEGs in BPs were significantly enriched in oxygen transport, bicarbonate transport, positive regulation of mast cell degranulation, angiotensin maturation, and positive regulation of cytosolic calcium ion concentration. Changes in CCs of upregulated DEGs were mainly enriched in nuclear chromatin, lamin filament, cilary transition zone, transcription factor complex, and nuclear inner membrane. Downregulated DEGs in CCs were mainly enriched in hemoglobin complex, haptoglobin-hemoglobin complex, integral component of plasma membrane, endocytic vesicle lumen, and blood microparticle. Changes in MFs of upregulated DEGs were mainly enriched in RNA polymerase II core promoter proximal region sequence-specific DNA binding, transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding, transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding, and transcription factor binding. Changes in MFs of downregulated DEGs were mainly enriched in oxygen transporter activity, oxygen binding, heme binding, iron ion binding, and haptoglobin binding (Table 3).

The pathways enriched by upregulated DEGs were mainly related to the TNF signaling pathway, Osteoclast differentiation, Herpes simplex infection, Prolactin signaling
pathway, and Hepatitis B. The pathways enriched by downregulated DEGs were mainly related to Asthma, Fc epsilon RI signaling pathway, and Hematopoietic cell lineage (Table 3).

Table 3. The top five GO and pathway enrichment analysis of DEGs.

| Category           | Term                  | Description                                               | Count | P. Value |
|--------------------|-----------------------|-----------------------------------------------------------|-------|----------|
| Upregulated genes  | GOTERM_BP: GO:0051591 | response to cAMP                                          | 3     | 0.000928 |
|                    | GOTERM_BP: GO:0045597 | positive regulation of cell differentiation               | 3     | 0.0003071|
|                    | GOTERM_BP: GO:0071277 | cellular response to calcium ion                           | 3     | 0.0010256|
|                    | GOTERM_BP: GO:0042493 | response to drug                                          | 3     | 0.0055356|
|                    | GOTERM_BP: GO:0051726 | regulation of cell cycle                                  | 3     | 0.0055356|
|                    | GOTERM_CC: GO:0000790 | nuclear chromatin                                         | 4     | 0.0023008|
|                    | GOTERM_CC: GO:00005638| lammin filament                                           | 2     | 0.0086248|
|                    | GOTERM_CC: GO:00035869| ciliary transition zone                                   | 2     | 0.0273487|
|                    | GOTERM_CC: GO:00025667| transcription factor complex                              | 3     | 0.0309618|
|                    | GOTERM_CC: GO:00005637| nuclear inner membrane                                    | 2     | 0.0374182|
|                    | GOTERM_MF: GO:0000978 | RNA polymerase II core promoter proximal region sequence-specific DNA binding | 6     | 0.0011178|
|                    | GOTERM_MF: GO:0000982 | transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding | 2     | 0.0293031|
|                    | GOTERM_MF: GO:0001077 | transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding | 3     | 0.0412042|
|                    | GOTERM_MF: GO:0008134 | transcription factor binding                               | 2     | 0.0483833|
|                    | KEGG_PATHWAY: ptr04668| TNF signaling pathway                                     | 4     | 0.0003942|
|                    | KEGG_PATHWAY: ptr04380| Osteoclast differentiation                                | 4     | 0.0007797|
|                    | KEGG_PATHWAY: ptr05168| Herpes simplex infection                                  | 4     | 0.0022599|
|                    | KEGG_PATHWAY: ptr04917| Prolactin signaling pathway                               | 3     | 0.0042762|
|                    | KEGG_PATHWAY: ptr05161| Hepatitis B                                              | 3     | 0.0186064|
| Downregulated genes| GOTERM_BP: GO:0015671 | oxygen transport                                          | 4     | 0.0000031|
|                    | GOTERM_BP: GO:0015701 | bicarbonate transport                                     | 3     | 0.0033648|
|                    | GOTERM_BP: GO:0043306 | positive regulation of mast cell degranulation            | 2     | 0.0214126|
|                    | GOTERM_BP: GO:0002003 | angiotensin maturation                                    | 2     | 0.0214126|
|                    | GOTERM_BP: GO:0007204 | positive regulation of cytosolic calcium ion concentration | 3     | 0.0284004|
|                    | GOTERM_CC: GO:0005833 | hemoglobin complex                                        | 4     | 0.0000015|
|                    | GOTERM_CC: GO:0031838 | haptoglobin-hemoglobin complex                            | 2     | 0.0078789|
|                    | GOTERM_CC: GO:0005887 | integral component of plasma membrane                     | 8     | 0.0190709|
|                    | GOTERM_CC: GO:0071682 | endocytic vesicle lumen                                   | 2     | 0.0311553|
|                    | GOTERM_CC: GO:0072562 | blood microparticle                                       | 3     | 0.0361847|
|                    | GOTERM_MF: GO:0005344 | oxygen transporter activity                                | 4     | 0.0000032|
|                    | GOTERM_MF: GO:0019825 | oxygen binding                                            | 4     | 0.0001354|
|                    | GOTERM_MF: GO:0020037 | heme binding                                              | 4     | 0.0036095|
|                    | GOTERM_MF: GO:0005506 | iron ion binding                                          | 4     | 0.0041867|
|                    | GOTERM_MF: GO:0031720 | haptoglobin binding                                       | 2     | 0.0063845|
|                    | KEGG_PATHWAY: hsa05310| Asthma                                                    | 3     | 0.0026938|
|                    | KEGG_PATHWAY: hsa04664| Fc epsilon RI signaling pathway                           | 3     | 0.0133022|
|                    | KEGG_PATHWAY: hsa04640| Hematopoietic cell lineage                                | 3     | 0.0212146|

If there were more than five terms enriched in this category, the top five terms were selected according to P-value.

3.3. PPI Network Construction and Hub Gene Identification

Protein interactions among the DEGs were predicted with STRING online database. A PPI network with 69 nodes and 68 edges was obtained and the PPI network was visualized by Cytoscape (Figure 2B). The cytoHubba plugin was then used to analyze hub genes with MCC, and genes with the top 10 scores were identified as hub genes (Figure 2C). As shown in Figure 2C, six upregulated genes (JUN, FOS, STAT3, SOCS3, JUNB, DUSP1) and four downregulated genes (IL4, FCER1A, MS4A2, CPA3) were identified. The gene symbols, full names, and scores of hub genes are shown in Table 4.
Bioinformatics Analysis Identifies Potential Key Genes of Peripheral Blood Mononuclear Cell in Idiopathic Pulmonary Fibrosis

A

B
Figure 2. PPI network construction and hub gene identification. (A) PPI network for DEGs. (B) Cytoscape network visualization of the 69 nodes and 68 edges that were obtained with interaction scores > 0.4 according to the STRING online database. The nodes represent genes, and the edges represent links between genes. Red represents upregulated genes, and green represents downregulated genes. (C) The hub genes with the top 10 scores.

Table 4. The top 10 hub genes with highest scores.

| Gene symbol | Full name                                                  | Score |
|-------------|------------------------------------------------------------|-------|
| JUN         | Jun proto-oncogene, AP-1 transcription factor subunit      | 267   |
| FOS         | Fos proto-oncogene, AP-1 transcription factor subunit      | 266   |
| STAT3       | signal transducer and activator of transcription 3         | 244   |
| SOCS3       | suppressor of cytokine signaling 3                         | 241   |
| JUNB        | JunB proto-oncogene, AP-1 transcription factor subunit     | 240   |
| IL4         | interleukin 4                                              | 148   |
| DUSP1       | dual specificity phosphatase 1                             | 120   |
| FCER1A      | Fc fragment of IgE receptor Ia                             | 37    |
| MS4A2       | membrane spanning 4-domains A2                            | 30    |
| CPA3        | carboxypeptidase A3                                        | 14    |

3.4. Go and Pathway Enrichment Analysis of Hub Genes

We performed a functional enrichment analysis for hub genes. The GO analysis demonstrated that changes in BPs were mainly enriched in cellular response to calcium ion, regulation of cell cycle, positive regulation of mast cell degranulation, response to muscle stretch, B cell activation, positive regulation of pri-miRNA transcription from RNA polymerase II promoter, positive regulation of cell differentiation, SMAD protein signal transduction, transforming growth factor beta receptor signaling pathway, and response to drug. Changes in CCs were significantly enriched in external side of plasma membrane and nucleoplasm. Changes in MFs for the hub genes were enriched mainly in transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding, and RNA polymerase II core promoter proximal region sequence-specific DNA binding.

The pathways enriched by hub genes were mainly related to Asthma, Fc epsilon RI signaling pathway, Inflammatory bowel disease (IBD), Leishmaniasis, T cell receptor signaling pathway, TNF signaling pathway, Osteoclast differentiation, Jak-STAT signaling pathway, Prolactin signaling pathway, Measles, Hepatitis B, Herpes simplex infection, and MAPK signaling pathway (Table 5).
3.5. Using Hub Genes for IPF Diagnosis

The diagnostic accuracy of the top 10 hub genes was assessed using ROC curve analysis (Figure 3). The areas under the ROC curves were 0.9797, 0.9629, 0.9654, 0.9687, and 0.9154 for JUN, FOS, JUNB, SOCS3, and STAT3, as shown in Figure 3A. The areas under the ROC curves were 0.9318, 0.9501, 0.8805, 0.932, and 0.9365 for DUSP1, IL4, FCER1A, MS4A2, and CPA3, as shown in Figure 3B.

![Figure 3. Validation of the diagnostic value of the hub genes for IPF. (A-B) Receiver operating characteristic curve of the hub genes for diagnosis of IPF.](image)

4. Discussion

In this study, we analyzed the DEGs in PBMCs from IPF patients and healthy controls. We performed independently for the male IPF and female IPF patients to pinpoint the potential genes. The results of the microarray analysis revealed the expression of 72 DEGs (28 upregulated genes and 44 downregulated genes). The associations between these genes were revealed by constructing a PPI network. The top 10 genes with the highest scores were identified, including JUN, FOS, STAT3, SOCS3, JUNB, DUSP1, IL4, FCER1A, MS4A2, and CPA3. There were six upregulated genes (JUN, FOS, STAT3, SOCS3, JUNB, DUSP1) and four downregulated genes (IL4,
bleomycin-induced fibrosis and pulmonary fibrosis was used as a biomarker in IPF. JUN and FOS are upregulated genes in PBMC of IPF. JUN and FOS are a subunit of the activator protein-1 (AP-1) [14]. AP-1 is a dimeric complex composed of the JUN (c-JUN, JUND, JUNB), FOS, ATF, and MAF protein families [14-16]. It can be seen that JUNB is a member of the JUN protein. Moreover, JUN and FOS are important members of the transcription AP-1 [17]. One study has reported that AP-1 induction may be associated with increased proliferation and extracellular matrix (ECM) production in IPF fibroblasts [18]. And Werning et al. have recently found that c-JUN and FOS are expressed in fibroblasts of IPF [19]. Besides, Chang et al. have shown that JUNB can regulate Epithelial-to-mesenchymal transition (EMT) [20]. One study has shown that EMT plays an important role in the pathogenesis of IPF [21]. Therefore, JUN, FOS, and JUNB may be involved in forming IPF.

STAT3 is a ubiquitously expressed latent cytoplasmic protein that regulates lung fibrosis [22]. Waters et al. have shown that STAT3 can promote fibroblast senescence to promote fibrosis [23]. Milara et al. have found that lung from patients with IPF expressed higher levels of STAT3, as well as phosphorylated [24]. Recently, it has been demonstrated that STAT3 regulates lung fibroblast-myofibroblast activation and differentiation in IPF [25]. Given these findings, STAT3 can be used as a biomarker in IPF.

SOCS3 is one of the most studied members of the SOCS family that consists of eight proteins (SOCS1-7 and cytokine-inducible SH2-containing protein, CISH) [26]. SOCS3 can be a major regulator of STAT3 activation [27]. One study has found that SOCS3 expression was shown to be elevated for up to 30 days in bleomycin-induced fibrosis [28]. A study by Akram et al. has shown that a significant increase in SOCS3 expression was observed in IPF AEC and macrophages compared to control lung tissue using dual immunohistochemical analysis [29]. Shocher et al. have found that primary human fibroblast culture from IPF (IPF-HLFs) expressed higher levels of SOCS3 when tested basal levels in HLFs [30]. Therefore, it may be hypothesized that SOCS3 may play a significant role in IPF.

DUSP1, also named mitogen-activated protein kinase (MAPK) phosphatase (MKP-1) that dephosphorylates and deactivates MAPKs, acts as a negative regulator of the MAPK signaling pathway [31, 32]. Redente et al. have shown that DUSP1-deficient mice reduced pulmonary fibrosis in bleomycin-induced fibrosis and pulmonary fibrosis was attenuated in mice given bleomycin using DUSP1 inhibitors [33]. Besides, one study has found that DUSP1 plays a critical role in promoting pulmonary fibrosis from macrophages to fibroblasts in vivo experiments [34]. These studies suggest that DUSP1 plays a crucial role in IPF and maybe a relevant therapeutic target.

IL-4, FCER1A, MS4A2, and CPA3 are downregulated genes in PBMC of IPF. IL-4 is a fibrogenic cytokine that increases collagen production by fibroblasts [35]. But one study has found that pulmonary fibrosis and lung injury and inflammation in the bleomycin-induced fibrosis model of IL-4-deficient mice were substantially less than wild-type mice [36]. These results suggest that IL-4 has both fibrogenic and anti-inflammatory in the context of bleomycin-induced lung fibrosis and injury. FCER1A gene encodes the a-subunit (FCER1a) of the high-affinity IgE receptor consisting of an α-chain (FCER1), an β-chain (MS4A2) and two γ-chains (FCER1G) [37-39]. FCER1 and MS4A2 expressing in mast cells (MCs) are associated with asthma [37, 38, 40]. But CPA3 is one of the MC-restricted proteases that are secreted by MCs [41]. It is also associated with asthma [41]. What’s more, it is demonstrated that FCER1A, MS4A2, and CPA3 may be favorable prognostic indicators in non-small cell lung cancer [42, 43]. However, there is no research on the relationship between FCER1A, MS4A2, and CPA3 and IPF so far. Based on our study, we can predict that FCER1A, MS4A2, and CPA3 may be PBMC markers in IPF.

In our study, all the AUC values of the top 10 genes were in the range 0.880-0.980 concerning the ROC curve. These genes possess high accuracy except that FCER1A indicated moderate accuracy [44]. KEGG enrichment analysis of these genes showed that these genes were mainly linked to the Fc epsilon RI signaling pathway, TNF signaling pathway, Jak-STAT signaling pathway, and MAPK signaling pathway. These signaling pathways play an essential role in the pathogenesis of IPF. We speculated that these genes might play an important role in IPF. But our study has a few limitations. Firstly, to fully identify the key genes in PBMC of IPF, it is better to combine venous blood samples and lung tissue to explore. Second, the sample size of the dataset we explored was too small. Therefore, it is necessary to increase the samples to improve the diagnostic accuracy of these hub genes in IPF. Third, a single microarray analysis was used in our study. It may result in a high false-positive rate. Thus, it is necessary to improve the detection capability by combining multiple individual data in future studies. What’s more, some key genes and pathways were not found in IPF in previous studies. For this reason, we need more experimental evidence to prove the relationship between these key genes and IPF.

5. Conclusion

Our study used bioinformatics analysis to identify the associated biological functions and pathways involved in IPF to explore the pathogenesis, diagnosis, and prognosis of IPF. Moreover, we identified 10 key genes as potential diagnostic biomarkers through PPI network analysis and the ROC curve analysis. However, more experiments are needed to validate the results of our study further.

Acknowledgements

The authors would like to express their sincere thanks to all those who have lent them hands in writing this paper.
References

[1] Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med, 2011, 183 (6): 788-824.

[2] Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. N Engl J Med, 2018, 378 (19): 1811-1823.

[3] Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. Lancet, 2017, 389 (10082): 1941-1952.

[4] Hutchinson J, Fogarty A, Hubbard R, et al. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. Eur Respir J, 2015, 46 (3): 795-806.

[5] American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med, 2002, 165 (2): 277-304.

[6] Sgalla G, Iovene B, Calvello M, et al. Idiopathic pulmonary fibrosis: pathogenesis and management. Respir Res, 2018, 19 (1): 32.

[7] Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. Annu Rev Pathol, 2014, 9: 157-79.

[8] Wolters PJ, Blackwell TS, Eickelberg O, et al. Time for a change: is idiopathic pulmonary fibrosis still idiopathic and only fibrotic? Lancet Respir Med, 2018, 6 (2): 154-160.

[9] Qiu CC, Su QS, Zhu SY, et al. Identification of Potential Biomarkers and Biological Pathways in Juvenile Dermatomyositis Based on miRNA-mRNA Network. Biomed Res Int, 2019, 2019: 7814287.

[10] Zhang G, Xu S, Zhang Z, et al. Identification of Key Genes and the Pathophysiology Associated With Major Depressive Disorder Patients Based on Integrated Bioinformatics Analysis. Front Psychiatry, 2020, 11: 192.

[11] Yi XH, Zhang B, Fu YR, et al. STAT1 and its related molecules as potential biomarkers in Mycobacterium tuberculosis infection. J Cell Mol Med, 2020, 24 (5): 2866-2878.

[12] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics, 2000, 25 (1): 25-29.

[13] Kancheisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 2000, 28 (1): 27-30.

[14] Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nat Rev Cancer, 2003, 3 (11): 859-68.

[15] Shaulian E. AP-1 — The Jun proteins: Oncogenes or tumor suppressors in disguise? Cellular Signalling, 2010, 22 (6): 894-899.

[16] Gazon H, Barbeau B, Mesnard JM, et al. Hijacking of the AP-1 Signaling Pathway during Development of ATL. Front Microbiol, 2017, 8: 2686.

[17] Jiang X, Xie H, Dou Y, et al. Expression and function of FRA1 protein in tumors. Mol Biol Rep, 2020, 47 (1): 737-752.

[18] Klymenko O, Huehn M, Wilhelm J, et al. Regulation and role of the ER stress transcription factor CHOP in alveolar epithelial type-II cells. J Mol Med (Berl), 2019, 97 (7): 973-990.

[19] Wernig G, Chen SY, Cui L, et al. Unifying mechanism for different fibrotic diseases. Proc Natl Acad Sci U S A, 2017, 114 (18): 4757-4762.

[20] Chang H, Liu Y, Xue M, et al. Synergistic action of master transcription factors controls epithelial-to-mesenchymal transition. Nucleic Acids Res, 2016, 44 (6): 2514-27.

[21] Salton F, Volpe MC, Confalonieri M. Epithelial-Mesenchymal Transition in the Pathogenesis of Idiopathic Pulmonary Fibrosis. Medicina (Kaunas), 2019, 55 (4).

[22] Knight D, Mutsaers SE, Prêle CM. STAT3 in tissue fibrosis: is there a role in the lung? Pulm Pharmacol Ther, 2011, 24 (2): 193-8.

[23] Waters DW, Blokland KEC, Pathinayake PS, et al. Fibroblast senescence in the pathology of idiopathic pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol, 2018, 315 (2): L162-172.

[24] Milara J, Hernandez G, Ballester B, et al. The JAK2 pathway is activated in idiopathic pulmonary fibrosis. Respir Res, 2018, 19 (1): 24.

[25] Peckovský DV, Prêle CM, Wong J, et al. STAT3-mediated signaling dysregulates lung fibroblast-myofibroblast activation and differentiation in UIP/IPF. Am J Pathol, 2012, 180 (4): 1398-412.

[26] Yoshimura A, Naka T, Kubo M. SOCS proteins, cytokine signalling and immune regulation. Nat Rev Immunol, 2007, 7 (6): 454-65.

[27] Rottenberg ME, Carow B, SOCS3 and STAT3, major controllers of the outcome of infection with Mycobacterium tuberculosis. Semin Immunol, 2014, 26 (6): 518-32.

[28] O’Donoghue RJ, Knight DA, Richards CD, et al. Genetic partitioning of interleukin-6 signalling in mice dissociates Stat3 from Smad3-mediated lung fibrosis. EMBO Mol Med, 2012, 4 (9): 939-51.

[29] Akram KM, Lomas NJ, Forsyth NR, et al. Alveolar epithelial cells in idiopathic pulmonary fibrosis display upregulation of TRAIL, DR4 and DR5 expression with simultaneous preferential over-expression of pro-apoptotic marker p53. Int J Clin Exp Pathol, 2014, 7 (2): 552-64.

[30] Shochet GE, Brook E, Bardenstein-Wald B, et al. TGF-β pathway activation by idiopathic pulmonary fibrosis (IPF) fibroblast derived soluble factors is mediated by IL-6 trans-signaling. Respir Res, 2020, 21 (1): 56.

[31] Moosavi SM, Prabhala P, Ammit AJ. Role and regulation of MKP-1 in airway inflammation. Respir Res, 2017, 18 (1): 154.

[32] Keyse SM. Dual-specificity MAP kinase phosphatases (MKPs) and cancer. Cancer Metastasis Rev, 2008, 27 (2): 253-61.

[33] Redente EF, Black BP, Aguilar MA, et al. DUSP1 inhibition impairs fibrosis development by altering macrophage programming. American Journal of Respiratory and Critical Care Medicine, 2017, 195.

[34] Goda C, Balli D, Black M, et al. Loss of FOXM1 in macrophages promotes pulmonary fibrosis by activating p38 MAPK signaling pathway. PLoS Genetics, 2020, 16 (4).
[35] Tsoutsou PG, Gourgoulianis KI, Petinaki E, et al. Cytokine levels in the sera of patients with idiopathic pulmonary fibrosis. Respir Med, 2006, 100 (5): 938-45.

[36] Huaux F, Liu T, McGarry B, et al. Dual roles of IL-4 in lung injury and fibrosis. J Immunol, 2003, 170 (4): 2083-92.

[37] Kraft S, Kinet JP. New developments in FcepsilonRI regulation, function and inhibition. Nat Rev Immunol, 2007, 7 (5): 365-78.

[38] Potaczek DP, Michel S, Sharma V, et al. Different FCER1A polymorphisms influence IgE levels in asthmatics and non-asthmatics. Pediatr Allergy Immunol, 2013, 24 (5): 441-9.

[39] Wu LC, Zarrin AA. The production and regulation of IgE by the immune system. Nat Rev Immunol, 2014, 14 (4): 247-59.

[40] Ferreira MA, Zhao ZZ, Thomsen SF, et al. Association and interaction analyses of eight genes under asthma linkage peaks. Allergy, 2009, 64 (11): 1623-8.

[41] Pejler G. The emerging role of mast cell proteases in asthma. Eur Respir J, 2019, 54 (4).

[42] Ly D, Zhu CQ, Cabanero M, et al. Role for High-Affinity IgE Receptor in Prognosis of Lung Adenocarcinoma Patients. Cancer Immunol Res, 2017, 5 (9): 821-829.

[43] Visser S, Hou J, Bezemer K, et al. Prediction of response to pemetrexed in non-small-cell lung cancer with immunohistochemical phenotyping based on gene expression profiles. BMC Cancer, 2019, 19 (1): 440.

[44] Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr, 2007, 96 (5): 644-7.