Int J Clin Exp Pathol 2020;13(7):1698-1706
www.ijcep.com /ISSN:1936-2625/IJCEP0107763

Original Article

Identification of transcriptomic markers for developing idiopathic pulmonary fibrosis: an integrative analysis of gene expression profiles

Diandian Li[1], Yi Liu[2], Bo Wang[1]

[1]*Department of Respiratory and Critical Care Medicine, West China Hospital of Sichuan University, Chengdu 610041, China;* [2]*West China School of Medicine, Sichuan University, Chengdu 610041, China*

Received January 12, 2020; Accepted March 6, 2020; Epub July 1, 2020; Published July 15, 2020

Abstract: Idiopathic pulmonary fibrosis (IPF) remains a lethal disease with unknown etiology and unmet medical need. The aim of this study was to perform an integrative analysis of multiple public microarray datasets to investigate gene expression patterns between IPF patients and healthy controls. Moreover, functional interpretation of differentially expressed genes (DEGs) was performed to assess the molecular mechanisms underlying IPF progression. DEGs between IPF and normal lung tissues were picked out by GEO2R tool and Venn diagram software. Database for Annotation, Visualization and Integrated Discovery (DAVID) was applied to analyze gene ontology (GO) and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway. Protein-protein interaction (PPI) of these DEGs was visualized by Cytoscape with Search Tool for the Retrieval of Interacting Genes (STRING). 5520 DEGs were identified in IPF based on six profile datasets, including 3714 up-regulated genes and 1806 down-regulated genes. Using Venn software, a total of 367 commonly altered DEGs were revealed, including 259 up-regulated genes mostly enriched in collagen catabolic process, heparin binding, and the extracellular region. For pathway analysis, up-regulated DEGs were mainly enriched in ECM-receptor interaction, protein digestion and absorption, and focal adhesion. Finally, 24 DEGs with degrees ≥10 were screened as hub genes from the PPI network, which were enriched in protein digestion and absorption, ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, amoebiasis, and platelet activation. The present integrative study identified DEGs and hub genes that may be diagnostic biomarkers or therapeutic targets, and provide novel insights into the pathogenesis of IPF.

Keywords: IPF, differentially expressed gene, co-expression network, transcriptomic markers, bioinformatic analysis

Introduction

Idiopathic interstitial pneumonia (IIP), with unknown etiology, has major implications for prognosis and management [1]. Idiopathic pulmonary fibrosis (IPF), a chronic and progressive interstitial fibrotic lung disease, where no cause can be identified, is one of the most common types of IIP and remains a fatal disease with a median survival time from diagnosis of 2 to 5 years [2]. Although the prognosis is poor, it is difficult to predict due to limited clinical biomarkers reliably reflecting disease progression. With regard to management, limited pharmacologic options--pirfenidone and nintedanib are supposed to be promising for IPF based on clinical trials [3], but the long-term efficacy is still unclear. Particularly, patients with severe IPF often develop acute exacerbations resulting in the rapid deterioration of lung function, requiring lung transplantation [4]. Thus, it is essential to elucidate the mechanisms of IPF pathogenesis and explore potential biomarkers for improving the treatment effect of IPF.

Currently, studies have demonstrated that environmental (such as smoking, metal or wood dust, sand, and spores from soil) and genetic factors variably contribute to disease susceptibility and outcomes [5]. A number of common gene variants with modest effect size increase the risk of disease in patients with sporadic IPF, while rare variants with large

effect size influence disease risk in patients with familial interstitial pneumonia [6]. In the past decades, rapid growth of high-throughput transcriptomic data largely enables quickly identification of differentially expressed genes (DEGs) in diseases and has been proved to be a reliable technique [7]. There are a large number of microarray gene expression datasets that are publicly available from the Gene Expression Omnibus (GEO) database, and many bioinformatic studies on IPF have shown that DEGs and biologic functional pathways participated in the development of this disease. However, single microarray analysis often brings a high false-positive rate, while analysis of multiple transcriptomic datasets may shed light on discovering robust candidates for diagnosis and treatment.

Therefore, the aim of present study was to perform an integrative analysis of multiple public microarray datasets and investigate gene expression patterns in IPF lung tissue in IPF patients and healthy controls. Moreover, DEGs identified in this study were further interpreted by enrichment analysis and co-expression network construction to assess the molecular mechanisms underlying IPF progression.

Materials and methods

Datasets search and selection

Microarray data from IPF-related mRNA expression profiles were retrieved and downloaded from the National Center for Biotechnology Information (NCBI) GEO database (http://www.ncbi.nlmNih.gov/geo). We searched public microarray datasets till Jun 1, 2019 using the keyword “idiopathic pulmonary fibrosis” or “IPF” restricted to Homo sapiens. The datasets obtained were further selected by our inclusion criteria: (1) case-control study; (2) sample size larger than 5 in each group; (3) dataset using lung tissues for gene expression analysis; (4) raw data or processed data were available in these datasets; (5) subjects 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. Two investigators (J.H. and Y.L.) extracted the data and reached a consensus on all items. Data retrieved from the dataset included GEO accession, publication year, sample size, sample source, platform, and raw gene expression data. Ethical approval for this study was not required because the data were freely available in public datasets.

Data processing of DEGs

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/), a web tool, can perform sophisticated R-based analyses of GEO data and presents the results as a table of DEGs [8]. In the present study, GEO2R online tools were applied to identify DEGs in lung tissues between IPF and healthy controls with absolute $\log_2$ Fold Change (logFC) >1 and adjusted *P*-value < 0.05. The DEGs with logFC >1 were considered as up-regulated genes, while the DEGs with logFC <-1 were considered as down-regulated genes [9]. The raw data downloaded in TXT format were imported in Venn software online to detect the commonly DEGs among the six datasets.

Gene ontology and pathway enrichment analysis

Functional interpretation/gene ontology (GO) biological process analysis and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway analysis) of DEGs was further performed using Database for Annotation, Visualization and Integrated Discovery (DAVID 6.8; available online: http://david.ncifcrf.gov). This online bioinformatic tool was designed to identify a large number of genes or proteins function [10]. By gene ontology (GO) analysis of high-throughput transcriptome or genome data, genes and their RNA or protein product can be defined to identify unique biologic properties [11]. KEGG is a collection of databases dealing with genomes, diseases, biologic pathways, drugs, and chemical materials [12]. We used DAVID to visualize the DEGs enrichment of biologic process (BP), molecular function (MF), cellular component (CC) and pathways ($P<0.05$).

Protein-protein interaction network construction

Information of protein-protein interaction (PPI) of the DEGs was evaluated by STRING (Search Tool for the Retrieval of Interacting Genes) online tool [13], with interactions of a combined score >0.4 considered statistically significant [14]. Cytoscape (version 3.6.1) was used to visualize molecular interaction networks

Table 1. Microarray datasets included in this analysis

GEO accession	Publication year	Sample size		Source	Platform
		IPF	Healthy control		
GSE110147	2018	22	11	Lung tissue	Affymetrix Human Gene 1.0 ST Array
GSE53845	2014	40	8	Lung tissue	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F
GSE24206	2011	11	6	Lung tissue	Affymetrix Human Genome U133 Plus 2.0 Array
GSE10667	2009	23	15	Lung tissue	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F
GSE32537	2013	119	50	Lung tissue	Affymetrix Human Gene 1.0 ST Array
GSE2052	2005	13	11	Lung tissue	Amersham Biosciences CodeLink Uniset Human I Bioarray

GEO, Gene Expression Omnibus; IPF, idiopathic pulmonary fibrosis.

[15]. In addition, the MCODE app in Cytoscape was used to check the most significant module in the PPI networks (degree cutoff = 2, node score cutoff = 0.2, max depth = 100, and k-core = 2).

Results

Identification of DEGs in IPF

The detailed information of the six datasets is presented in **Table 1**. There were 228 lung tissues from IPF patients and 101 normal lung tissues in our study. By GEO2R tools, a total of 5520 DEGs were detected in six microarray datasets (GSE110147, GSE53845, GSE10667, GSE24206, GSE32537 and GSE2052) [16-21], including 3714 up-regulated genes (logFC >1) and 1806 down-regulated genes (logFC <1). Then, we identified 367 commonly DEGs (overlapped in at least three datasets) by Venn diagram software, among which 259 genes were up-regulated and 108 were down-regulated. Besides, 9 overlapping genes (MMP7, TRIM2, ASPN, SULF1, CXCL14, DCLK1, IL13RA2, TP63, CRTAC1) in the six datasets were also identified (Table S1).

GO enrichment and KEGG pathway analysis of DEGs

Advanced analyses were carried out for further functional investigation of the DEGs. GO analysis results showed that changes in BP of up-regulated DEGs were significantly enriched in the collagen catabolic process, extracellular matrix organization and cell adhesion, while down-regulated DEGs were particularly enriched in astrocyte development, angiogenesis, and response to hypoxia. For MF, up-regulated DEGs were mainly enriched in heparin binding, calcium ion binding and extracellular matrix structural constituent, and down-regulated DEGs were in calcium ion binding, calcium-dependent phospholipase A2 activity, and arachidonic acid binding. Moreover, changes in CC of up-regulated DEGs were mainly enriched in the extracellular region,space and matrix. **Table 2** presented a summary of the GO biologic process analysis results. Results of KEGG analysis are shown in **Table 3**, which suggest that up-regulated DEGs were mainly enriched in ECM-receptor interaction, protein digestion and absorption, and focal adhesion, while down-regulated DEGs were related to dilated cardiomyopathy and neuroactive ligand-receptor interaction.

PPI network and modular analysis

A total of 304 DEGs were imported into the PPI network complex of DEGs, including 304 nodes and 837 edges (Figure S1). MCODE app in Cytotype was used for further analysis, which identified 24 central nodes as hub genes (**Figure 1**).

Re-analysis of hub genes via KEGG pathway enrichment

To explore the biological classifications of hub genes, functional and pathway enrichment analyses were performed by DAVID. Results of GO analysis suggested that changes in the BP of hub genes were significantly enriched in collagen catabolic process, collagen fibril organization and cellular response to amino acid stimulus. Changes in the MF of hub genes were mainly enriched in extracellular matrix structural constituent and calcium ion binding. Changes in CC of hub genes were mainly enriched in proteinaceous extracellular matrix, collagen trimer and extracellular space. KEGG pathway analysis demonstrated that the hub genes were mainly enriched in protein digestion and absorption, ECM-receptor interaction, and focal adhesion (**Table 4**).

Table 2. Gene ontology analysis of differentially expressed genes (DEGs)

Expression	Category	Term	Count	*p*-Value	FDR
Up-regulated	GOTERM_BP_DIRECT	GO:0030574~collagen catabolic process	14	2.03E-12	3.31E-09
		GO:0030198~extracellular matrix organization	20	1.53E-11	2.49E-08
		GO:0007155~cell adhesion	28	9.35E-11	1.52E-07
		GO:0030199~collagen fibril organization	8	7.11E-07	0.001158
		GO:0001649~osteoblast differentiation	9	7.36E-05	0.119878
		GO:0001501~skeletal system development	10	9.00E-05	0.146568
	GOTERM_CC_DIRECT	GO:0005576~extracellular region	64	2.12E-15	2.69E-12
		GO:0005615~extracellular space	58	2.28E-15	2.96E-12
		GO:0031012~extracellular matrix	28	3.07E-15	3.95E-12
		GO:0005578~proteinaceous extracellular matrix	25	1.74E-13	2.22E-10
		GO:0005581~collagen trimer	13	3.25E-09	4.14E-06
		GO:0005788~endoplasmic reticulum lumen	13	1.03E-05	0.01307
	GOTERM_MF_DIRECT	GO:0008201~heparin binding	17	2.91E-10	3.94E-07
		GO:0005509~calcium ion binding	31	1.47E-08	1.99E-05
		GO:0005201~extracellular matrix structural constituent	10	1.97E-07	2.67E-04
		GO:0004252~serine-type endopeptidase activity	13	1.29E-04	0.175317
		GO:0005178~integrin binding	8	4.36E-04	0.589961
		GO:0017147~Wnt-protein binding	5	6.59E-04	0.889743
Down-regulated	GOTERM_BP_DIRECT	GO:0014002~astrocyte development	4	9.56E-05	0.137416
		GO:0001525~angiogenesis	5	0.003404	4.784321
		GO:0001666~response to hypoxia	4	0.007062	9.686593
		GO:2001244~positive regulation of intrinsic apoptotic signaling pathway	3	0.007917	10.79746
		GO:0070488~neutrophil aggregation	2	0.012224	16.20669
		GO:0002793~positive regulation of peptide secretion	2	0.012224	16.20669
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	15	5.87E-04	0.629604
		GO:0005887~integral component of plasma membrane	11	0.010539	10.77307
		GO:0016021~integral component of membrane	32	0.012844	12.98409
		GO:0005901~caveola	3	0.018758	18.43186
		GO:0005886~plasma membrane	16	0.026289	24.92018
		GO:0031090~organelle membrane	2	0.04719	40.55265
	GOTERM_MF_DIRECT	GO:0005509~calcium ion binding	10	0.006725	7.577132
		GO:0047498~calcium-dependent phospholipase A2 activity	2	0.017469	18.59931
		GO:0050544~arachidonic acid binding	2	0.017469	18.59931

GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function.

Table 3. KEGG pathway analysis of differentially expressed genes in IPF

Expression	Term	Count	*p*-Value	Genes
Up-regulated	ECM-receptor interaction	11	5.59E-07	ITGB8, COMP, TNC, COL3A1, COL6A3, ITGA7, COL1A2, COL1A1, THBS2, COL5A2, SPP1
	Protein digestion and absorption	10	5.70E-06	KCNN4, COL17A1, COL14A1, COL3A1, COL6A3, COL1A2, COL15A1, COL1A1, COL5A2, COL10A1
	Focal adhesion	12	2.37E-04	ITGB8, COMP, TNC, COL3A1, COL6A3, ITGA7, COL1A2, IGF1, COL1A1, THBS2, COL5A2, SPP1
Down-regulated	Dilated cardiomyopathy	4	0.015	ADRB1, TNNC1, TTN, CACNA2D2
	Neuroactive ligand-receptor interaction	6	0.02	EDNRB, S1PR1, ADRB1, RXFP1, GRIA1, VIPR1

KEGG, Kyoto Encyclopedia of Gene and Genome; IPF, idiopathic pulmonary fibrosis; ECM, excess extracellular matrix.

Discussion

In the present analysis, 5520 DEGs were identified in IPF based on six profile datasets, including 3714 up-regulated genes and 1806 down-regulated genes. Furthermore, a total of 367 commonly changed DEGs were revealed via Venn software, including 259 up-regulated genes and 108 down-regulated genes. Notably, 8 DEGs were upregulated in all datasets (MMP7, TRIM2, ASPN, SULF1, CXCL14, DCLK1, IL13RA2, TP63), and some of these have been

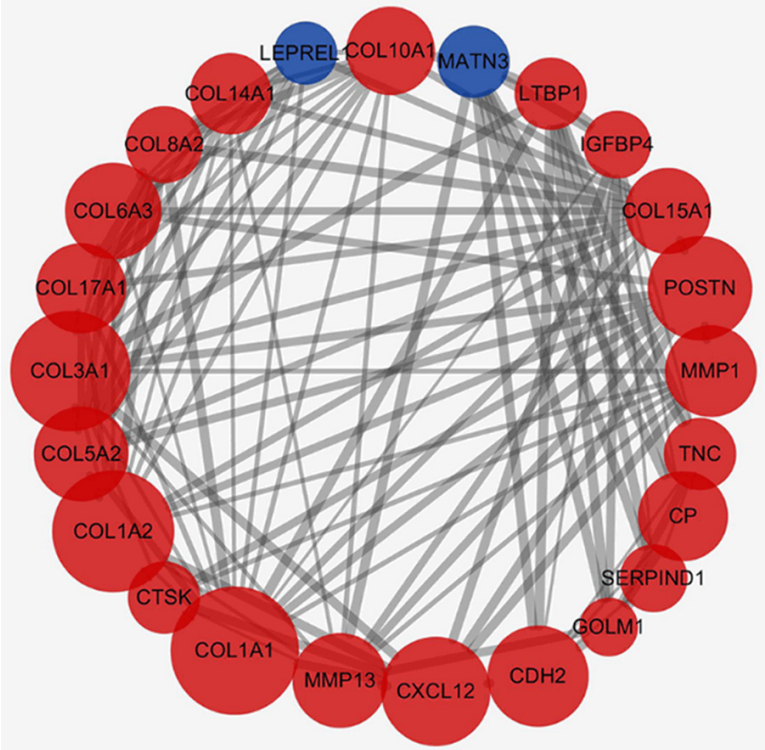

Figure 1. Module analysis identified 24 central nodes as hub genes by MCODE app in Cytotype software. The nodes indicate proteins; the edges indicate the interaction of proteins; red circles represent up-regulated DEGs and blue circles represented down-regulated DEGs. (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. Depth = 100).

validated to be related to pathogenesis of pulmonary fibrosis. Better understanding of the molecular mechanisms of overlapping genes and hub genes in IPF may provide novel insights into the diagnosis and treatment of IPF. Among these overlapping genes, matrix metalloproteinase (MMP)-7, an extensively studied biomarker of IPF, has been reported to be expressed in serum and bronchoalveolar lavage fluid from patients with IPF [22]. MMP7 is also one of the most promising prognostic biomarkers in IPF, with serum levels reproducibly correlating with lung function, clinical predictors of all-cause mortality, and transplant free survival in several studies [23, 24]. Heparan sulfate (HS) 6-O-endosulfatase 1 (SULF1) is a transforming growth factor-β1 (TGF-β1)-responsive gene in normal human lung fibroblasts and functions as a negative feedback regulator of TGF-β1. Overexpression of SULF1 promoted silica-induced proliferative and fibrogenic gene expression, and collagen production [25], while SULF1 siRNA transfection enhanced the anti-proliferative effect of TGF-β1 [26]. CXCL14 has been shown to overexpress in whole lung tissue of IPF patients [17, 27]. Circulating CXCL14 protein levels were significantly higher in plasma from IPF patients than controls [28]. Knockdown of CXCL14 prevented LPS-induced fibrogenesis of L929 cells through inhibiting cell proliferation and decreasing the expression of MMP2/9, Hyp, collagen I/III [29]. Likewise, IL-13RA2 gene silencing or blockade of IL-13RA2 signaling led to marked downregulation of TGF-β1 production and collagen deposition in bleomycin-induced lung fibrosis [30]. Our findings, consistent with previous studies, imply that the above DEGs might be potential diagnostic or prognostic biomarkers, but the therapeutic value needs future validation based on more *in vivo* research and clinical trials. Also, a set of DEGs (TRIM2, ASPN, DCLK1, TP63), without previous studies in IPF were also identified in our analysis, although the exact contributions of them are not clear yet. A recent quantitative proteomic study suggested that ASPN may be a novel ECM protein with unknown function deposited in IPF lung tissue [31]. Further research is necessary to explore how those genes participate in IPF progression as they could be transcriptomic markers.

In the GO analysis, the up-regulated DEGs were particularly enriched in a series of biological processes (BP) such as collagen catabolic process, extracellular matrix organization and cell adhesion, while changes in BP of down-regulated DEGs were significantly enriched in astrocyte development, angiogenesis, and response to hypoxia. For CC, both up- and down-regulated DEGs were predominantly enriched in extracellular region or space, while for MF, calcium ion binding was in the top 3 GO terms. These results are concordant with previous studies which has demonstrated that

Table 4. Re-analysis of 24 hub genes by KEGG pathway enrichment

Term	Count	*p*-Value	Genes
Protein digestion and absorption	9	9.97E-12	COL17A1, COL14A1, COL3A1, COL6A3, COL1A2, COL15A1, COL1A1, COL5A2, COL10A1
ECM-receptor interaction	6	5.69E-07	TNC, COL3A1, COL6A3, COL1A2, COL1A1, COL5A2
Focal adhesion	6	3.86E-05	TNC, COL3A1, COL6A3, COL1A2, COL1A1, COL5A2
PI3K-Akt signaling pathway	6	3.24E-04	TNC, COL3A1, COL6A3, COL1A2, COL1A1, COL5A2
Amoebiasis	4	0.001272	COL3A1, COL1A2, COL1A1, COL5A2
Platelet activation	4	0.001975	COL3A1, COL1A2, COL1A1, COL5A2

KEGG, Kyoto Encyclopedia of Gene and Genome; ECM, excess extracellular matrix.

pulmonary fibrosis is characterized by aberrant fibro-proliferation, excess extracellular matrix (ECM) accumulation, and increased collagen deposition in tissues [32]. Accordingly, for pathway analysis, up-regulated DEGs were mainly enriched in ECM-receptor interaction, protein digestion and absorption, and focal adhesion, while ECM-receptor interaction and focal adhesion have been proven to be closely related to the regulation of IPF [33].

DEGs PPI network complex of 304 nodes and 837 edges was constructed via the STRING online database and Cytoscape software. Using Cytotype MCODE analysis, 24 DEGs with degrees ≥10 were screened as hub genes from the PPI network complex. Finally, we re-analyzed the hub genes using DAVID for KEGG pathway enrichment and found that 10 genes (COL17A1, COL14A1, COL3A1, COL6A3, COL1A2, COL15A1, COL1A1, COL5A2, COL10A1, TNC) enriched in protein digestion and absorption, ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, amoebiasis and platelet activation had a significance ($P<0.05$), which was highly consistent with the results from GO and KEGG enrichment analyses for commonly DEGs identified in the six datasets, indicating that these pathways play crucial roles in the progression of IPF and may also be potential targets for the treatment of IPF.

In IPF, activated myofibroblasts secrete exaggerated levels of ECM within the alveolar space at the site of epithelial cell loss and the IPF ECM itself is also fibrogenic [34]. Multiple ECM proteins are involved in fibrosis, including collagens, tenasin-c (TNC), fibronectin and periostin [35]. Research over the past two decades has elucidated several pathological processes that are integral to disease initiation such as transforming growth factor (TGF)-β and Smad signaling [36], but the molecular mechanisms that drive disease progression remain less well defined. In the present study, commonly DEGs identified and hub genes screened were mainly enriched in the downstream biologic processes or pathways like collagen catabolism, ECM organization, ECM-receptor interaction, and protein digestion, which seem to be related to disease progression other than initiation. There are several reasons that would explain these findings. First, samples in the six datasets were all from IPF patients who have been diagnosed with definite histologic usual interstitial pneumonia (UIP) pattern, or those with end-stage disease undergoing transplantation. As the molecular mechanisms that mediate disease initiation and disease progression may be distinct, transcriptomic data from IPF patients with earlier-stage disease may provide different information. Another explanation is based on a novel view that remodeling of ECM is not only a consequence of IPF, but also a cause. It has been reported that the IPF fibroblastic focus is a polarized structure where active ECM synthesis often takes place has been recently established [37, 38]. As the number of collagen cross-links double, ECM stiffness can increase, switching TGF-β signaling from apoptosis to epithelial-mesenchymal transition in a PI3K/Akt-dependent manner [39, 40]. In our study, hub genes up-regulated were also enriched in PI3K-Akt signaling pathway, along with protein digestion, ECM-receptor interaction and focal adhesion, suggesting that the formation of this positive feedback loop might be pivotal in IPF progression.

Recent studies have explored DEGs participating in IPF by bioinformatic approach. For example, Wang et al. identified 350 DEGs genetically associated with IPF from three microarray datasets including 54 IPF samples and 34 normal samples [41]. Fan et al. predicted potential microRNA-target interac-

tions between 17 DE-miRNAs and 17 DEGs by using three web-available microarray datasets [42]. Wang et al. analyzed gene expression data from 119 patients with IPF and 50 controls from GSE32537 datasets using weighted gene coexpression network analysis (WGCNA), and demonstrated four hub genes (COL14A1, TSHZ2, IL1R2 and SLCO4A1) as potential biomarkers for IPF [43]. However, the results of previous integrative analyses were variant, let alone single transcriptomic studies. High false-positive rates may be observed in small-sample microarray analyses, and different sample types could contribute to discordant results. Compared with the previous studies, our research was based on six GEO datasets comprising 329 lung tissue samples, and 367 commonly DEGs were identified, which is a more comprehensive bioinformatic analysis to identify possible biomarkers and elucidate molecular mechanisms of IPF.

Nevertheless, the findings of our study should be considered with caution due to several limitations. First, a functional deficit or acquisition study are warranted to clarify the functions of DEGs and hub genes in IPF, especially for those genes without previous studies in IPF. Second, hub genes and their interactions were constructed by bioinformatic approach. To figure out the underlying link among these molecules, further investigation are necessary. Third, the current study identified DEGs in lung tissues, but for diagnostic purpose, lung tissue is not readily accessible. Integrative analyses for identification of potential transcriptomic markers in IPF should also be performed based on other same types such as peripheral blood and sputum. Fourth, lack of subgroup analyses based on potential influential factors, including age, sex, disease severity, and platform usage limits our further investigation on complex relationships between gene expression profiles and phenotypes. Also, biological knowledge base and pathway information are far from being complete at present. Thus, stratified analyses on different factors such as age, sex, disease severity, and platform are needed to reach a more convincing conclusion, and a more complete biologic knowledge base and pathway information are required.

Conclusion

The integrative analysis of gene expression profiles included six microarray datasets and identified 367 commonly changed DEGs and 24 hub genes that may be diagnostic biomarkers or therapeutic targets of IPF. Although these predictions should be verified by a series of experiments in the future, our results provided novel insights into the pathogenesis of IPF. Such insights into the molecular biology of IPF may ultimately lead to effective anti-fibrosis therapeutics.

Acknowledgements

This work was supported by National Natural Science Foundation of China (81800042 and 81701586). The funders had no role in study design, data collection or analysis, decision to publish, and manuscript preparation.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Bo Wang, Department of Respiratory and Critical Care Medicine, West China Hospital of Sichuan University, Chengdu 610041, China. E-mail: wangbo31hx@163.com

References

[1] Park JH, Kim DS, Park IN, Jang SJ, Kitaichi M, Nicholson AG and Colby TV. Prognosis of fibrotic interstitial pneumonia: idiopathic versus collagen vascular disease-related subtypes. Am J Respir Crit Care Med 2007; 175: 705-711.

[2] American Thoracic Society; European Respiratory Society. 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: 277-304.

[3] Kishaba T. Evaluation and management of idiopathic pulmonary fibrosis. Respir Investig 2019; 57: 300-311.

[4] Sivakumar P, Thompson JR, Ammar R, Porteous M, McCoubrey C, Cantu E 3rd, Ravi K, Zhang Y, Luo Y, Streltsov D, Beers MF, Jarai G and Christie JD. RNA sequencing of transplant-stage idiopathic pulmonary fibrosis lung reveals unique pathway regulation. ERJ Open Res 2019; 5: 00117-2019.

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

[6] Barros A, Oldham J and Noth I. Genetics of idiopathic pulmonary fibrosis. Am J Med Sci 2019; 357: 379-383.
[7] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr and Kinzler KW. Cancer genome landscapes. Science 2013; 339: 1546-1558.
[8] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S and Soboleva A. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013; 41: D991-995.
[9] Feng H, Gu ZY, Li Q, Liu QH, Yang XY and Zhang JJ. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. J Ovarian Res 2019; 12: 35.
[10] Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.
[11] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet 2000; 25: 25-29.
[12] Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28: 27-30.
[13] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-452.
[14] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.
[15] Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 2011; 27: 431-432.
[16] Cecchini MJ, Hosein K, Howlett CJ, Joseph M and Mura M. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. Respir Res 2018; 19: 153.
[17] DePianto DJ, Chandriani S, Abbas AR, Jia G, N'Diaye EN, Caplazi P, Kauder SE, Biswas S, Karnik SK, Ha C, Modrusan Z, Matthay MA, Kukreja J, Collard HR, Egen JG, Wolters PJ and Arron JR. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. Thorax 2015; 70: 48-56.
[18] Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW, Kim DS and Kaminski N. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2009; 180: 167-175.
[19] Meltzer EB, Barry WT, D'Amico TA, Davis RD, Lin SS, Onaitis MW, Morrison LD, Sporn TA, Steele MP and Noble PW. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. BMC Med Genomics 2011; 4: 70.
[20] Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, Rosen R, Neidermyer AJ, McKean DF, Groshong SD, Cool C, Cosgrove GP, Lynch DA, Brown KK, Schwarz MI, Fingerlin TE and Schwartz DA. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. Thorax 2013; 68: 1114-1121.
[21] Wang XM, Zhang Y, Kim HP, Zhou Z, Feghali-Bostwick CA, Liu F, Ifedigbo E, Xu X, Oury TD, Kaminski N and Choi AM. Caveolin-1: a critical regulator of lung fibrosis in idiopathic pulmonary fibrosis. J Exp Med 2006; 203: 2895-2906.
[22] Borensztajn K, Crestani B and Kolb M. Idiopathic pulmonary fibrosis: from epithelial injury to biomarkers--insights from the bench side. Respiration 2013; 86: 441-452.
[23] Tzouvelekis A, Kouliatsis G, Anevlavis S and Bouros D. Serum biomarkers in interstitial lung diseases. Respir Res 2005; 6: 78.
[24] Song JW, Do KH, Jang SJ, Colby TV, Han S and Kim DS. Blood biomarkers MMP-7 and SP-A: predictors of outcome in idiopathic pulmonary fibrosis. Chest 2013; 143: 1422-1429.
[25] Perkins TN, Peeters PM, Albrecht C, Schins RPF, Dentener MA, Mossman BT, Wouters EFM and Reynaert NL. Crystalline silica alters Sulfatase-1 expression in rat lungs which influences hyper-proliferative and fibrogenic effects in human lung epithelial cells. Toxicol Appl Pharmacol 2018; 348: 43-53.
[26] Yue X, Li X, Nguyen HT, Chin DR, Sullivan DE and Lasky JA. Transforming growth factor-beta1 induces heparan sulfate 6-O-endosulfatase 1 expression in vitro and in vivo. J Biol Chem 2008; 283: 20397-20407.
[27] Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, Aziz N, Kaminski N and Zlotnik A. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity

pneumonitis. Am J Respir Crit Care Med 2006; 173: 188-198.
[28] Jia G, Chandriani S, Abbas AR, DePianto DJ, N'Diaye EN, Yaylaoglu MB, Moore HM, Peng I, DeVoss J, Collard HR, Wolters PJ, Egen JG and Arron JR. CXCL14 is a candidate biomarker for hedgehog signalling in idiopathic pulmonary fibrosis. Thorax 2017; 72: 780-787.
[29] Li L, Li Q, Wei L, Wang Z, Ma W, Liu F, Shen Y, Zhang S, Zhang X, Li H and Qian Y. Chemokine (C-X-C motif) ligand 14 contributes to lipopolysaccharide-induced fibrogenesis in mouse L929 fibroblasts via modulating PPM1A. J Cell Biochem 2019; 120: 13372-13381.
[30] Fichtner-Feigl S, Strober W, Kawakami K, Puri RK and Kitani A. IL-13 signaling through the IL-13alpha2 receptor is involved in induction of TGF-beta1 production and fibrosis. Nat Med 2006; 12: 99-106.
[31] Tian Y, Li H, Gao Y, Liu C, Qiu T, Wu H, Cao M, Zhang Y, Ding H, Chen J and Cai H. Quantitative proteomic characterization of lung tissue in idiopathic pulmonary fibrosis. Clin Proteomics 2019; 16: 6.
[32] King TE Jr, Pardo A and Selman M. Idiopathic pulmonary fibrosis. Lancet 2011; 378: 1949-1961.
[33] Gvaramia D, Blaauboer ME, Hanemaaijer R and Everts V. Role of caveolin-1 in fibrotic diseases. Matrix Biol 2013; 32: 307-315.
[34] Parker MW, Rossi D, Peterson M, Smith K, Sikström K, White ES, Connett JE, Henke CA, Larsson O and Bitterman PB. Fibrotic extracellular matrix activates a profibrotic positive feedback loop. J Clin Invest 2014; 124: 1622-1635.
[35] Liu G, Cooley MA, Jarnicki AG, Borghuis T, Nair PM, Tjin G, Hsu AC, Haw TJ, Fricker M, Harrison CL, Jones B, Hansbro NG, Wark PA, Horvat JC, Argraves WS, Oliver BG, Knight DA, Burgess JK and Hansbro PM. Fibulin-1c regulates transforming growth factor-beta activation in pulmonary tissue fibrosis. JCI Insight 2019; 5: e124529.
[36] King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, Gorina E, Hopkins PM, Kardatzke D, Lancaster L, Lederer DJ, Nathan SD, Pereira CA, Sahn SA, Sussman R, Swigris JJ and Noble PW; ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med 2014; 370: 2083-2092.
[37] Xia H, Gilbertsen A, Herrera J, Racila E, Smith K, Peterson M, Griffin T, Benyumov A, Yang L, Bitterman PB and Henke CA. Calcium-binding protein S100A4 confers mesenchymal progenitor cell fibrogenicity in idiopathic pulmonary fibrosis. J Clin Invest 2017; 127: 2586-2597.
[38] Yang L, Herrera J, Gilbertsen A, Xia H, Smith K, Benyumov A, Bitterman PB and Henke CA. IL-8 mediates idiopathic pulmonary fibrosis mesenchymal progenitor cell fibrogenicity. Am J Physiol Lung Cell Mol Physiol 2018; 314: L127-L136.
[39] Leight JL, Wozniak MA, Chen S, Lynch ML and Chen CS. Matrix rigidity regulates a switch between TGF-beta1-induced apoptosis and epithelial-mesenchymal transition. Mol Biol Cell 2012; 23: 781-791.
[40] Han H, Wecker T, Grehn F and Schlunck G. Elasticity-dependent modulation of TGF-beta responses in human trabecular meshwork cells. Invest Ophthalmol Vis Sci 2011; 52: 2889-2896.
[41] Wang H, Xie Q, Ou-Yang W and Zhang M. Integrative analyses of genes associated with idiopathic pulmonary fibrosis. J Cell Biochem 2018; [Epub ahead of print].
[42] Fan L, Yu X, Huang Z, Zheng S, Zhou Y, Lv H, Zeng Y, Xu JF, Zhu X and Yi X. Analysis of microarray-identified genes and microRNAs associated with idiopathic pulmonary fibrosis. Mediators Inflamm 2017; 2017: 1804240.
[43] Wang Z, Zhu J, Chen F and Ma L. Weighted gene coexpression network analysis identifies key genes and pathways associated with idiopathic pulmonary fibrosis. Med Sci Monit 2019; 25: 4285-4304.

Table S1. 367 common DEGs in IPF (overlapped in at least three datasets) by Venn diagram software

GEO accession	Number of DEGs	Gene name
Up-regulated	259	
GSE10667 GSE110147 GSE2052 GSE24206 GSE32537 GSE53845	8	MMP7 TRIM2 ASPN SULF1 CXCL14 DCLK1 IL13RA2 TP63
GSE10667 GSE2052 GSE24206 GSE32537 GSE53845	3	COMP SLN COL15A1
GSE10667 GSE110147 GSE2052 GSE32537 GSE53845	7	IGF1 SPP1 POSTN CLDN1 TMEM45A TNC TDO2
GSE10667 GSE110147 GSE24206 GSE32537 GSE53845	21	LRRN1 COL1A1 MMP16 LTBP1 SFRP2 ITGB8 CDH3 SPATA18 DIO2 SLAMF7 PSD3 GOLM1 SYNPO2 PLN LRRC17 CFH NRP2 COL3A1 EPHA3 CD24 COL14A1
GSE10667 GSE2052 GSE24206 GSE53845	3	ALDH1A3 FHL2 SCG5
GSE10667 GSE2052 GSE32537 GSE53845	1	TGFB3
GSE10667 GSE110147 GSE2052 GSE32537	2	SERPINB5 DSC3
GSE10667 GSE110147 GSE2052 GSE53845	5	LGI2 S100A2 CFI CFB GPR87
GSE10667 GSE24206 GSE32537 GSE53845	11	FNDC1 PCDH7 C12orf75 THY1 MXRA5 PLPPR4 ST6GALNAC1 SERPIND1 ZNF521 LAMP5 ECM2
GSE110147 GSE24206 GSE32537 GSE53845	4	MUC5B OGN VCAM1 CCDC146
GSE10667 GSE110147 GSE24206 GSE32537	12	KLHL13 TNFRSF19 RPGRIP1L SYTL2 GPX8 PTGFRN HSPA4L CCDC170 CRISPLD1 CNTN3 SLFN13 FANK1
GSE10667 GSE110147 GSE24206 GSE53845	7	KCNMA1 PGM2L1 TSHZ2 FRMD6 MOXD1 BICC1 FERMT1
GSE10667 GSE110147 GSE32537 GSE53845	23	CXCL13 VTCN1 KRT15 VSIG1 TMPRSS4 MMP1 BPIFB1 SERPINB3 ABCA13 MMP13 CDH2 AMPD1 SFRP4 MUC4 FAP CCDC80 PROM1 COL17A1 COL6A3 KRT5 CP CYP24A1 CLIC6
GSE10667 GSE2052 GSE24206	4	DOK5 MEOX1 ENC1 ZKSCAN7
GSE10667 GSE2052 GSE53845	10	PCSK1 LGALS7 CHRDL2 RARRES1 ITGA7 TWIST1 TMEM158 PLEKHA4 KCNN4 SFRP1
GSE10667 GSE110147 GSE2052	2	PLA2G2A FAM198B
GSE24206 GSE32537 GSE53845	2	MS4A2 TPPP3
GSE10667 GSE24206 GSE32537	5	GLT8D2 GXYLT2 PDE7B CCL18 FGF14
GSE110147 GSE24206 GSE32537	10	MNS1 EFHC1 RASSF9 CAPS2 HMCN1 SLITRK6 PRSS12 DZIP3 CEP126 SPEF2
GSE10667 GSE24206 GSE53845	28	CPXM2 BACE2 SULF2 IL17RD PDLIM4 CTSE FBLN2 MFAP2 NHS KIAA1211 COL8A2 COL1A2 TMEM176B KIF26B TSPAN11 PROM2 LDLRAD4 HS6ST2 ROBO1 CDCA7 IGDCC4 CXCL12 MDK COL10A1 BCL11A IGFBP4 CTHRC1 SLC4A11
GSE110147 GSE24206 GSE53845	3	BCHE ITGBL1 RGS5
GSE10667 GSE110147 GSE24206	10	ANLN ABCC5 FZD3 SLC28A3 CROT CSRNP3 SERPINE2 TOP2A TRIM59 AHNAK2
GSE10667 GSE32537 GSE53845	15	TRIM29 CTSK PLPP2 IGKC SOX2 CCL13 GSTA1 UBXN10 CAPN13 PAMR1 SIX4 TSPAN1 FAM83D PIP SIX1
GSE110147 GSE32537 GSE53845	3	HAS2 NELL2 GEM
GSE10667 GSE110147 GSE32537	34	STOX1 CCDC190 CFAP206 HHLA2 CCDC113 FAM81B C9orf135 CFAP47 DNAH7 CFAP53 RGS22 DNAH10 DNAH3 SLC27A2 C11orf70 ZNF385D SPATA17 DNAJA4 MYH11 IQCG CLCA2 PRUNE2 EFHC2 NEK11 EFCAB10 BPIFA1 CHST9 EFHB PLEKHS1 EFCAB1 DNAH5 MDH1B C6 WDR63
GSE10667 GSE110147 GSE53845	26	CLMP LMAN1 ADGRF1 KRT17 LXN COL5A2 TNFRSF17 SMOC2 PLA2G7 FCRL5 MMP10 BAAT CYP1B1 PTPRZ1 MRVI1 ARNTL2 GREM1 DFNA5 PDE1A THBS2 STEAP2 DST AIM2 CFHR3 LCN2 BMS1P20
Down-regulated	108	
GSE10667 GSE110147 GSE2052 GSE24206 GSE32537 GSE53845	1	CRTAC1
GSE10667 GSE2052 GSE24206 GSE32537 GSE53845	1	SLC39A8
GSE110147 GSE2052 GSE24206 GSE32537 GSE53845	2	S1PR1 PLLP
GSE10667 GSE110147 GSE2052 GSE24206 GSE53845	1	CA4

GSE10667 GSE110147 GSE2052 GSE32537 GSE53845	2	TMEM100 P3H2
GSE10667 GSE110147 GSE24206 GSE32537 GSE53845	5	BTNL9 PLA2G1B NECAB1 HHIP HSD17B6
GSE2052 GSE24206 GSE32537 GSE53845	1	PTPRB
GSE110147 GSE2052 GSE24206 GSE32537	1	LRRC32
GSE10667 GSE2052 GSE24206 GSE53845	2	EDNRB CCK
GSE110147 GSE2052 GSE24206 GSE53845	1	CLDN5
GSE10667 GSE2052 GSE32537 GSE53845	1	HECW2
GSE110147 GSE2052 GSE32537 GSE53845	4	ABCA3 ACVRL1 SLCO2A1 SMAD6
GSE10667 GSE110147 GSE2052 GSE32537	1	AGER
GSE10667 GSE110147 GSE2052 GSE53845	8	CDH13 EMP2 TNNC1 DAPK2 CLIC5 GPM6A STXBP6 FGFBP2
GSE110147 GSE24206 GSE32537 GSE53845	9	PEBP4 PAPSS2 HIF3A NPR1 MT1M S100A8 SLCO4A1 GPR4 S100A12
GSE10667 GSE110147 GSE24206 GSE53845	2	GPIHBP1 FAM107A
GSE10667 GSE110147 GSE32537 GSE53845	4	RTKN2 SLC6A4 ITLN2 VIPR1
GSE2052 GSE24206 GSE32537	2	CSRNP1 CHI3L2
GSE2052 GSE24206 GSE53845	2	SDPR DENND3
GSE110147 GSE2052 GSE24206	1	MATN3
GSE10667 GSE2052 GSE53845	2	OLFML2A PRX
GSE110147 GSE2052 GSE53845	8	TMEM204 KCNK3 ANXA3 LAMC3 SH2D3C EMCN CAV1 HEY1
GSE10667 GSE110147 GSE2052	2	EDN1 SLCO1A2
GSE24206 GSE32537 GSE53845	16	FASN CLEC4E ZNF385B RNASE2 ACADL IL1RL1 SDR16C5 IL18RAP CD163 IL1R2 FMO5 TTN HMGCS1 SERPINA3 SULT1B1 SLCO4C1
GSE110147 GSE24206 GSE32537	3	CSF3R S100A9 AFF3
GSE10667 GSE24206 GSE53845	1	EPB41L5
GSE110147 GSE24206 GSE53845	4	TIMP3 MGST1 STX11 GKN2
GSE10667 GSE110147 GSE24206	2	ADRB1 STC1
GSE10667 GSE32537 GSE53845	1	GALNT18
GSE110147 GSE32537 GSE53845	9	MT1JP ID1 CPB2 SLC19A3 DUOX1 CACNA2D2 PLA2G4F PGC FCN3
GSE10667 GSE110147 GSE32537	3	GRIA1 MS4A15 RXFP1
GSE10667 GSE110147 GSE53845	6	FAM167A EFR3B NCKAP5 MYRF SERTM1 SOSTDC1

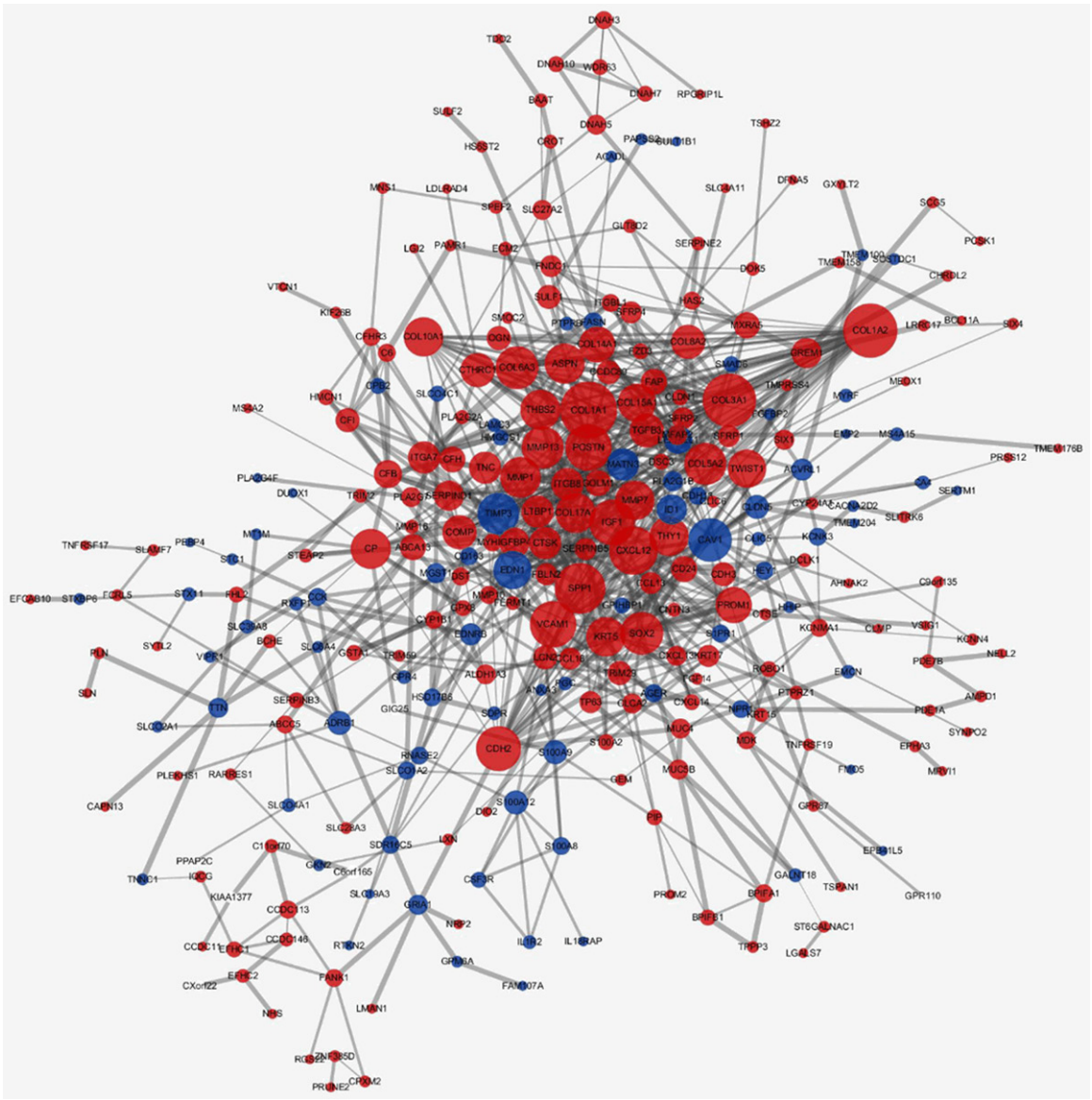

Figure S1. PPI network of commonly DEGs constructed by STRING online database. A total of 107 DEGs in the DEGs PPI network complex. The nodes indicate proteins; the edges indicate the interaction of proteins; red circles represent up-regulated DEGs and blue circles represent down-regulated DEGs.