

Identification of Differentially Expressed Genes Triggered by Aberrant Methylation in Idiopathic Pulmonary Fibrosis Using Integrated Bioinformatic Analysis

Shuaijun Chen

Huazhong University of Science and Technology - Main Campus: Huazhong University of Science and Technology <https://orcid.org/0000-0002-8302-2728>

Jun Zhang

Huazhong University of Science and Technology - Main Campus: Huazhong University of Science and Technology

Wanli Ma

Huazhong University of Science and Technology - Main Campus: Huazhong University of Science and Technology

Hong Ye (✉ yehmwl@hust.edu.cn)

Huazhong University of Science and Technology - Main Campus: Huazhong University of Science and Technology

Research

Keywords: Idiopathic pulmonary fibrosis, IPF, methylation, biomarker, GEO

Posted Date: December 31st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-136553/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Idiopathic pulmonary fibrosis (IPF) is a relentlessly progressive and fatal fibrotic lung disease all over the world, and specific pathogenesis is still not well understood. DNA methylation is an essential epigenetic mechanism, which likely contributes to the progress of IPF. The purpose of this study is to identify aberrantly methylated differentially expressed genes (DEGs) in IPF and to explore the underlying mechanisms of IPF by using integrated bioinformatics analysis.

Method

Gene expression profiles and gene methylation profile were downloaded and analyzed to identify the aberrantly methylated-differentially expressed genes. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Search Tool for the Retrieval of Interacting Genes Database (STRING) and Gene set enrichment analysis (GSEA) were used to evaluate function of DEGs. RT-PCR was used to verify the mRNA levels of DEGs in mice with pulmonary fibrosis.

Results

By analyzing the differentially expressed genes of the three IPF expression profiles, and taking the intersection, we got 143 co-upregulated genes and 104 co-downregulated genes; GO and KEGG pathway analysis of the DEGs suggested these genes involved in the extracellular matrix organization, multicellular organismal homeostasis. Combining the sequencing data of two IPF methylation chips, we have identified genes that may be regulated by methylation in IPF. Finally, we obtained the mRNA expression of DEGs using a mouse model of pulmonary fibrosis.

Conclusion

Through integrated analysis and experimental verification, we found a series of biomarkers which were regulated by methylation should be potential therapeutic targets for IPF.

Introduction

IPF is a chronic progressive interstitial lung disease, which is characterized by the formation of scar tissue and the destruction of lung tissue structure, which ultimately leads to gas exchange dysfunction and restricted ventilation [13]. To date, the pathogenesis of IPF is still unclear, but it may be affected by the complex interaction between environmental and genetic factors. The prognosis of idiopathic pulmonary fibrosis is poor, with a median survival time of 3–5 years after diagnosis and no curative drug therapy. Therefore, exploring specific biomarkers and therapeutic targets in IPF plays a vital role in diagnosing and treating diseases [14].

With the development of sequencing technology, more and more research reveals the occurrence and progression of diseases through multi-omics. A large amount of sequencing data are mostly stored in the GEO database, which can be easily analyzed by other researchers. Ji-Hoon Cho. et al. used array sequencing to analyze 23 IPF patients and 5 normal lung tissues, revealing that several disease-associated modules involving genes from the TGF-beta, Wnt, focal adhesion, and smooth muscle actin pathways may be involved in advancing fibrosis[15]. Daryle J DePianto. et al. analyzed 40 IPF and 8 control lung tissues and performed chip sequencing analysis. The results suggest microscopic pathological heterogeneity in IPF lung tissue corresponds to specific gene expression patterns related to bronchodilation and lymphoid aggregates[1]. Marissa J Schafer et al. used high-throughput sequencing to examine lung tissue from 19 normal and 20 IPF patients and found that senescence markers significantly increased in pulmonary fibrosis development[2].

The epigenetic modification includes DNA methylation, hydroxymethylation, histone modification, non-coding RNA (ncRNA), etc. Among them, DNA methylation plays an essential role in the repression of transposons and genes. In some cases, it is also involved in the activation of genes[16]. DNA methylation is closely related to human growth and development, social behavior, metabolism, tumors, etc. [17–20]. Changes in the level of DNA methylation are common in the occurrence and development of diseases. It can identify specific biomarkers and provide crucial clues to the diagnosis, prognosis, and treatment of diseases. More and more evidence shows that abnormal DNA methylation levels are closely related to the occurrence and development of IPF and provide guidance for the diagnosis and treatment of IPF[16, 21, 22]. For example, Atsushi Hata et al. used methylation microarray sequencing to explore the methylation profiles of normal and fibrotic lung tissue. They found that the low-methylation subgroup significantly correlated with IPF[4]. However, so far there has not been any studies to analyze methylation patterns in pulmonary fibrosis on the gene expression profile.

In our study, We analyzed the sequencing data sets of the IPF mRNA expression spectrum in GEO data and the 450K methylation sequencing data sets. Through the integrated analysis of them, we found that the changes in genomic methylation modification significantly promoted the progress of pulmonary fibrosis.

Materials And Methods

Microarray data

In the study, we downloaded three gene expression profiles (GSE53845[1], GSE92592[2], and GSE124685[3]) and two gene methylation profiles (GSE107226 and GSE121849[4]) from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>).

Data processing and identification of DEGs and DMGS in IPF

We downloaded the expression matrix of IPF patients from GEO. For the chip sequencing data, we performed gene symbol conversion analysis on the expression matrix according to the soft file. For the expression matrix of high-throughput sequencing, we unified the gene name into the form of a gene symbol. For the chip sequencing data, we used the "Limma" package for differential gene screening[5], and for the expression matrix of high-throughput sequencing, we used the "edgR" package for differential analysis[6]. For DEGs, $|\log_2FC| > 1$ and $P < 0.05$ were set as the cutoff criteria. Similarly, we also downloaded the methylation chip sequencing data of IPF patients and used the "ChAMP" package for annotation and difference analysis of methylation sites[7], the differential β value $|\Delta\beta|$ where gene with a positive $\Delta\beta$ value signifies hypermethylation and gene with a negative $\Delta\beta$ value signifies hypomethylation can be applied in subsequent analysis, $|\Delta\beta| > 0.2$ and $P < 0.05$ were set as the cutoff criteria. All calculations were done using R4.0.

GO term and KEGG pathway enrichment analysis

Gene Ontology (GO) enrichment analysis was employed for functional analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was applied to analyze the biological functions and pathways of the DEGs and DMGS[8]. KOBAS is a widely used gene set enrichment (GSE) analysis tool[9], and we assessed GO and KEGG analysis on KOBAS 3.0(<http://kobas.cbi.pku.edu.cn/kobas3/?t=1>). $P < 0.05$ and counts ≥ 2 were used as the significance threshold.

PPI network construction and hub gene identification

We utilized the Search Tool for the Retrieval of Interacting Genes (STRING) database to establish a protein-protein interaction (PPI) network and reveal the relationship among the DEGs[10]. The interaction score was set at 0.4 in the STRING database. Cytoscape was conducted to enhance the legibility of the PPI network on the basis of interaction data obtained from the STRING database. We analyzed the hub genes of the PPI network through the MCODE tool in Cytoscape, a plug-in for exploring the core modules in the gene network[11].

Gene set enrichment analysis (GSEA)

Gene set enrichment analysis (GSEA) is a bioinformatics method that inspects the statistical significance of a priori defined sets of genes and validates the differences between two biological states[12]. We divided samples into two phenotype subgroups, normal and IPF. Genes from the expression profile were ranked in a list according to the degree of divergence between the two subgroups through GSEA software 4.0. Then, Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets were analyzed to identify functional terms and pathways enriched in each phenotype subgroup. Gene set permutations were executed 1000 times for each analysis. The criteria of significantly enriched pathways were normalized p-value < 0.05 and the absolute value of normalized enrichment score (NES) > 1.5 .

Bleomycin-Induced Murine Pulmonary Fibrosis Model

All experimental mice were provided by the Experimental Animal Center of Huazhong University of Science and Technology, and this animal experiment method was approved by the Animal Care and Use Committee (IACUC) of Tongji Medical College of Huazhong University of Science and Technology. The C57BL/6 J mice (18– 20 g, 6-8 weeks of age) were used in this study, and placed in a standard freely accessible water and rodent laboratory food environment. All mice were divided into two groups (bleomycin group and saline group). Bleomycin dissolved in saline solution (concentration, 4 mg/ml) was administered by intraperitoneal injection at a dose of 40 mg/kg on day 1, 5, 8, 11, and 15. In the saline group, the saline treatment was given at the same time point. Forty days later, the mice were euthanized, and lung tissues were taken for RNA extraction and tissue staining.

Masson's trichrome staining

The mouse lung tissues with the largest cross-section were embedded in paraffin and cut into 5- μ M-thick tissue sections. And then, the tissue sections were stained with Masson's Trichrome Stain kit (Servicebio, G1006-100ML). Finally, Olympus DP50-CU microscope was employed to observe the blue collagen deposition in the tissue sections and take pictures for morphological analysis and evaluation of collagen deposition.

RNA extraction and quantitative real-time RT-PCR.

After the mouse was dissected, the lung tissue was taken out and frozen in liquid nitrogen immediately and then placed in a refrigerator at -80°C until it was used. The tissues were lysed with TRIzol reagent (Vazyme) and then extracted according to the instructions provided by the manufacturer. Next, cDNA was synthesized using HiScript II Q RT SuperMix (Vazyme, R223-01), and quantitative real-time PCR was performed using ChamQ SYBR qPCR Master Mix (Vazyme, Q331-02/03). The primers for and GAPDH are as follows:

Statistical analysis

Results are presented as mean \pm SEM. All the data was analyzed by one-way analysis of variance (ANOVA) and Tukey's test for multiple comparisons. All p-value of less than 0.05 was considered statistically significant. Data analysis was processed by GraphPad Prism 8 software (La Jolla, CA, USA).

Results

Differentially expressed genes between IPF patients and normal people

We performed a differential expression of the gene sequencing data of three different IPF patient cohorts. According to our screening criteria ($\log_2|FC| > 1$ and $p \text{ value} < 0.05$), for the GSE53945 data set, we got 964 up-regulated genes and 1123 down-regulated genes. In the data set GSE92592, we have 1698 up-regulated genes and 1309 down-regulated genes. In the data set GSE124685, we have 617 up-regulated genes and 395 down-regulated genes. The volcano map shows the up-regulated and down-regulated genes (Figure 1A). Each heatmap shows the overall expression level of the differentially expressed genes between IPF patients and the normal population (Figure 1B, Supplementary Table S1). To obtain more common and representative differentially expressed genes in IPF patients, we took the intersection of these three different data sets and finally got 143 co-upregulated genes and 104 co-downregulated genes (Figure 1C).

GO function enrichment analysis of differentially expressed genes

To further explore the functions of these differentially expressed genes, we performed GO function enrichment analysis on the up-regulated genes and the down-regulated genes, respectively. As the results showed, the up-regulated genes were enriched in response to vitamin, response to nutrient, regeneration, osteoblast differentiation, ossification, hemidesmosome assembly, extracellular matrix structural constituent, extracellular structure organization, extracellular matrix disassembly, collagen trimer, collagen metabolic process, collagen fibril organization, collagen-containing extracellular matrix, chondrocyte development, cartilage development, bone development, basement membrane (Figure 2A). We could find that these up-regulated differential genes were enriched in multiple collagen-related GO terms. The genes involved include SFRP2, COL14A1, COL15A1, MMP11, SPP1, LOXL1, MMP1, TNC, ITGA11, COL17A1, COL10A1, MFAP2, VCAM1, FBLN2, SULF1, FAP, COL1A1, MMP13, CCDC80, GREM1, TTR, POSTN, MMP7, MMP10, COMP. Also, down-regulated genes were enriched in regulation of calcium-mediated signaling, positive regulation of integrin-mediated signaling pathway, positive regulation of endothelial cell migration, positive regulation of cold-induced thermogenesis, PDZ domain binding, negative regulation of apoptotic process, N-methyltransferase activity, integral component of membrane, G protein-coupled peptide receptor activity, extracellular region, calcium ion binding, activation of adenylate cyclase activity and so on (Figure 2B, Supplementary Table S2). These downregulated GO terms may be involved in the occurrence and progression of IPF.

KEGG function enrichment analysis of differentially expressed genes

To explore the altered signaling pathways in IPF, we further performed KEGG enrichment analysis on the up-regulated genes and the downregulated genes, respectively. The results suggested that the up-regulated genes were enriched in Protein digestion and absorption, PI3K-Akt signaling pathway, Phagosome, Pathogenic Escherichia coli infection, Focal adhesion, ECM-receptor interaction, Cytokine-cytokine receptor interaction, Cell adhesion molecules (CAMs) (Figure 3A). The down-regulated genes were enriched in pathways in cancer, Neuroactive ligand-receptor interaction, metabolic pathways, dilated

cardiomyopathy (DCM), cGMP-PKG signaling pathway, cAMP signaling pathway(Figure 3B, Supplementary Table S3).

Protein interaction network of differentially expressed genes

To further investigate the molecular mechanism of the development of IPF, we attempted to describe the interaction network between these different genes based on the STRING tool. The result indicated that these genes form an intertwined network with each other(Figure 4A). On the other hand, we used the tool MCODE to screen out two clusters of core genes from these differential genes,cluster1 included COL17A1, COL10A1, LCOL14A1, COL15A1, MMP13, SPP1, MMP1, and cluster2 included P2RY6, NTS, ADRB1, VIPR1, RXFP1, EDNRE, BDKRB2, RXFP1, Some of these genes are known to play a key role in the progression of pulmonary fibrosis(Figure 4B). Other genes may be involved in the formation of pulmonary fibrosis.

GSEA analysis of the mRNA expression profile

We carried out GSEA analysis on the expression profile of IPF patients and normal tissues to explore the signaling pathways changed during the formation and development of IPF. The results showed that Asthma, Type I diabetes mellitus,Lupus erythematosus, intestinal immune network for IgA production,p53 signaling pathway were activated in the tissues of IPF while aldosterone regulated sodium reabsorption was activated in the tissues of normal(Figure 5).

The role of gene methylation modification in pulmonary fibrosis

To further explore the molecular mechanism in the process of lung fibrosis, we analyzed the results of two methylation chips from IPF and normal samples. As shown in the heatmap, there are different gene methylation patterns between IPF tissue and normal lung tissue. The differentially methylated genes could well distinguish IPF tissue from normal lung tissue and were expected to become a new methylation marker for IPF(Figure 6A, Supplementary Table S4). Taking the intersection between the differentially expressed gene set and the differentially methylated gene set, we finally got 8 genes with low methylation and high expression at the same time, including CXCL14, DAPL1, DOK5, FNDC4, MMP7, MMP10, MMP11, SPP1 (Figure 6B).

GO function enrichment analysis of methylated genes

We carried out GO enrichment analysis of hypermethylated and hypomethylated genes, respectively. The results indicated that the hypermethylated genes were enriched in transforming growth factor-beta receptor signaling pathway, transcription regulator complex, SMAD protein signal transduction, response to endoplasmic reticulum stress, negative regulation of the apoptotic process, cell migration, and so on

(Figure 7A). The hypomethylated genes were enriched in regulation of small GTPase mediated signal transduction, positive regulation of MAPK cascade, positive regulation of adenylate cyclase activity, negative regulation of cell population proliferation, inflammatory response, extracellular matrix organization, extracellular exosome, DNA-binding transcription factor activity, collagen catabolic process (Figure 7B, Supplementary Table S5).

KEGG function enrichment analysis of methylated genes

KEGG enrichment analysis was performed to investigate the signaling pathways that these differentially methylated genes might be involved in. The results indicated that the hypermethylated genes were enriched in the TGF-beta signaling pathway, Ras signaling pathway, Rap1 signaling pathway, PI3K-Akt signaling pathway, HIF-1 signaling pathway, Focal adhesion, Chemokine signaling pathway, and so on (Figure 8A). The hypomethylated genes were enriched in Toll-like receptor signaling pathway, p53 signaling pathway, Hippo signaling pathway, Cytokine-cytokine receptor interaction, Cell adhesion molecules (CAMs), cAMP signaling pathway, C-type lectin receptor signaling pathway (Figure 8B, Supplementary Table S6).

Validation in a mouse model

To better verify whether these hypomethylated genes were highly expressed in pulmonary fibrosis, mice models of pulmonary fibrosis were constructed by intraperitoneal injection of bleomycin. Masson's trichrome staining results indicated that the bleomycin treatment group showed apparent collagen deposition and pluralization characteristics (Figure 9A). RT-PCR was performed to detect the mRNA levels of these genes in pulmonary fibrosis tissues.

The results indicated that the expression of CXCL14, DAPL1, DOK5, FNDC4, MMP7, MMP10, MMP11, SPP1 were significantly increased in the IPF group compared with the control (Figure 9B)

Discussion

Mortality in IPF is high, with a reported median survival of 2–3 years from diagnosis, based on historical data[23], and more recent evidence shows no improvement in survival[24–26]. It is well recognized that IPF is a heterogeneous disease with a variable disease course, and it's very difficult for predicting disease outcomes [27]. Sequencing technology has provided convenience for us to explore the internal mechanism of pulmonary fibrosis development. We can find the biological markers of pulmonary fibrosis according to the change of gene expression level or epigenetic modification and provide guidance for our subsequent research[28].

Fundamental pathological changes in pulmonary fibrosis include the accumulation of extracellular matrices, such as collagen and fibronectin, in the lung interstitium leading to respiratory failure[29].

Therefore, the expression of many collagen proteins was increased in pulmonary fibrosis tissues. Moreover, when hub gene screening was carried out, several collagen genes occupied very core positions.

The primary function of MMPs(metalloproteinases) is to degrade extracellular matrix proteins; the role of some matrix metalloproteinases (MMPs) is profibrotic, where others have anti-fibrotic functions[30]. Ivan O Rosas et al. analyzed the concentrations of 49 proteins in the plasma of 74 patients with IPF and the plasma of 53 control individuals. They found that MMP1 and MMP7 were significantly increased and were potential peripheral blood biomarkers in idiopathic pulmonary fibrosis[31]. Takwi Nkyimbeng et al. reported that in IPF, MMPs-1, 2, 7, 9, and 13 were significantly up-regulated in model mice and patients with pulmonary fibrosis[32]. Akihiko Sokai et al. reported that serum MMP-7 and -10 levels markedly correlated with both the percentage of predicted forced vital capacity and the percentage of predicted diffusing capacity of the lung for carbon monoxide and serum MMP-10 predicted clinical deterioration within 6 months and overall survival of IPF patients[33]. In our study, we found that MMP1, MMP7, MMP10, MMP11, and MMP13 were significantly up-regulated in three different data sets, among which MMP7, MMP10, and MMP11 were hypomethylated, suggesting that the change in methylation level may be a driving factor for the change in MMPs expression during pulmonary fibrosis. We also demonstrated elevated MMP7, MMP10, and MMP11 expression in bleomycin-induced pulmonary fibrosis in mice.

SPP1(Secreted Phosphoprotein 1) is involved in the attachment of osteoclasts to the mineralized bone matrix, and the encoded protein is secreted and binds hydroxyapatite with high affinity[34]. Christina Morse et al. found that in the lung tissue of lung fibers, macrophages with high SPP1 expression can aggravate pulmonary fibrosis[35]. SPP1 has been found to be a useful biomarker for several immune-mediated diseases, including multiple sclerosis, SLE, rheumatoid arthritis, atherosclerosis, cardiovascular disease, inflammatory bowel disease, and asthma[36–40]. Our study found that SPP1 was increased in multiple IPF data sets, and mRNA levels of SPP1 were also increased in the mice with pulmonary fibrosis model, suggesting that SPP1 may be involved in the development of IPF, but at present, SPP1 may be linked to the development of IPF.

DAPL1(Death Associated Protein Like 1) was reported to autonomously suppresses retinal pigment epithelium proliferation in vivo and in vitro[41]. Felix Grassmann et al. found that DAPL1 was associated with Age-related macular degeneration in a joint analysis of 3,229 cases and 2,835 controls from five studies[42]. At present, the biological function of DAPL1 is still unclear. Our research found that DAPL1 may be a biomarker of IPF, but the specific mechanism still needs to be further explored.

DOK5(Docking Protein 5) is a member of the DOK family of membrane proteins, which are adapter proteins involved in signal transduction. Hidekata Yasuoka et al. reported that DOK5 is up-regulated in systemic sclerosis and associated with IGFBP-5-induced fibrosis[43]. Lei Shi et al. found that Dok5 was the substrate of TrkB and TrkC receptors and involved in neurotrophin-induced MAPK activation[43]. At the same time, MAPK was considered to associate with the development of IPF[44].

FNDC4(Fibronectin Type III Domain Containing 4) is a member of the fibronectin type III domain family of proteins[45]. It was reported that FNDC4 signaled to macrophages to downregulate inflammation and

FNDC4 supplementation reduced disease severity in a mouse model for colitis, indicating therapeutic potential[46]. Our study found that the expression of FNDC4 was elevated in IPF and might be involved in the progress of IPF.

In conclusion, we screened out genes that were differentially expressed in three different IPF data sets by the methods of bioinformatics. Furthermore, GO and KEGG analysis suggested these genes involved in the extracellular matrix organization, and we found the hub genes related to the development of IPF. GSEA analysis indicated that in IPF disease, Asthma, Type I diabetes mellitus, Lupus erythematosus, the intestinal immune network for IgA production, p53 signaling pathway were highly activated; Besides, IPF-related genes regulated by methylation were identified by combining IPF methylation sequencing chips, and these genes were demonstrated in bleomycin-induced murine pulmonary fibrosis model.

Conclusion

In summary, through the integrated analysis of the mRNA sequencing data sets and methylation chip sequencing data sets of pulmonary fibrosis, we found that several hub genes play an essential role in the fibrosis process while the other 8 key genes were increased due to the decrease of methylation level, which promoted the progression of pulmonary fibrosis. These genes have potential as new biomarkers or therapeutic targets for IPF, but the exact mechanisms by which they affect pulmonary fibrosis progression need to be further explored.

Abbreviations

IPF: Idiopathic pulmonary fibrosis; DEGs: differentially expressed genes; GEO: Gene Expression Omnibus; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: protein-protein interaction; GSEA: Gene set enrichment analysis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All analyzed data are included in this published article and its supplementary information file. The original data are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Contributions

Shuaijun Chen and Jun Zhang carried out most of the experimental work and the analysis of data; Hong Ye and Wanli Ma provided scientific and administrative oversight for the conduct of the research and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We appreciate the Gene Expression Omnibus for the open data.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. DePianto DJ, Chandriani S, Abbas AR, Jia G, N'Diaye EN, Caplazi P, Kauder SE, Biswas S, Karnik SK, Ha C, et al: **Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis.** *Thorax* 2015, **70**:48-56.
2. Schafer MJ, White TA, Iijima K, Haak AJ, Ligresti G, Atkinson EJ, Oberg AL, Birch J, Salmonowicz H, Zhu Y, et al: **Cellular senescence mediates fibrotic pulmonary disease.** *Nat Commun* 2017, **8**:14532.
3. McDonough JE, Ahangari F, Li Q, Jain S, Verleden SE, Herazo-Maya J, Vukmirovic M, Deluliis G, Tzouveleakis A, Tanabe N, et al: **Transcriptional regulatory model of fibrosis progression in the human lung.** *JCI Insight* 2019, **4**.
4. Hata A, Nakajima T, Matsusaka K, Fukuyo M, Morimoto J, Yamamoto T, Sakairi Y, Rahmutulla B, Ota S, Wada H, et al: **A low DNA methylation epigenotype in lung squamous cell carcinoma and its association with idiopathic pulmonary fibrosis and poorer prognosis.** *Int J Cancer* 2020, **146**:388-399.
5. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**:e47.
6. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**:139-140.
7. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S: **ChAMP: 450k Chip Analysis Methylation Pipeline.** *Bioinformatics* 2014, **30**:428-430.

8. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 1999, **27**:29-34.
9. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L: **KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.** *Nucleic Acids Res* 2011, **39**:W316-322.
10. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.** *Nucleic Acids Res* 2019, **47**:D607-d613.
11. Bader GD, Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4**:2.
12. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for Gene Set Enrichment Analysis.** *Bioinformatics* 2007, **23**:3251-3253.
13. Richeldi L, Collard HR, Jones MG: **Idiopathic pulmonary fibrosis.** *Lancet* 2017, **389**:1941-1952.
14. Bonella F, Stowasser S, Wollin L: **Idiopathic pulmonary fibrosis: current treatment options and critical appraisal of nintedanib.** *Drug Des Devel Ther* 2015, **9**:6407-6419.
15. Cho JH, Gelinis R, Wang K, Etheridge A, Piper MG, Batte K, Dakhallah D, Price J, Bornman D, Zhang S, et al: **Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes.** *BMC Med Genomics* 2011, **4**:8.
16. Greenberg MVC, Bourc'his D: **The diverse roles of DNA methylation in mammalian development and disease.** *Nat Rev Mol Cell Biol* 2019, **20**:590-607.
17. Seebacher F, Krause J: **Epigenetics of Social Behaviour.** *Trends Ecol Evol* 2019, **34**:818-830.
18. Mafficini A, Scarpa A: **Genetics and Epigenetics of Gastroenteropancreatic Neuroendocrine Neoplasms.** *Endocr Rev* 2019, **40**:506-536.
19. Aloia L, McKie MA, Vernaz G, Cordero-Espinoza L, Aleksieva N, van den Ameele J, Antonica F, Font-Cunill B, Raven A, Aiese Cigliano R, et al: **Epigenetic remodelling licences adult cholangiocytes for organoid formation and liver regeneration.** *Nat Cell Biol* 2019, **21**:1321-1333.
20. Sun D, Zhang T, Su S, Hao G, Chen T, Li QZ, Bazzano L, He J, Wang X, Li S, Chen W: **Body Mass Index Drives Changes in DNA Methylation: A Longitudinal Study.** *Circ Res* 2019, **125**:824-833.
21. Perez MF, Lehner B: **Intergenerational and transgenerational epigenetic inheritance in animals.** *Nat Cell Biol* 2019, **21**:143-151.
22. Loo Yau H, Ettayebi I, De Carvalho DD: **The Cancer Epigenome: Exploiting Its Vulnerabilities for Immunotherapy.** *Trends Cell Biol* 2019, **29**:31-43.
23. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, 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**:788-824.

24. Raghu G, Chen SY, Yeh WS, Maroni B, Li Q, Lee YC, Collard HR: **Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11.** *Lancet Respir Med* 2014, **2**:566-572.
25. Strongman H, Kausar I, Maher TM: **Incidence, Prevalence, and Survival of Patients with Idiopathic Pulmonary Fibrosis in the UK.** *Adv Ther* 2018, **35**:724-736.
26. Kaunisto J, Kelloniemi K, Sutinen E, Hodgson U, Piilonen A, Kaarteenaho R, Mäkitaro R, Purokivi M, Lappi-Blanco E, Saarelainen S, et al: **Re-evaluation of diagnostic parameters is crucial for obtaining accurate data on idiopathic pulmonary fibrosis.** *BMC Pulm Med* 2015, **15**:92.
27. Ley B, Collard HR, King TE, Jr.: **Clinical course and prediction of survival in idiopathic pulmonary fibrosis.** *Am J Respir Crit Care Med* 2011, **183**:431-440.
28. Barratt SL, Creamer A, Hayton C, Chaudhuri N: **Idiopathic Pulmonary Fibrosis (IPF): An Overview.** *J Clin Med* 2018, **7**.
29. Selman M, Pardo A: **Idiopathic pulmonary fibrosis: an epithelial/fibroblastic cross-talk disorder.** *Respir Res* 2002, **3**:3.
30. Mahalanobish S, Saha S, Dutta S, Sil PC: **Matrix metalloproteinase: An upcoming therapeutic approach for idiopathic pulmonary fibrosis.** *Pharmacol Res* 2020, **152**:104591.
31. Rosas IO, Richards TJ, Konishi K, Zhang Y, Gibson K, Lokshin AE, Lindell KO, Cisneros J, Macdonald SD, Pardo A, et al: **MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis.** *PLoS Med* 2008, **5**:e93.
32. Nkyimbeng T, Ruppert C, Shiomi T, Dahal B, Lang G, Seeger W, Okada Y, D'Armiento J, Günther A: **Pivotal role of matrix metalloproteinase 13 in extracellular matrix turnover in idiopathic pulmonary fibrosis.** *PLoS One* 2013, **8**:e73279.
33. Sokai A, Handa T, Tanizawa K, Oga T, Uno K, Tsuruyama T, Kubo T, Ikezoe K, Nakatsuka Y, Tanimura K, et al: **Matrix metalloproteinase-10: a novel biomarker for idiopathic pulmonary fibrosis.** *Respir Res* 2015, **16**:120.
34. Fisher LW, Torchia DA, Fohr B, Young MF, Fedarko NS: **Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin.** *Biochem Biophys Res Commun* 2001, **280**:460-465.
35. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, Jiang Y, Kass DJ, Gibson K, Chen W, et al: **Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis.** *Eur Respir J* 2019, **54**.
36. Carecchio M, Comi C: **The role of osteopontin in neurodegenerative diseases.** *J Alzheimers Dis* 2011, **25**:179-185.
37. Murugaiyan G, Mittal A, Weiner HL: **Increased osteopontin expression in dendritic cells amplifies IL-17 production by CD4+ T cells in experimental autoimmune encephalomyelitis and in multiple sclerosis.** *J Immunol* 2008, **181**:7480-7488.
38. Singh M, Ananthula S, Milhorn DM, Krishnaswamy G, Singh K: **Osteopontin: a novel inflammatory mediator of cardiovascular disease.** *Front Biosci* 2007, **12**:214-221.

39. Glas J, Seiderer J, Bayrle C, Wetzke M, Fries C, Tillack C, Olszak T, Beigel F, Steib C, Friedrich M, et al: **The role of osteopontin (OPN/SPP1) haplotypes in the susceptibility to Crohn's disease.** *PLoS One* 2011, **6**:e29309.
40. Frenzel DF, Weiss JM: **Osteopontin and allergic disease: pathophysiology and implications for diagnostics and therapy.** *Expert Rev Clin Immunol* 2011, **7**:93-109.
41. Ma X, Li H, Wang Y, Wang J, Zheng Q, Hua J, Yang J, Pan L, Lu F, Qu J, Hou L: **DAPL1, a susceptibility locus for age-related macular degeneration, acts as a novel suppressor of cell proliferation in the retinal pigment epithelium.** *Hum Mol Genet* 2017, **26**:1612-1621.
42. Grassmann F, Friedrich U, Fauser S, Schick T, Milenkovic A, Schulz HL, von Strachwitz CN, Bettecken T, Lichtner P, Meitinger T, et al: **A Candidate Gene Association Study Identifies DAPL1 as a Female-Specific Susceptibility Locus for Age-Related Macular Degeneration (AMD).** *Neuromolecular Med* 2015, **17**:111-120.
43. Yasuoka H, Yamaguchi Y, Feghali-Bostwick CA: **The membrane-associated adaptor protein DOK5 is upregulated in systemic sclerosis and associated with IGFBP-5-induced fibrosis.** *PLoS One* 2014, **9**:e87754.
44. Nie Y, Yang Y, Zhang J, Cai G, Chang Y, Chai G, Guo C: **Shikonin suppresses pulmonary fibroblasts proliferation and activation by regulating Akt and p38 MAPK signaling pathways.** *Biomed Pharmacother* 2017, **95**:1119-1128.
45. Teufel A, Malik N, Mukhopadhyay M, Westphal H: **Frcp1 and Frcp2, two novel fibronectin type III repeat containing genes.** *Gene* 2002, **297**:79-83.
46. Bosma M, Gerling M, Pasto J, Georgiadi A, Graham E, Shilkova O, Iwata Y, Almer S, Söderman J, Toftgård R, et al: **FNDC4 acts as an anti-inflammatory factor on macrophages and improves colitis in mice.** *Nat Commun* 2016, **7**:11314.

Figures

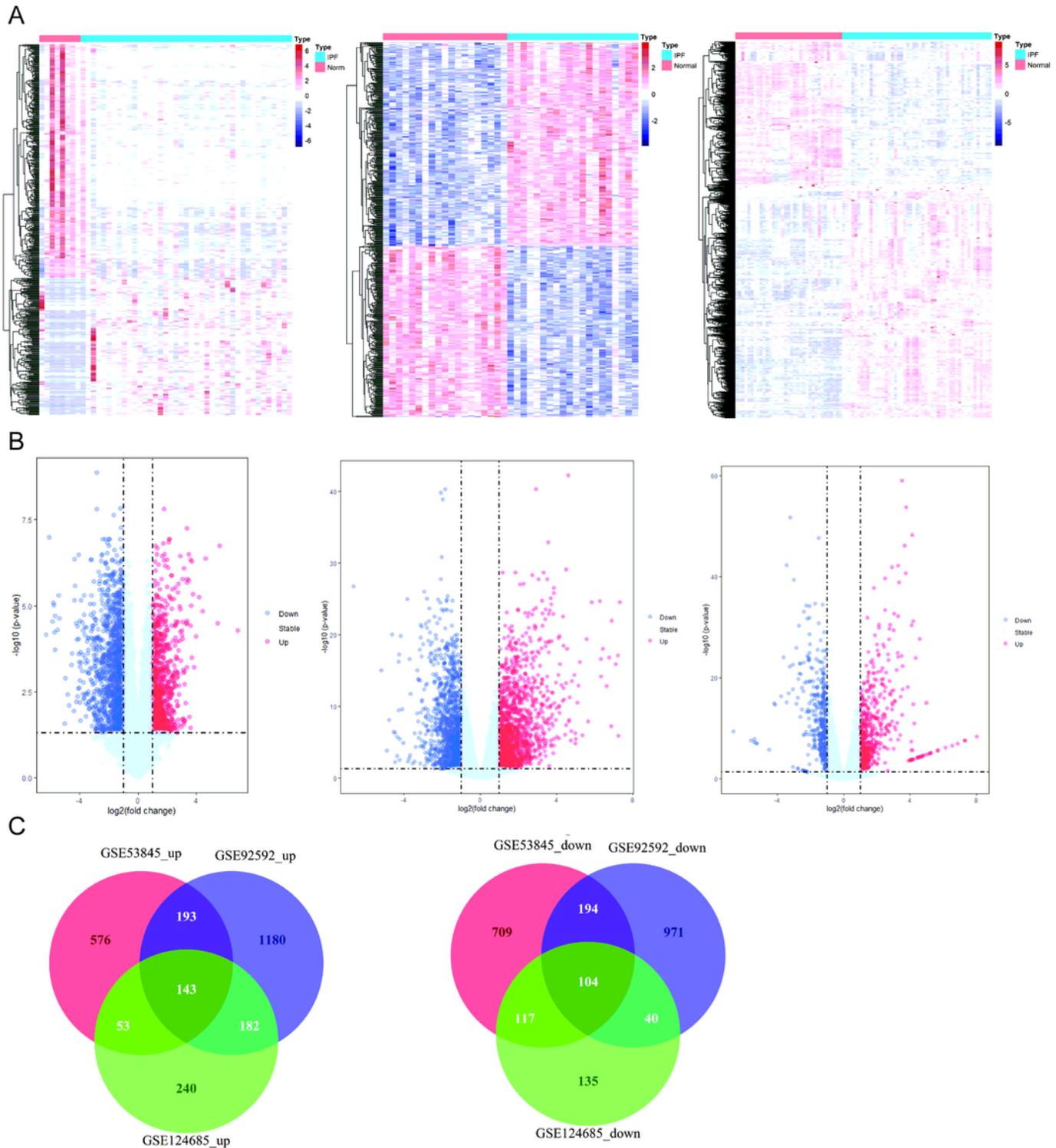
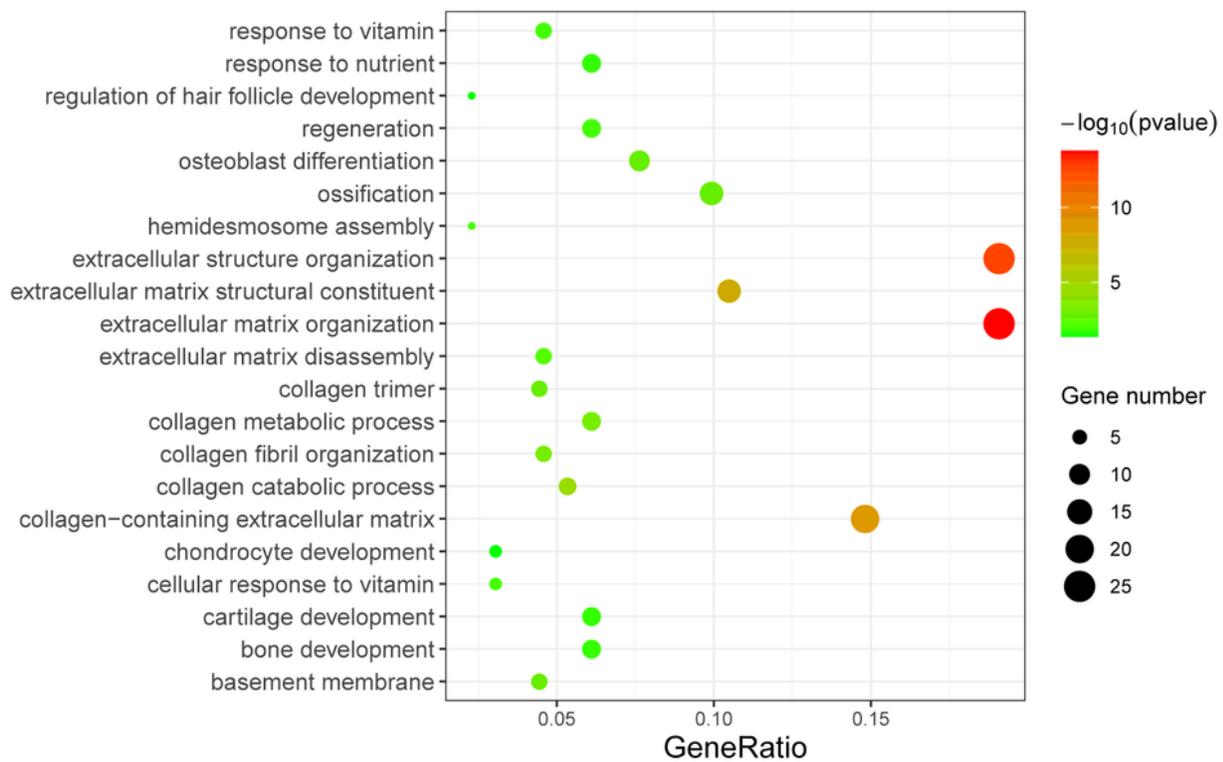


Figure 1

Differentially expressed genes between IPF patients and normal people. (A-C) The Volcano of differentially expressed genes in GSE92592, GSE124685, GSE53845, red dots represent high expression genes, and green dots represent low expression genes. (D-F) The heatmap of differential genes in GSE92592, GSE124685, GSE53845, red areas represent high expression genes, and green areas represent low expression genes. (G-H) The Venn diagram of differentially expressed genes between three gene sets.

A



B

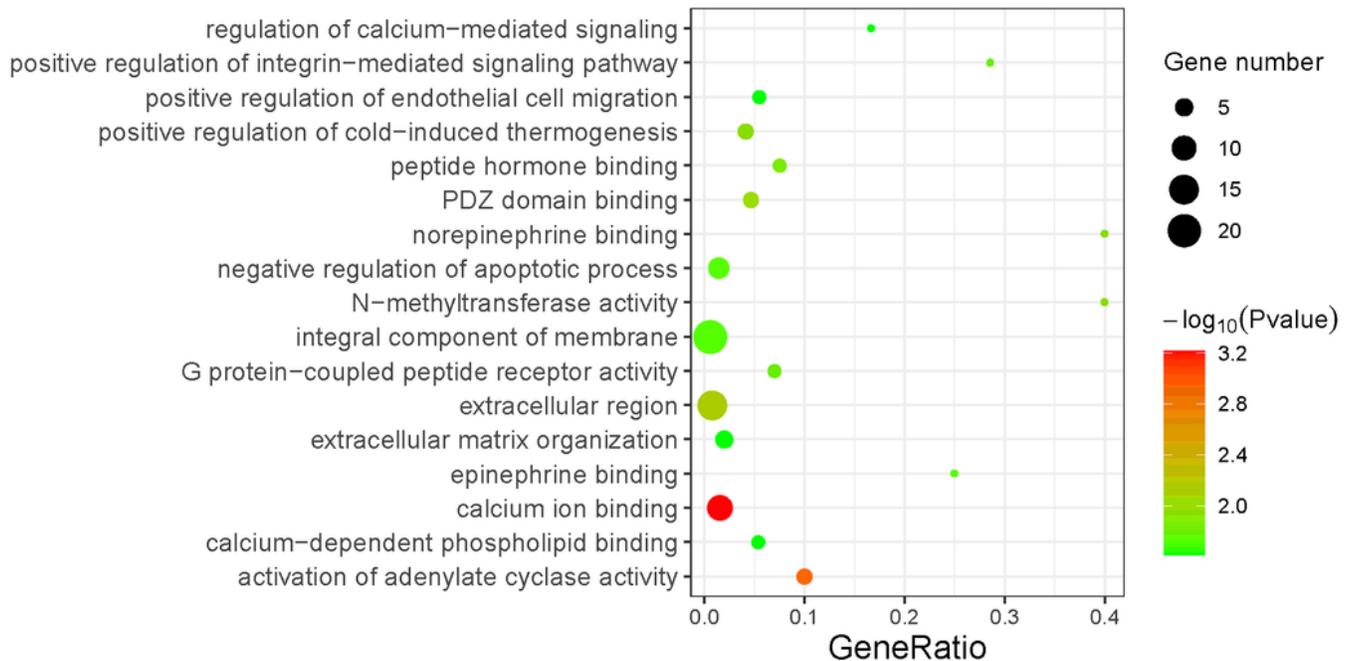
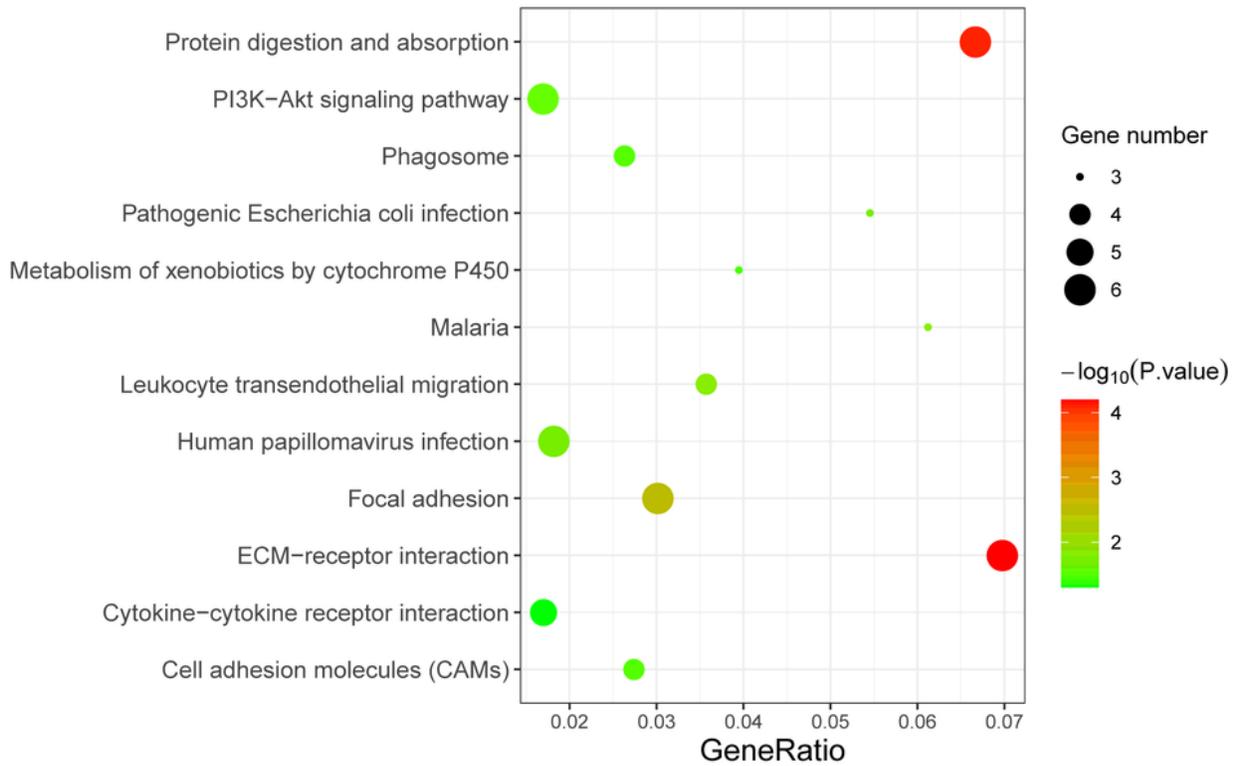


Figure 2

GO function enrichment analysis of differentially expressed genes. (A) The GO enrichment analysis of the up-regulated expressed genes; (B) The GO enrichment analysis of the downregulated expressed genes. Bubble plot of significant GO terms, the left is the name of the GO term, the bubble size on the right indicates the number of genes contained, and the color indicates the adjusted P-value.

A



B

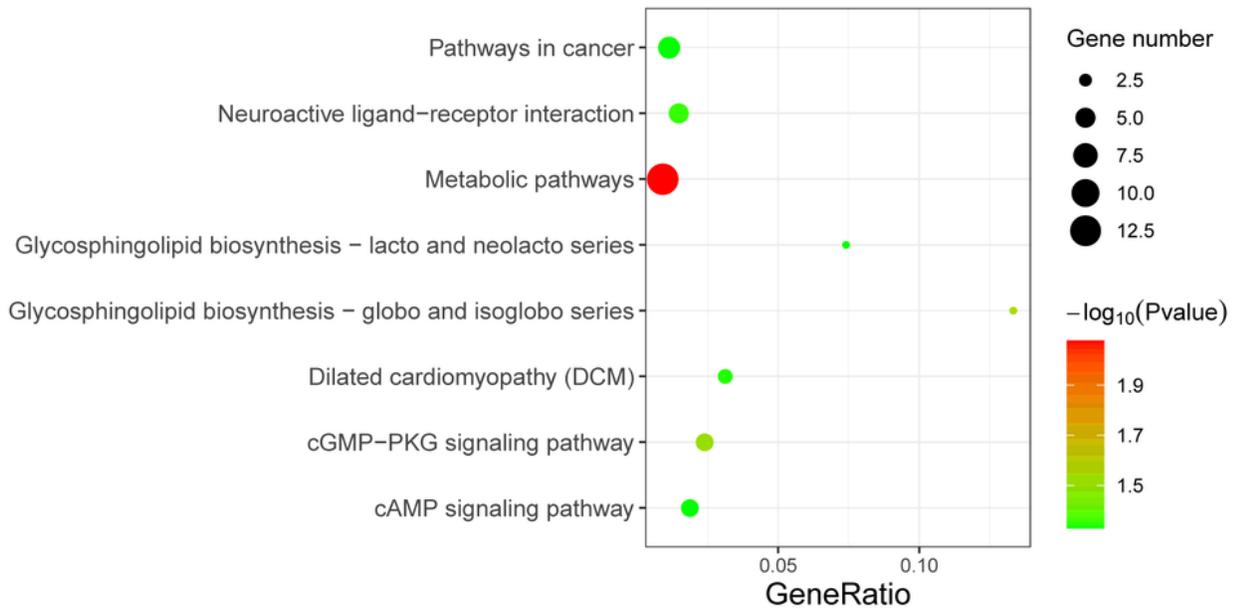
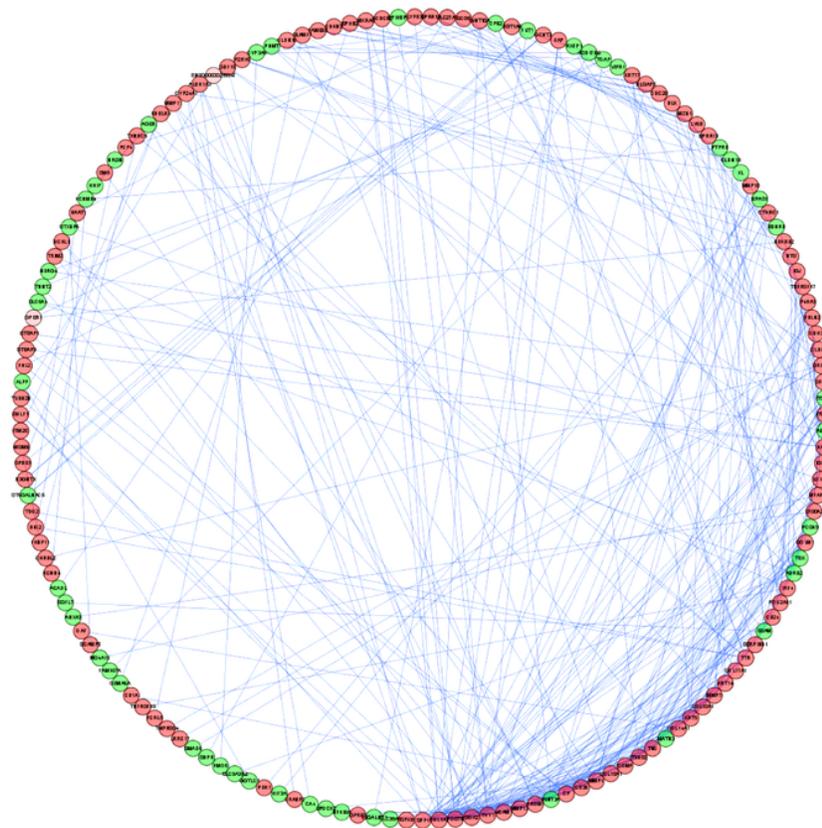


Figure 3

KEGG enrichment analysis of differentially expressed genes. (A) The KEGG enrichment analysis of the up-regulated expressed genes; (B) The KEGG enrichment analysis of the down-regulated expressed genes.

A



B

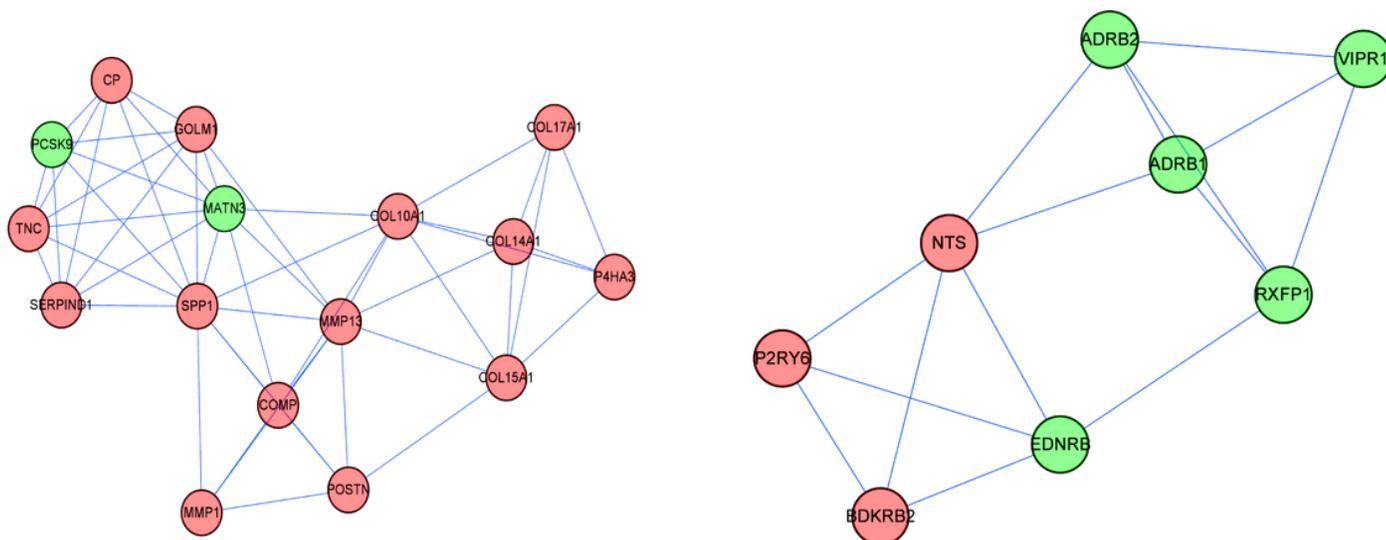


Figure 4

PPI network and the hub genes defined by MCODE. (A) PPI network of the differentially expressed genes; (B) Two subnetworks of hub genes defined by MCODE. The red circles represent genes with high expression, and the green circles represent genes with low expression.

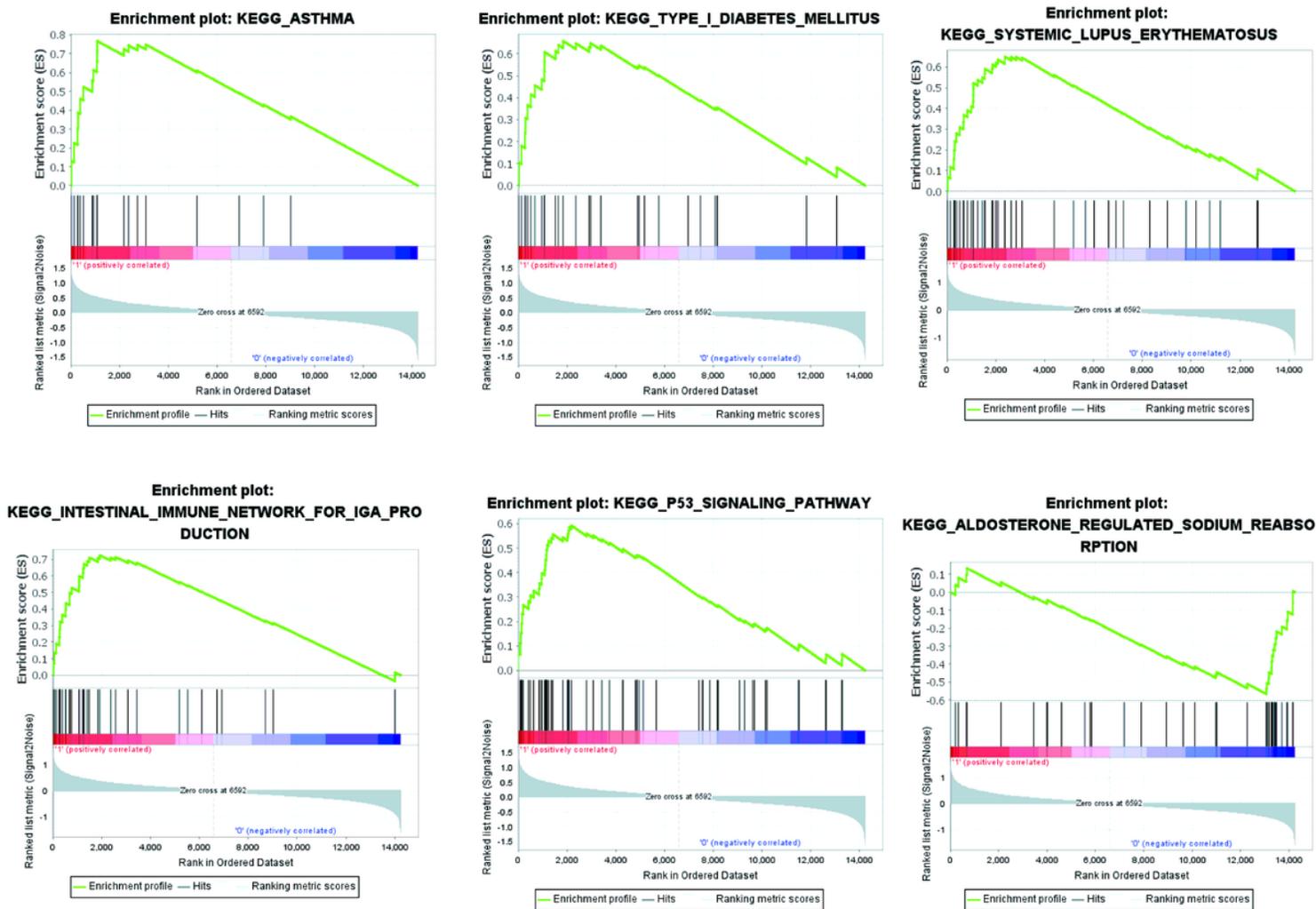


Figure 5

GSEA analysis of the IPF expression profile. (A-F) GSEA analysis of the IPF and Normal expression profile, The criteria of significantly enriched pathways were normalized p-value < 0.05 and the absolute value of normalized enrichment score (NES) > 1.5

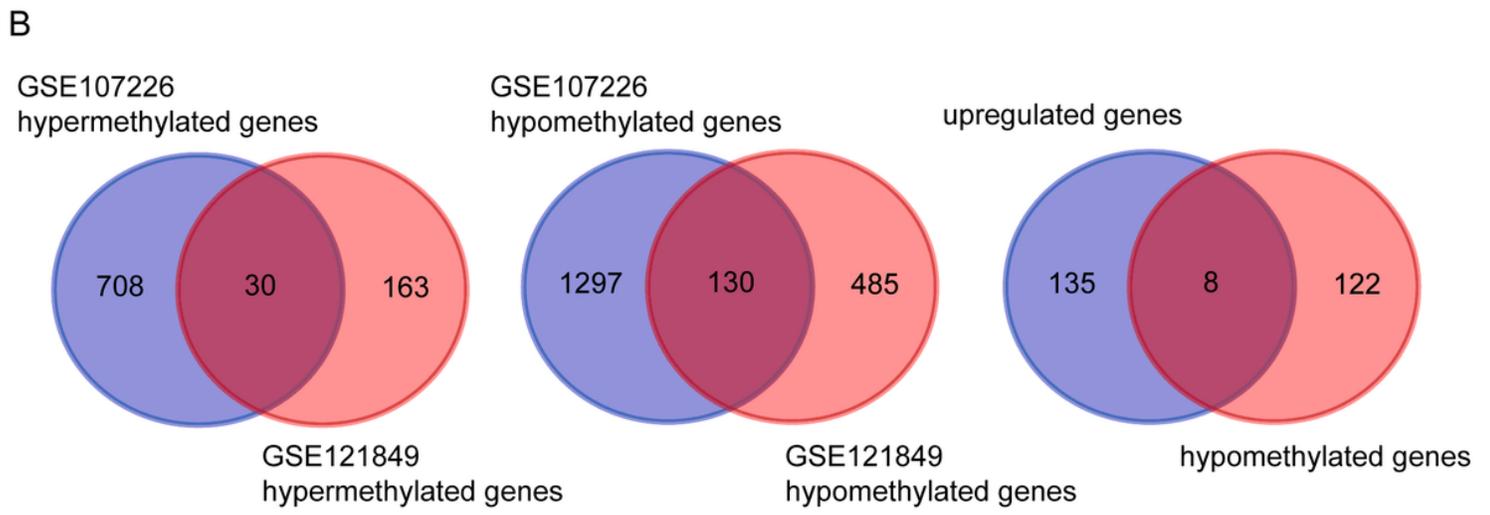
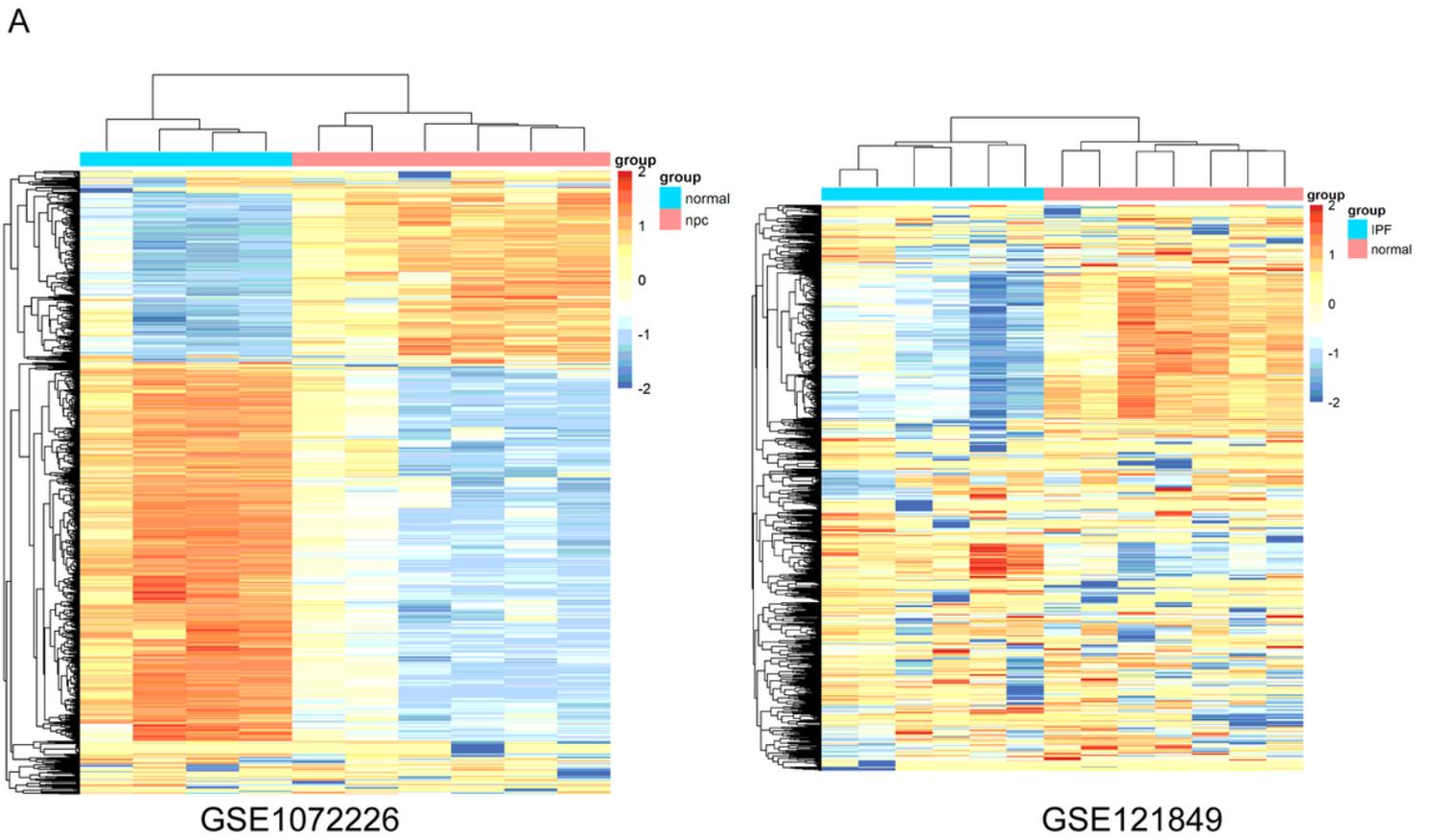


Figure 6

Differential methylated genes. (A) Heatmap of differential methylated genes between the IPF and normal lung tissues;(B) Venn diagram of differentially expressed genes and differentially methylated genes. (C) Venn diagram of up-regulated genes and hypomethylated genes.

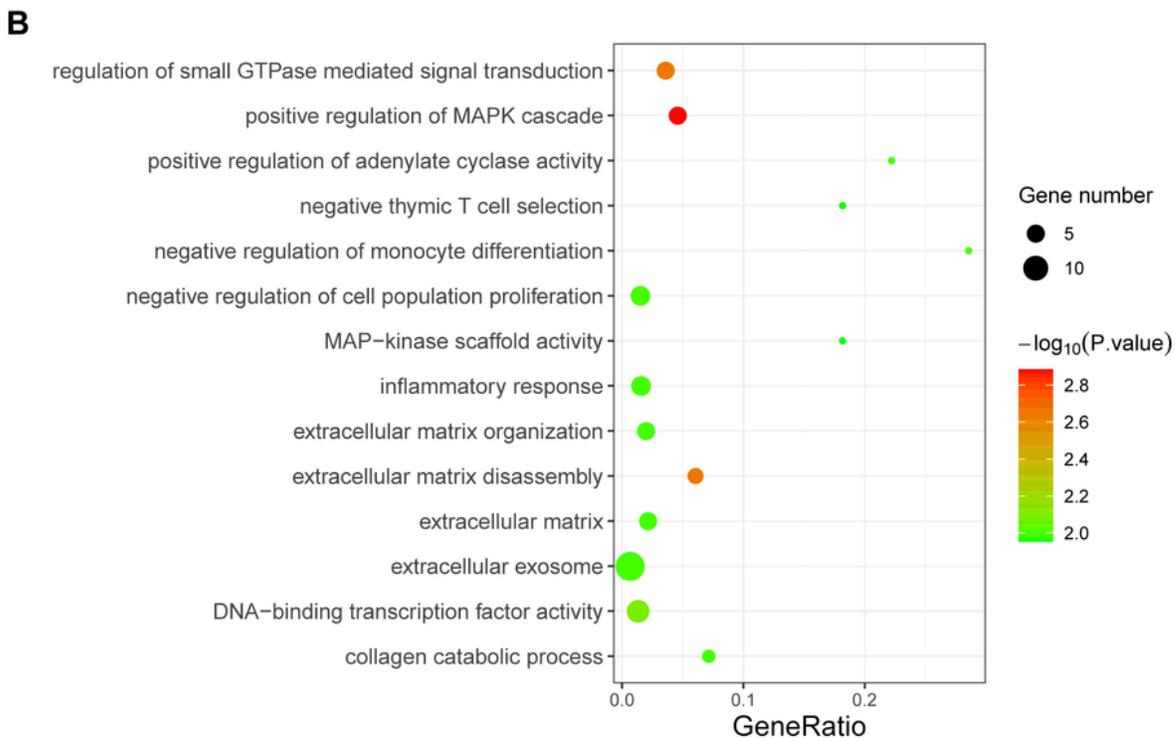
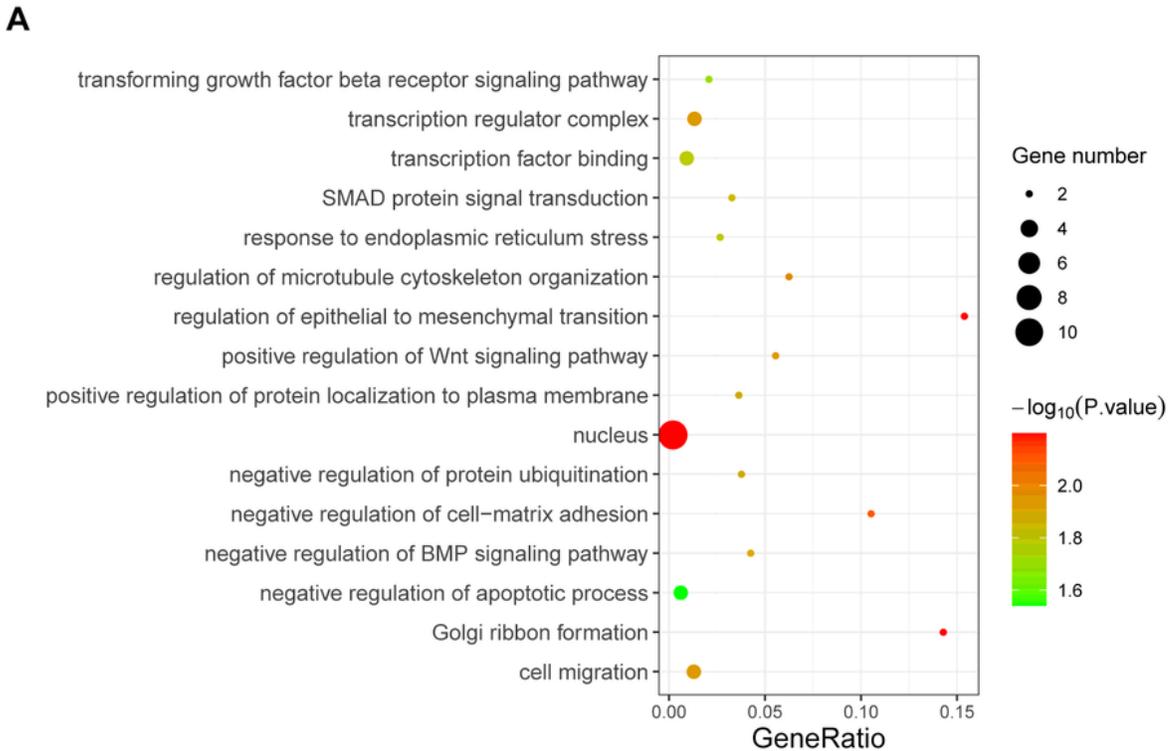


Figure 7

GO function enrichment analysis of differentially methylated genes. (A) The GO enrichment analysis of the hypermethylated genes; (B) The GO enrichment analysis of the hypomethylated genes.

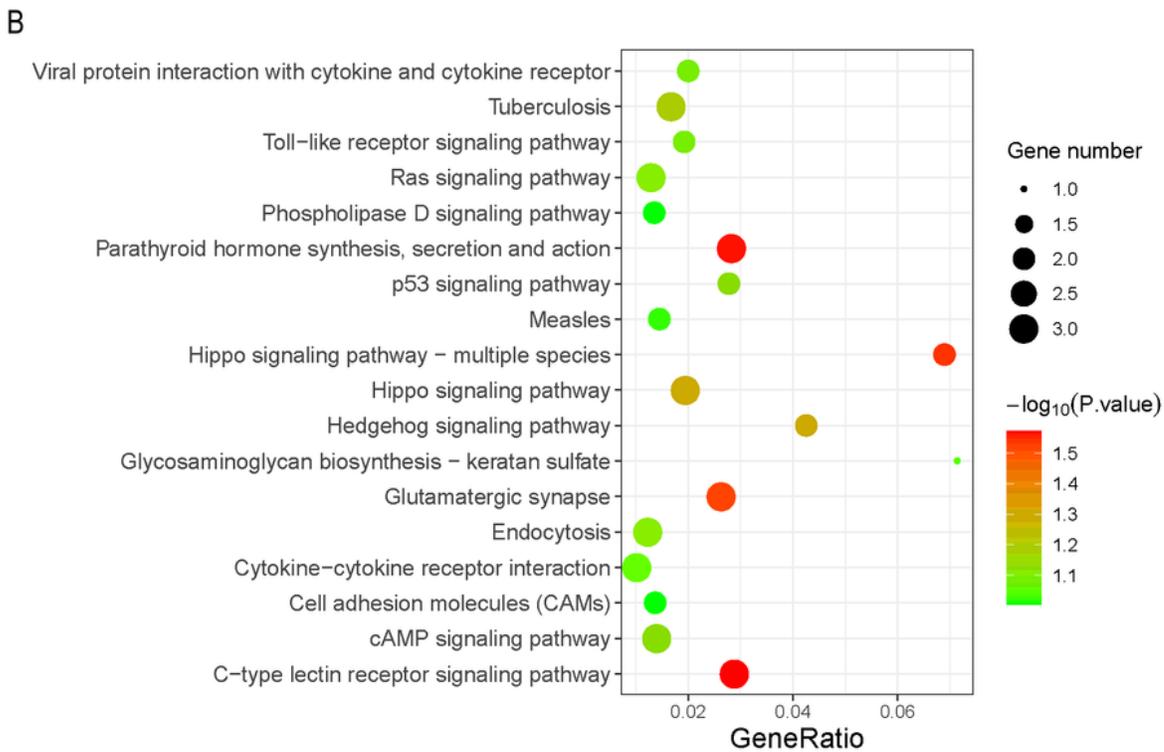
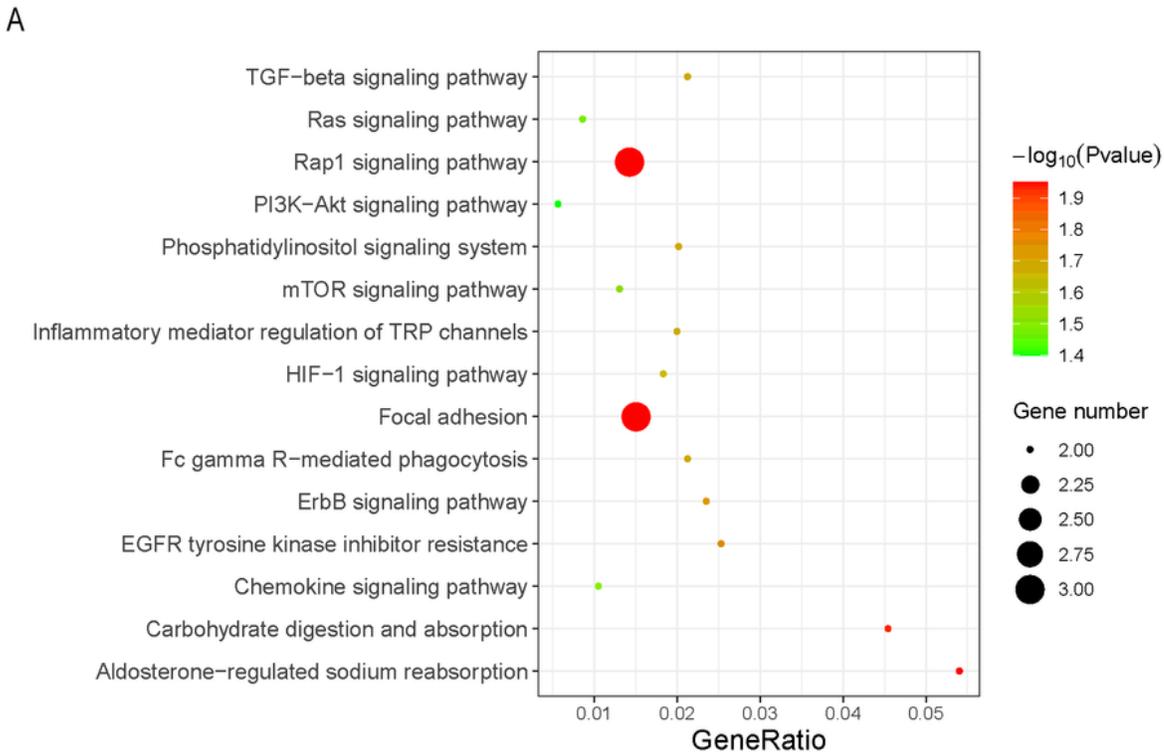


Figure 8

KEGG enrichment analysis of differentially methylated genes. (A) The KEGG enrichment analysis of the hypermethylated genes; (B) The KEGG enrichment analysis of the hypomethylated genes.

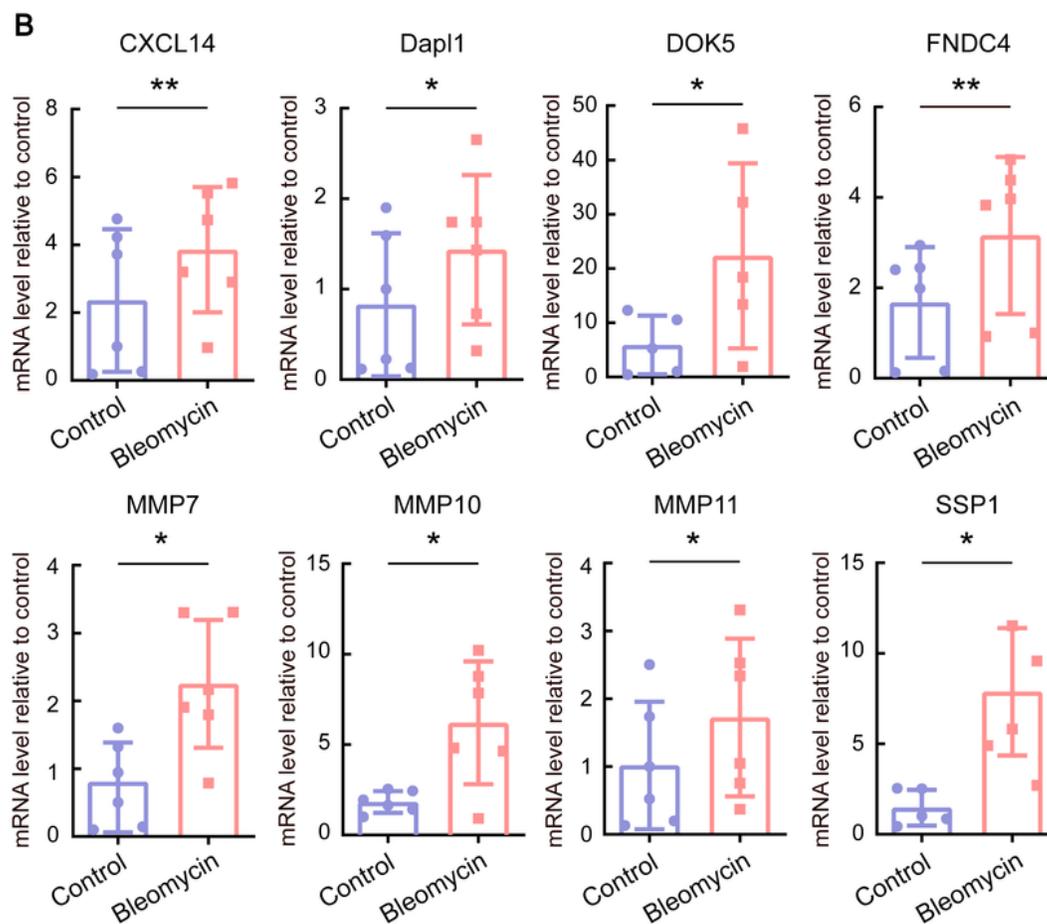
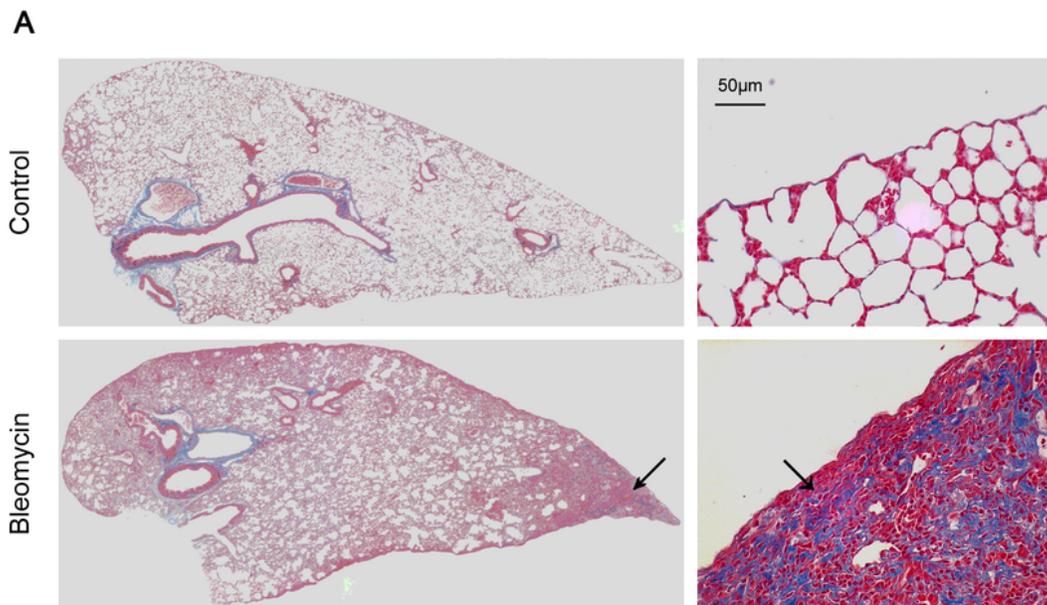


Figure 9

Validation in a mouse model. (A) Masson's trichrome staining of mouse lung tissues in which collagen was dyed blue;(B) mRNA levels of 8 genes in lung tissues of mice were detected by qPCR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1Listofdifferentiallyexpressedgenes.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6KEGG.xlsx](#)