

Transcriptomic and network analysis identifies shared and unique pathways and immune changes across fibrotic interstitial lung diseases

Wenhao Liu^{1,*}, Kangping Huang^{1,*}, Xin-Zhuang Yang², Ping Wang³

¹Eight-Year Program of Clinical Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

²Center for Bioinformatics, National Infrastructures for Translational Medicine, Institute of Clinical Medicine and Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China

³Department of Pulmonary and Critical Care Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100730, China

*Co-first author

Correspondence to: Ping Wang, Xin-Zhuang Yang; **email:** wangping6243@pumch.cn, xinzhuang_yang@163.com, <https://orcid.org/0000-0001-8345-9519>

Keywords: fibrotic interstitial lung diseases, idiopathic pulmonary fibrosis, fibrotic hypersensitivity pneumonitis, connective tissue disease-associated interstitial lung diseases, transcriptome data

Received: August 7, 2023

Accepted: December 1, 2023

Published: February 12, 2024

Copyright: © 2024 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Interstitial lung disease (ILD) encompasses a diverse group of disorders characterized by chronic inflammation and fibrosis of the pulmonary interstitium. Three ILDs, namely idiopathic pulmonary fibrosis (IPF), fibrotic hypersensitivity pneumonitis (fHP), and connective tissue disease-associated ILD (CTD-ILD), exhibit similar progressive fibrosis phenotypes, yet possess distinct etiologies, encouraging us to explore their different underlying mechanisms.

Methods: Transcriptome data of fibrotic lung tissues from patients with IPF, fHP, and CTD-ILD were subjected to functional annotation, network, and pathway analyses. Additionally, we employed the xCell deconvolution algorithm to predict immune cell infiltration in patients with fibrotic ILDs and healthy controls.

Results: We identified a shared progressive fibrosis-related module in these diseases which was related to extracellular matrix (ECM) degradation and production and potentially regulated by the p53 family transcription factors. In IPF, neuron-related processes emerged as a critical specific mechanism in functional enrichment. In fHP, we observed that B cell signaling and immunoglobulin A (IgA) production may act as predominant processes, which was further verified by B cell infiltration and the central role of CD19 gene. In CTD-ILD, active chemokine processes were enriched, and active dendritic cells (aDCs) were predicted to infiltrate the lung tissues.

Conclusions: This study revealed shared and specific molecular and cellular pathways among IPF, fHP, and CTD-ILD, providing a basis for understanding their pathogenesis and identifying potential therapeutic targets.

INTRODUCTION

Interstitial lung disease (ILD) refers to a group of lung diseases characterized by fibrosis and/or inflammation of

the lung interstitium. ILD can be idiopathic or associated with known causes such as connective tissue disease (CTD) and environmental exposures. Idiopathic pulmonary fibrosis (IPF), fibrotic hypersensitivity

pneumonitis (fHP), and CTD-associated ILD (CTD-ILD) are the most common three subtypes of fibrotic ILD, leading to progressive pulmonary fibrosis and respiratory failure [1]. Therefore, a profound understanding and comparison of their pathogenesis is indispensable.

The pathogenesis of IPF often includes pathological fibrosis in response to chronic lung tissue injury [2]. Fibroblasts and myofibroblasts play the most important roles in IPF, which overproduce extracellular matrix (ECM) through several dysfunctional signaling pathways [3]. Previous studies identified well-established mediators of IPF including transforming growth factor- β (TGF- β), fibroblast growth factor (FGF), CXCL12, CCL2, and tumor necrosis factor- α (TNF- α) [4]. The poor prognosis of IPF necessitates more investigation into its pathogenesis, including interactions of previously known mechanisms and new mechanisms [5].

fHP is a type of ILDs caused by continuous exposure to external antigens. fHP and IPF have some similar clinical, radiological, and pathological features, including progressing fibrosis. As antigens contribute to triggering of fHP, T lymphocytes and granuloma are involved in the pathophysiology of fHP [6], Th1 inflammatory response is canonical in the development of fHP while the Th17 response might be also involved in its pathogenesis [7]. Recent bioinformatic studies found that fHP and IPF have common molecular patterns, and HP-specific genes may be *BGN* and *CXCL9*, which are related to inflammation initiation and granulomatous diseases [8]. However, the specific molecular features of fHP remain to be explored.

CTD-ILD is a heterogeneous conditions and can exhibit similar clinical and imaging patterns with IPF. Systemic sclerosis, rheumatoid arthritis, and mixed connective tissue disease are the most common conditions that may concur with progressive pulmonary fibrosis. However, the treatment of CTD-ILD and IPF is different, and our understanding of this disparity is limited [9]. Previous mechanistic studies demonstrated that myofibroblasts possess profibrotic and anti-apoptotic properties in systemic sclerosis associated ILD (SSc-ILD), while alveolar epithelial cells play a more significant role in IPF [10]. Immune mechanisms including innate and adaptive immunity, especially profibrotic and anti-fibrotic immune cells, were also reported to play a role in CTD-ILD [11, 12]. So far, no definitive pattern has been identified to distinguish CTD-ILD from IPF, thus it is crucial to evaluate the distinct features of CTD-ILD.

Bioinformatics has provided new insight into different fibrotic ILDs [13–15]. However, many transcriptome studies of lung tissues were limited by small sample sizes and a narrow focus on specific types of fibrotic

ILDs, potentially introducing bias. Hence, there is an urgent need for a comprehensive investigation encompassing multiple ILDs with distinct etiologies yet sharing similar pathological patterns. Furthermore, previous studies have not adequately integrated the expression patterns, regulatory mechanisms, and cellular alterations within lung tissue samples. In this study, we utilized various databases to perform network and pathway analyses. We also incorporated transcription factors (TFs) data mining to investigate gene expression regulation patterns, and immune infiltration analysis to discover cellular changes. Through this comprehensive approach, our objective is to elucidate the differences and commonalities of the pathogenesis of the three prevalent fibrotic ILDs, thereby enabling more precise and targeted treatments for patients.

MATERIALS AND METHODS

Datasets employed in this study

The microarray datasets used in this study were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). The criteria for retrieval were: A) samples were from human lung tissue, B) gene expression pattern was profiled, C) datasets contained both patients and healthy individuals without a history of pulmonary fibrosis, D) the diagnosis of IPF was based on standard American Thoracic Society criteria [16], E) The diagnosis of fHP was made according to accepted guidelines or by multidisciplinary consensus using ATS criteria [8, 17], F) The RA-UIP subjects had evidence of UIP on lung biopsy and rheumatoid arthritis (RA) defined by the American College of Rheumatology and European League Against Rheumatism guidelines [18]. In addition, all patients with systemic sclerosis (SSc) met the American College of Rheumatology criteria for the diagnosis of SSc [19].

To ensure the consistency and completeness of the datasets, we manually identified relevant literature using keywords filters and applied R programming language (version: 4.1.3) for subsequent analysis. Finally, IPF dataset (GSE175457) [20], fHP dataset (GSE150910) [8], and CTD-ILD datasets (SSc-ILD: GSE48149 [21, 22]; RA-ILD: GSE199152 [23]) were included, each contributing unique insights into various lung conditions. We merged CTD-ILD datasets and corrected the batch effects using the “combat” function in the SVA package (version: 3.38.0) [24]. Next, we normalized the merged datasets and adjusted for covariates using the “Normalize between arrays” and “remove Batch Effect” functions in the limma package (version: 3.46.0) [25]. Table 1 summarizes the included datasets.

Table 1. The information of all the datasets employed in this study.

Data sets*	Data		Tissue sources	References	Category	GPL
	Case	Control				
GSE48149	44 ^a	9	Lung tissue	Hsu et al., 2011; Renaud et al., 2020	CTD-ILD (SSc-ILD)	GPL16221
GSE199152	23 ^b	4	Lung tissue	Vassallo et al., 2022	CTD-ILD (RA-ILD)	GPL16791
GSE150910	185 ^c	103	Lung tissue	Furusawa et al. 2020	CHP	GPL24676
GSE175457	234	188	Lung tissue	Borie et al., 2022	IPF	GPL24676

*: All data sets used in this study contain a total of 646 samples, among which there were 342 cases and 304 controls. All samples were collected in the lung tissue. ^a: 23 lung samples from SSc-ILD patients were employed in this study. ^b: 3 lung samples from RA-ILD patients were employed in this study. ^c: 82 lung samples from CHP patients were employed in this study.

Identification of differentially expressed genes and functional annotation

To identify differentially expressed genes (DEGs) in lung tissue samples from ILD patients and healthy controls, we performed differential expression analysis using the limma package (version: 3.46.0), controlling for age. The threshold for screening DEGs was $|\log_2 \text{FC (fold change)}| > 1$ and false discovery rate (FDR) < 0.01 . Upregulated and downregulated common DEGs for three ILDs were then subjected to functional annotation.

Enrichment analysis of Gene Ontology (GO) and Disease Ontology (DO) was performed on common DEGs using the clusterProfiler package (version: 3.18.1) [26]. Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>) and gene set enrichment analysis (GSEA) were performed for common DEGs. The threshold for significant differences in the aforementioned enrichment analyses was set at $\text{FDR} < 0.05$.

Network and pathway analysis

To explore the functional interaction between the common DEGs of IPF, fHP, and CTD-ILD, PPI network analysis was done using the STRINGdb package (version: 2.6.5) with a confidence score of ≥ 0.7 [0,1]. Entrez gene identifiers of DEGs were imported into STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (Version: 11.5) [27] for network and pathway analyses. We imported the information of the PPI network into the igraph package (version: 1.12.1) and selected the minimum connected network for further pathway analyses. To explore the unique biological pathways, we individually used the pathway analysis data derived from the KEGG database on the PPI network of IPF, fHP, and CTD-ILD.

Transcription factors analysis

The common DEGs were imported into Cytospace for network analysis of TFs [28]. RcisTarget package [29]

was used to acquire information on TFs and gene targets, and an adjusted P-value < 0.05 was considered significant. Subsequently, we separately identified the unique TFs in IPF, fHP, and CTD-ILD.

Immune infiltration analysis

The xCell deconvolution algorithm was applied in this study [30]. We used the “immunedeconv” package to comprehensively analyze tissue immune infiltration based on gene expression profiles [31]. Following the application of the xCell deconvolution algorithm, we obtained the estimated proportions of various immune cells in the lung tissues of ILD patients and healthy controls. Student’s t-test was used to investigate the differences in immune cell infiltration between these groups.

Statistical analysis

For this study, statistical computations were performed using R (version 4.2.2). Microarray data, sourced from the GEO database, were selected based on strict criteria ensuring sample relevance and quality. Differential expression analysis was executed via the limma package, with age as a covariate, identifying DEGs at a $|\log_2 \text{FC}| > 1$ and $\text{FDR} < 0.01$ threshold. Functional annotations were derived using the clusterProfiler package, with an $\text{FDR} < 0.05$ denoting significance. PPI networks, constructed using STRINGdb with a confidence score ≥ 0.7 , were analyzed in igraph to discern minimum connected networks for pathway analysis against KEGG database entries. Transcription factor associations were determined using the RcisTarget package, with significance set at an adjusted P-value < 0.05 . Immune cell proportions in lung tissue were estimated using the xCell algorithm within the immunedeconv package, with Student’s t-tests assessing differences in infiltration between patient and control groups. These R scripts for replication of this research can be found in our dedicated online repository at <https://github.com/Robinwhliu/IPF-fHP-CTD-ILD>.

Data available statement

Lists of genes with varying expression across three diseases are presented in Supplementary Tables 1–3. Comprehensive gene differential and functional enrichment analyses are included in Supplementary Tables 4–10. The full expression matrix employed in this research can be found in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). The R programming script for reproducing our analyses and graphical representations is also open for download on GitHub at the provided URL (<https://github.com/uuihbgg1/IPF-fHD-CTD>).

RESULTS

Differentially expressed genes

Four independent studies met our inclusion criteria (Table 1) were identified. First, a dataset consisting of 26 patients with CTD-ILD and 13 matched healthy controls was generated by merging two IS datasets: GSE48149 and GSE199152 (Table 2). Concurrently, the fundamental clinical characteristics of the CHP and IPF datasets are presented in Tables 3, 4, respectively. We controlled batch effects and normalized different subsets to ensure data consistency. The results showed that data pre-processing was effective and reliable (Supplementary Figure 1). The results of PCA analysis showed that data pre-processing was effective and reliable (Supplementary Figure 2). In total, 1059 DEGs for IPF, 602 DEGs for fHP, and 321 DEGs for CTD-ILD were identified between patients and healthy controls (Figure 1). We found 105 common DEGs between IPF, fHP, and CTD-ILD, among which 74 DEGs were upregulated, and 29 DEGs were downregulated (Figure 2).

Shared biological pathways in IPF, fHP, and CTD-ILD

For biological processes, 30 pathways reached statistical significance (Figure 3). Upregulated genes were associated with extracellular matrix changes while downregulated genes were relevant to cognition and sensory perception. For cellular components, upregulated genes were involved in the extracellular matrix, collagen trimming, and intermediate filament. The molecular function significantly associated with common upregulated DEGs was extracellular matrix structural constituent, structural constituent of the cytoskeleton, glycosaminoglycan binding, and heparin binding. Peptide, amide, carbohydrate binding, neurotransmitter receptor activity, and glutamate receptor activity were associated with the enrichment of downregulated genes. More information for GO and DO

enrichment analyses is presented in Figure 3A, 3B and Supplementary Figures 3, 4.

Furthermore, the GSEA results demonstrated that the enriched molecular pathways related to the upregulated DEGs were degradation of the extracellular matrix and collagen, formation of the cornified envelope and collagen, activation of matrix metalloproteinases, and keratinization (Figure 3C). Meanwhile, downregulated DEGs were associated with monitoring of oxygen-dependent proline hydroxylation of hypoxia-inducible factor α (HIF-1 α), pluripotent stem cells, GPCR ligand binding, G α (q) signaling, transmission across chemical synapses, O₂/CO₂ exchange, and pexophagy. These results, particularly the results upregulated DEGs, were consistent with those from GO enrichment analysis, indicating extracellular environment involvement.

Unique biological pathways in IPF, fHP, and CTD-ILD

Network analysis of upregulated genes identified in the differential analysis of the IPF dataset resulted in a unique network centered on Gremlin 2 (GREM2). GREM2 encodes an antagonist of bone morphogenetic protein (Figure 4A). This network was predominantly enriched in genes associated with neuron-related pathways, including axon development, axonogenesis, axon guidance, and neuron projection guidance (Figure 4C). Network analysis of downregulated genes identified in the differential analysis of the IPF dataset resulted in a particular network centered on chromosome 1 open reading frame 115 (C1orf115) and odontogenic ameloblast-associated protein (ODAM) (Figure 4B). Downregulated genes were predominantly enriched in angiogenesis, epithelial cell migration, and endothelial cell migration (Figure 4D).

The same network analysis was performed for the fHP dataset. The unique network of upregulated genes in fHP was centered on the cluster of differentiation 19 (CD19) and matrix metalloproteinase 9 (MMP9) and was enriched predominantly in immune-related pathways (Figure 5A, 5C). Among them, B cell receptor signaling pathway, primary immunodeficiency, and intestinal immune network for IgA production and cell adhesion molecules were the upregulated pathways. Downregulated genes formed a network centered on vascular endothelial growth factor A (VEGFA), protein tyrosine kinase 2 (PTK2), and kinase insert domain receptor (KDR), which were enriched in protein digestion and absorption, arginine biosynthesis, adherens junction, and PI3K/Akt signaling pathway (Figure 5B, 5D).

Analysis of upregulated genes in CTD-ILD resulted in a unique network centered on ATP binding cassette

Table 2. Clinical characteristics of the merged CTD-ILD data set.

	CTD-ILD (N=26)	Controls(N=13)	P-value
Gender (Male/Female)	9/17	5/8	1
Age at study visit (mean, years)	49.5 ± 10.2	53.3 ± 6.1	0.7510
Smoking history	10/16	4/9	0.7334

Table 3. Clinical characters of the fHP data set.

Characteristic	fHP (n = 82)	Control (n = 103)	P-Value
Age, years	59.4 ± 10.6	59.9 ± 10.2	0.98
Sex	n = 81	n = 103	0.23
Male	38 (47%)	45 (44%)	
Female	43 (53%)	58 (56%)	
Race	n = 75	n = 103	
Non-Hispanic white	59 (79%)	87 (84%)	0.055
Hispanic	10 (13%)	4 (4%)	—
Asian	4 (5%)	3 (3%)	—
Black	1 (1%)	9 (9%)	—
Other	1 (1%)	0 (0%)	—
Smoke	n = 73	n = 96	0.9
Ever	33 (45%)	43 (45%)	
Never	40 (55%)	53 (55%)	
Sampling method			0.51
Surgical lung biopsy	26 (32%)	41 (40%)	
Transplant	56 (68%)	62 (60%)	
MUC5B genotype			<0.001
GG	52 (63%)	80 (78%)	
GT	29 (35%)	21 (20%)	
TT	1 (2%)	2 (2%)	
Minor allele frequency	0.19	0.12	—

Data are presented as mean ± standard deviation or number (%).

fHP, fibrotic hypersensitivity pneumonitis; yr, years; M, male; F, female.

Table 4. Clinical characters of the IPF data set.

Characteristic	IPF (n = 234)	Control (n = 188)	P-Value
Age, years	61.4 ± 7.5	55.3 ± 16.8	9.92 x 10 ⁻⁶
Sex	n = 234	n = 188	1.11x10 ⁻⁵
Male	184 (79%)	110 (59%)	
Female	50 (21%)	78 (41%)	
Race	n = 234	n = 188	0.42
White	200 (85%)	155 (82%)	—
Non-White	34 (15%)	33 (18%)	—
Smoke	n = 234	n = 188	5.7 x10 ⁻³
Ever	140 (60%)	86 (46%)	
Never	75 (32%)	89 (47%)	
Unknown	19 (8%)	13 (7%)	

Data are presented as mean ± standard deviation or number (%).

IPF, idiopathic pulmonary fibrosis.

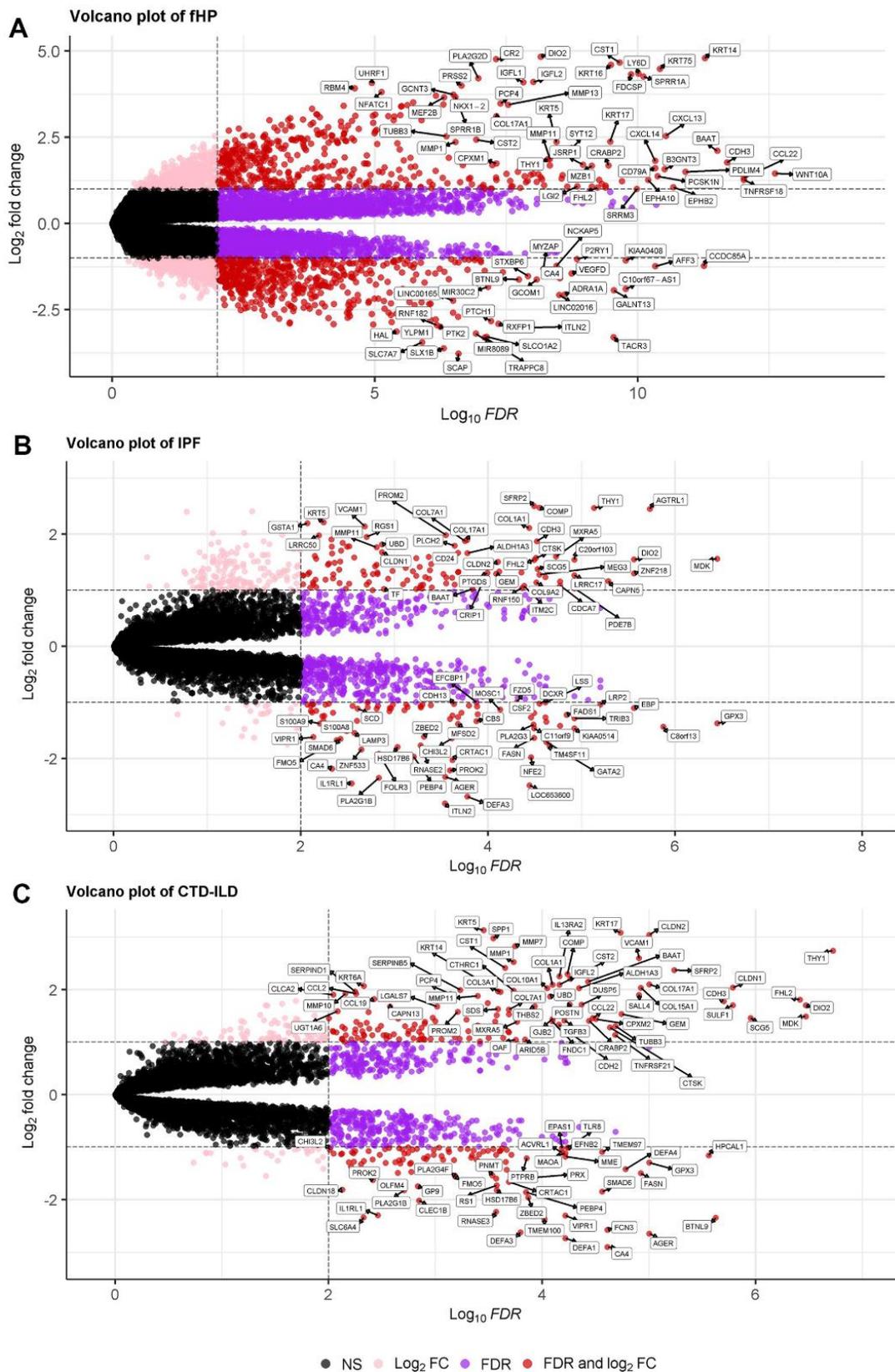


Figure 1. (A–C) Volcano plot demonstrating an overview of the differential expression of all genes in fHP, IPF and CTD-ILD. The threshold in the volcano plot was $-\log_{10}$ adjusted $P > 2$ and $|\log_2$ fold change > 0.5 ; red dots indicate significant differentially expressed genes. Note: CTD-ILD expression matrix was the merge of SSC-ILD and RA-ILD datasets.

subfamily C member 3 (ABCC3) and was enriched in pathways related to cell chemotaxis (Figure 6A, 6C). Network analysis of downregulated genes in CTD-ILD exhibited a network centered on agouti-related neuropeptide (AGRP) and midline 1 interacting protein 1 (MID1IP1), and enriched predominantly in mucosal immune response (Figure 6B, 6D).

Identification of regulatory transcription factors

We found six possible critical TFs regulating the expression of these common DEGs including two TFs from the p53 family (Figure 7A). We next investigated unique TFs among the three types of ILDs (Figure 7B). Venn diagram analysis showed that CTD-ILD has the most significant number of specific TFs (33 specific TFs), while fHP and IPF have 12 and 9 specific TFs, respectively.

Evaluation of tissue-infiltrating immune cells

The results of the violin plot showed the top 5 significant differences in immune cell proportions between healthy individuals and patients with IPF, fHP, and CTD-ILD (Figure 8). The increased abundance of DCs, including induced-DCs, was common in the three ILDs. Meanwhile, patients with IPF and fHP exhibited a decrease in natural killer T cell abundance, and patients with IPF and CTD-ILD had decreased abundance of microvascular endothelial cells. Lymphatic endothelial cells were more abundant in IPF; total B cells and class-switched memory B cells were more abundant in fHP; and activated dendritic cells (aDC) and neutrophils were more abundant in CTD-ILD. Furthermore, we observed unique variants of immune cells in the three ILDs (Supplementary Figure 5).

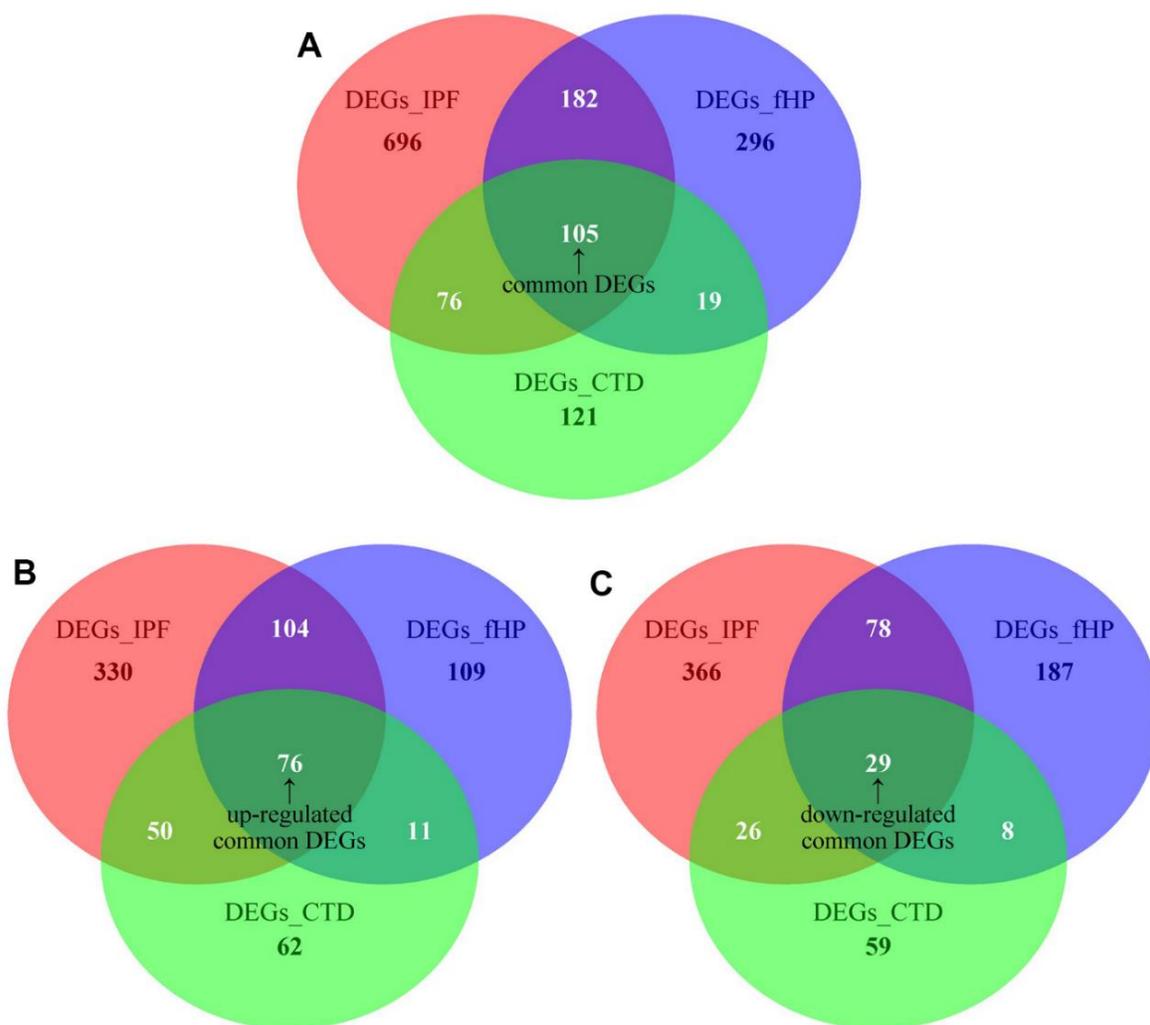


Figure 2. (A) Venn diagram demonstrates the common DEGs of fHP, IPF and CTD-ILD. (B, C) Venn diagrams demonstrate the up-regulated and down-regulated common DEGs of fHP, IPF and CTD-ILD, respectively.

DISCUSSION

This study revealed that IPF, fHP, and CTD-ILD exhibit comparable downstream molecular patterns but divergent upstream drivers at the transcriptome level. Surprisingly, IPF demonstrated upregulated biological processes associated with neuronal and developmental processes, while fHP showed a prominent involvement of the B cell signaling pathway and intestinal IgA production. CTD-ILD had positive regulation in chemotaxis and weakened response to various pathogens and external stimuli. The diversity of cellular changes may

contribute to the different clinical phenotypes, with endothelial cells being more prominent in IPF, total B cells and class-switched memory B cells in fHP, and aDCs and neutrophils in CTD-ILD.

The analysis of shared DEGs revealed that the same molecular mechanisms in IPF, fHP, and CTD-ILD lead to progressive fibrosis. The results of GO enrichment and GSEA further supported this observation, indicating significant alterations in profibrotic ECM changes, such as ECM structure organization and collagen metabolism. Progressive fibrosis in pathological assessment of IPF,

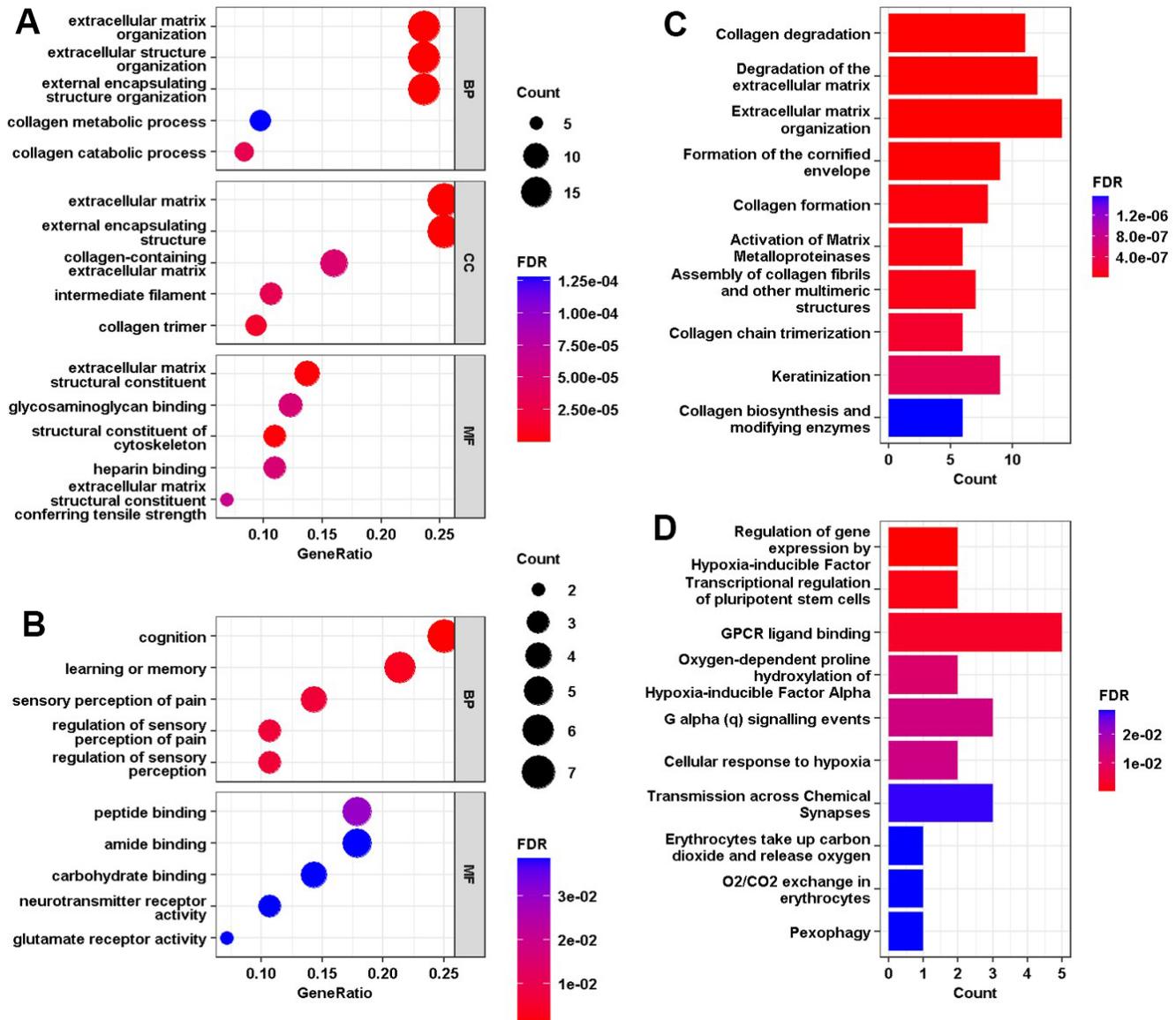


Figure 3. (A, B) GO enrichment analysis of the up-regulated and down-regulated common DEGs, where the horizontal axis represents the proportion of DEGs under the GO term. Top 5 pathways with most significant P-value were shown and ordered by gene ratio. BP, biological process; CC, cellular component; MF, molecular function. (C, D) Gene set enrichment analysis of the up-regulated and down-regulated common DEGs, where the horizontal axis represents the number of DEGs under the gene set enrichment analysis (GSEA) terms.

fHP, and CTD-ILD can verify these findings [32]. Oxygen-dependent proline hydroxylation of HIF-1 α in the GSEA indicated disinhibition of HIF-1 α in all three ILDs [33]. Erythrocyte-related processes, including carbon dioxide transfer, oxygen release, and O₂/CO₂

exchange were suppressed in all three ILDs, verifying the role of hypoxia in all three ILDs. In the TF analysis, common genes were found to be associated with p53 (encoded by *TP53*) and a p53 family member, p63, suggesting that these TFs may play a role in pulmonary

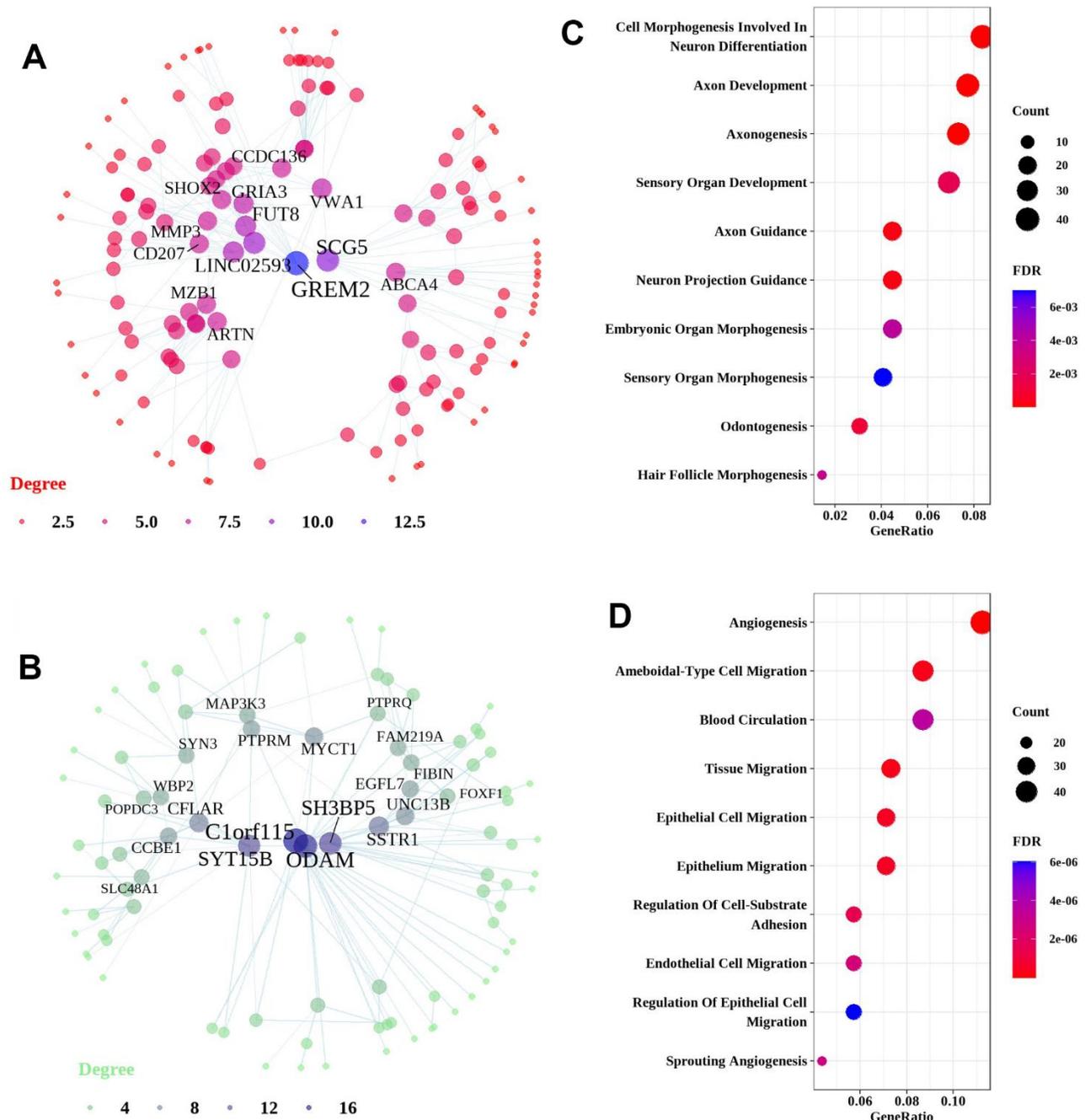


Figure 4. Network and pathway analyses of DEGs in the lung tissues of IPF patients. Network and pathways were performed on the dataset from IPF patients compared to healthy controls to identify unique dysregulated pathways in the lung tissues of IPF. (A, B) Network analysis of up-regulated genes and down-regulated genes. Most significant hub genes, according to degree and betweenness centrality, with the highest number of connections, are enclosed in ovals. (C, D) Function enrichment analysis of the up-regulated and down-regulated common DEGs, where the horizontal axis represents the proportion of DEGs under the functional terms. Top 5 pathways with most significant P-value were shown and ordered by gene ratio.

fibrosis. The differential analysis further demonstrated the up-regulation of p53 family members in IPF patients, indicating the activation of the p53 family (Supplementary Figure 6). Former evidences reported that these p53 may promote fibrosis by inducing autophagy resistance in alveolar epithelial cells [34, 35]. The involvement of p53 signaling pathway was previously reported in human lung samples of SSc-ILD and IPF, but not fHP and other CTD-ILDs [10, 36, 37],

and its activation was proven to promote lung fibrosis resolution in aged mice [38], which is also reflected in our research. Consistently, our study indicated that p53 may contribute to progressive fibrosis in three ILDs and is a therapeutic target. We displayed the pathway view of collagen degradation to illustrate a shared extracellular environment involvement across three interstitial lung diseases, underscoring a common pathological thread (Supplementary Figure 7).

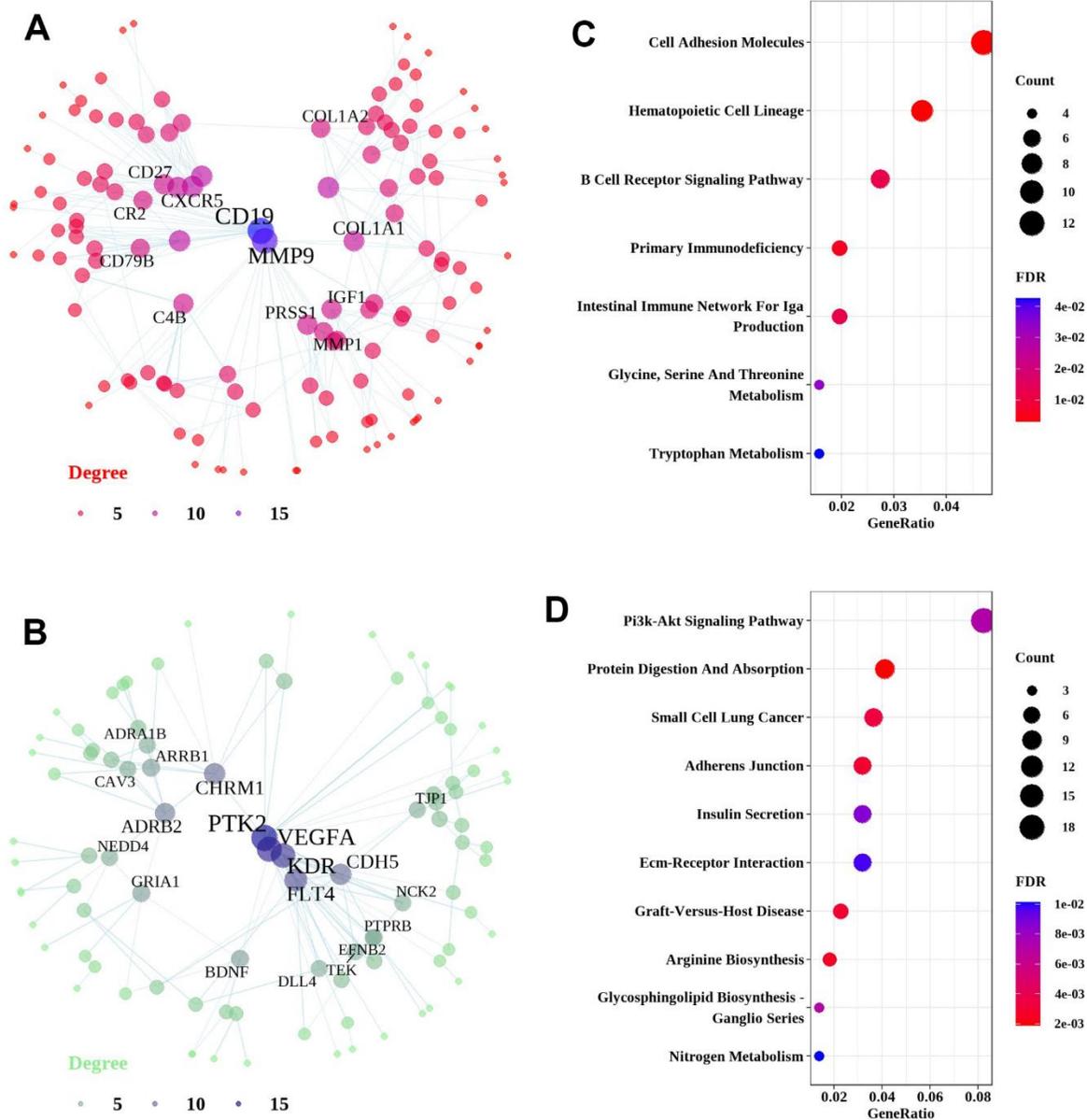


Figure 5. Network and pathway analyses of common DEGs in the lung tissues of fHP patients. Network and pathways were performed on the dataset from fHP patients compared to non-demented controls to identify unique dysregulated pathways in the lung tissues of fHP. (A, B) Network analysis of up-regulated genes and down-regulated genes. Most significant hub genes, according to degree and betweenness centrality, with the highest number of connections, are enclosed in ovals. (C, D) Function enrichment analysis of the up-regulated and down-regulated common DEGs, where the horizontal axis represents the proportion of DEGs under the functional terms. Top 5 pathways with most significant P-value were shown and ordered by gene ratio.

IPF-specific DEGs were significantly enriched in neuron-related processes like axon guidance (Supplementary Figure 8) and neuron differentiation. Neuronal guidance proteins including netrins, semaphorin, ephrins, and neurophilins are increasingly studied at the interface of

injury and repair [39]. It is reported that macrophage-derived netrin-1 (NTN1) drives the development of experimentally induced lung fibrosis via their axon-related functions involving adrenergic nerves remodeling, noradrenaline secretion, and $\alpha 1$ adrenoreceptors [40].

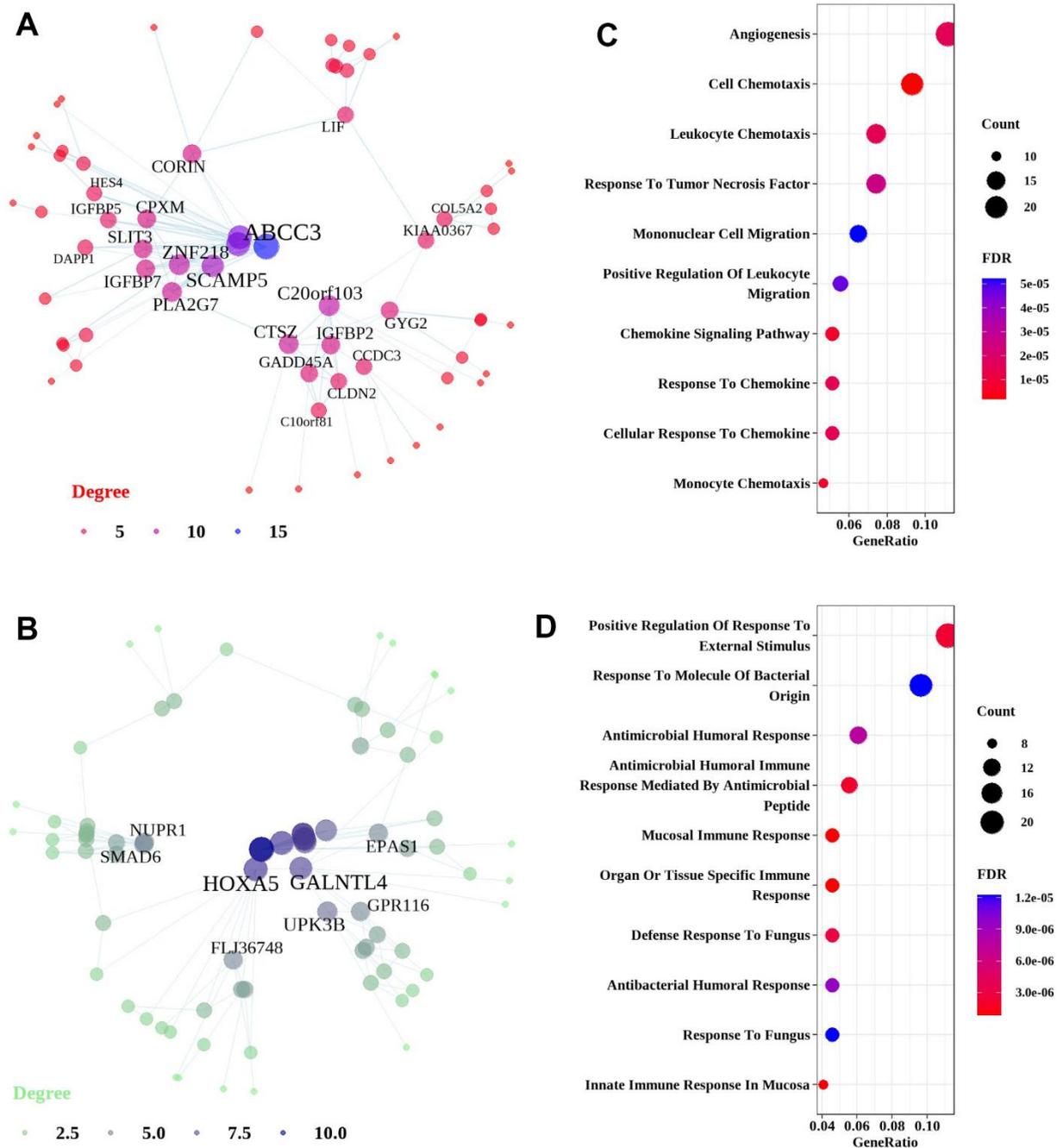


Figure 6. Network and pathway analyses of common DEGs in the lung tissues of CTD-ILD patients. Network and pathways were performed on the dataset from CTD-ILD patients compared to non-demented controls to identify unique dysregulated pathways in the lung tissues of CTD-ILD. (A, B) Network analysis of up-regulated genes and down-regulated genes. Most significant hub genes, according to degree and betweenness centrality, with the highest number of connections, are enclosed in ovals. (C, D) Function enrichment analysis of the up-regulated and down-regulated common DEGs, where the horizontal axis represents the proportion of DEGs under the functional terms. Top 5 pathways with most significant P-value were shown and ordered by gene ratio.

NTN1 deficiency or adrenergic denervation attenuated experimentally induced lung fibrosis [40, 41]. Furthermore, it is confirmed that NTN1 protein expression increased in IPF macrophages and IPF lungs are enriched for noradrenalin [40]. Use of $\alpha 1$ adrenoceptor antagonists is associated with improved survival in IPF patients [40]. Therefore, our study highlighted that neuron-related processes may play a unique important role in IPF pathogenesis and worth more investigation.

In the analysis of fHP-specific DEGs, three different methods showed B cell receptor signaling pathway (Supplementary Figure 9) as the main culprit. We found that B cells and antibody-related pathways are upregulated and may play more important roles in fHP, though a higher CD4⁺:CD8⁺ ratio and Th1 to Th2 switch are considered to be important in the pathogenesis of

fHP in former studies [42]. Consistently, prior study has reported the predominance of B cells in the transcriptome of most severely affected lung zones [43]. CD19, CXCR5, and CCR7 were identified as key regulators of fHP, all of which are all associated with B cell response. Furthermore, the active TFs in HP, such as Myb proto-oncogene protein (MYB), have been implicated in B cell differentiation under specific circumstances [44]. Specifically, increased IgA production and B cell class switching were dominant in fHP, while increased IgA levels were only reported in bronchoalveolar lavage of HP patients before [45, 46]. IgA can possibly induce lung injury by forming immune complexes [47]. Our results illuminated the role of B cells in fHP. This may also help the diagnosis of fHP, as fHP patients are frequently misdiagnosed as IPF or CTD-ILD due to their similar radiologic and/or histopathologic findings.

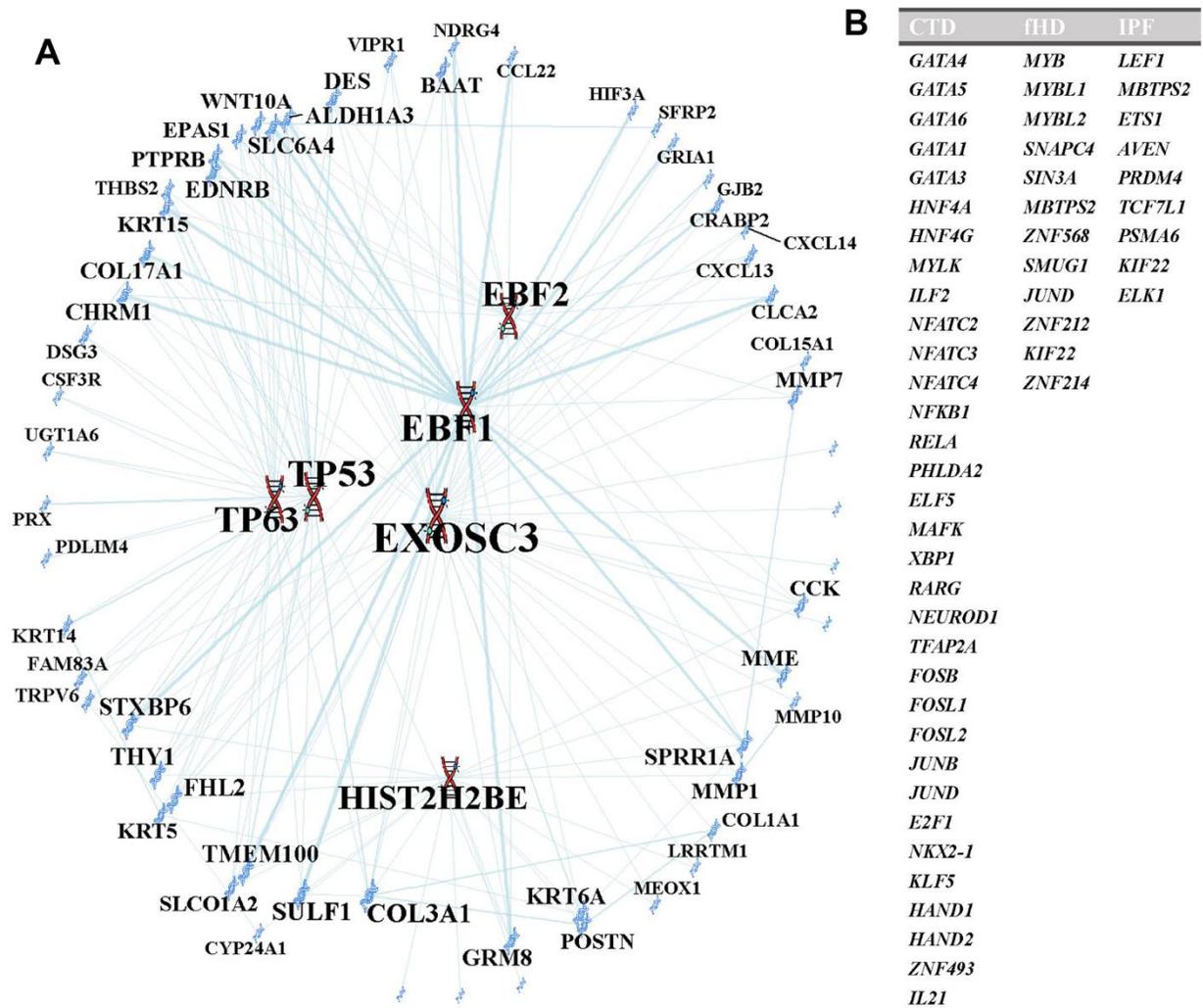


Figure 7. Transcription factors analysis. (A) TFs regulatory network. TFs were marked in red, and the DEGs were marked in blue. Meanwhile, TFs were enclosed in ovals. (B) The transcription factors in this table are unique to each pulmonary fibrosis analysis.

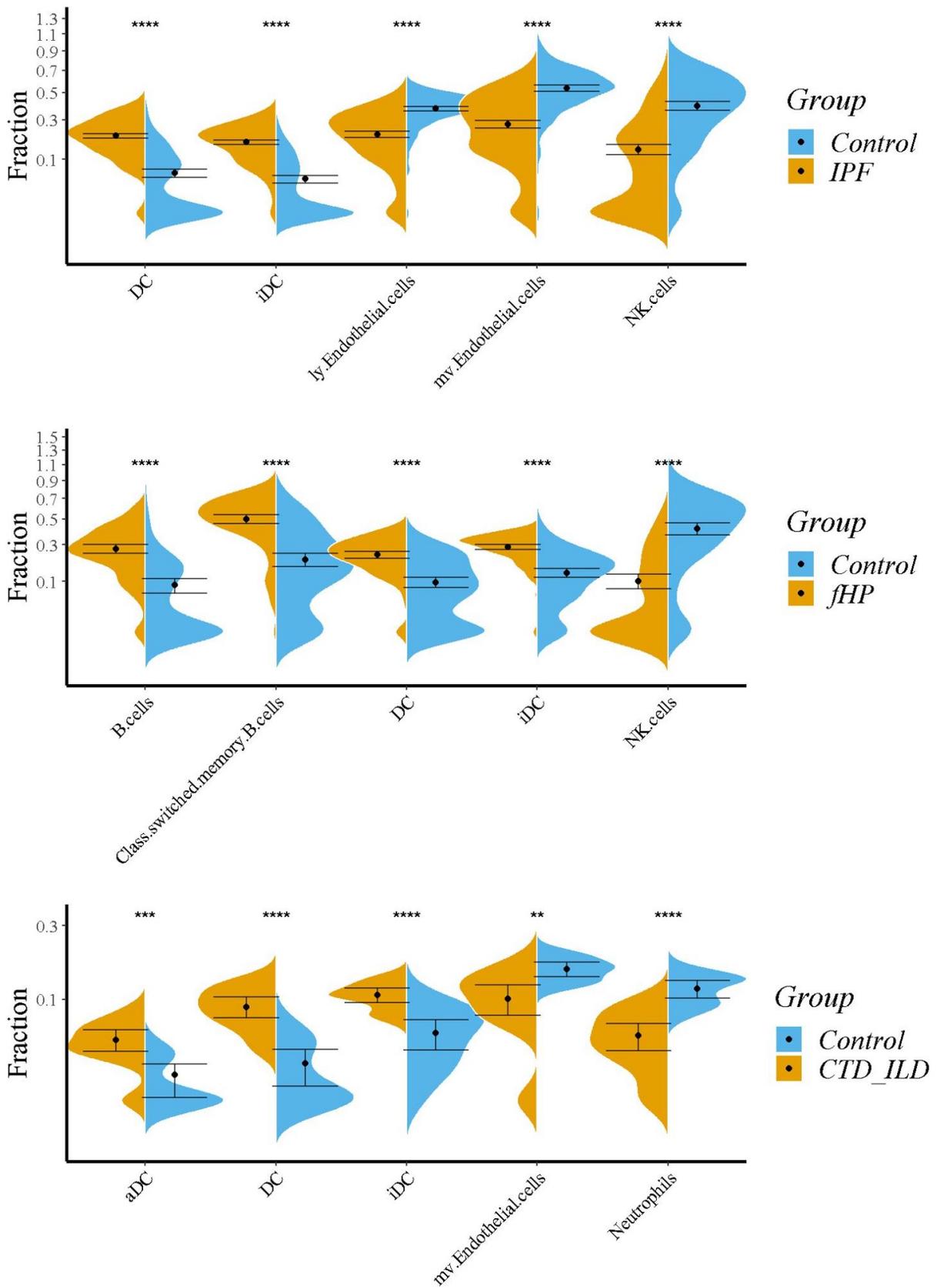


Figure 8. Immune infiltration analysis by xCell. Violin plots showing the comparison between 3 ILDs and healthy controls. Top 5 significant variations in immune cell proportions were shown. ns: $P \geq 0.05$; *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$.

We also demonstrated that activated DCs play an important role in the pathogenesis of pulmonary fibrosis associated with CTD-ILD. Enrichment analysis showed that chemokine-related pathways (Supplementary Figure 10), particularly mononuclear cell-related processes are active in CTD-ILD. DEGs including CCL19, CCL22, CCL2, and CCL18 were related to DC's function. Though total DC abundance was elevated in all three ILDs, activated DCs infiltrated the CTD-ILD lung tissue and were a possible source or target of these chemokines [48]. Previous studies primarily attributed CTD-ILD to autoantibodies and resultant dysregulation of innate immunity. However, the specific role of aDCs remains largely unexplored [49, 50]. In contrast to fHP, in CTD-ILD, DCs are activated by endogenous antigens originating from collagen fibers, neutrophil extracellular trap, or immune complexes [51]. Notably, targeting activated follicular DC and inducing DCs' tolerance has shown promise in alleviating autoimmune conditions [52, 53]. Consequently, interventions aimed at modulating aDCs or chemotaxis in the context of CTD-ILD may hold therapeutic potential for improving patient outcomes.

The strengths of this study were inclusion of most common subtypes of fibrotic ILD and integrated analysis. IPF, fHP, and CTD-ILD have similar pulmonary fibrosis but different causes, allowing for exploration of diverse fibrosis pathways. Additionally, the study analyzed data on different levels, including signaling pathways, protein interactions, immune infiltration, and TFs.

However, there were some limitations to this study. First, the number of samples was not equal among groups because lung tissue from CTD-ILD patients was relatively rare. Second, some subtypes of CTD-ILD, like idiopathic inflammatory myopathy associated-ILD and primary Sjögren syndrome-associated ILD, were missing, though two most prevalent subtypes, SSC- and RA-ILD were included.

In summary, the objective of this study was to compare distinct molecular patterns among three common types of fibrotic ILDs. Transcriptome data from public databases was collected and integrated. 1059 DEGs for IPF, 602 DEGs for fHP, and 321 DEGs for CTD-ILD were identified, and 105 DEGs were found to share among these 3 ILDs. PPI network and functional enrichment were conducted to identify functional modules within each set of DEGs. Interestingly, a shared ECM-related module among 3 ILDs was discovered and possibly regulated by p53 family members. Neuronal response seems to be involved in IPF and B cell signaling pathway was found active in

fHP, based on functional enrichment of disease-specific DEGs. Computational immune filtration was performed to predict the cellular changes in the lung tissue. The results indicated an enrichment of aDCs in CTD-ILD, while B cells seemed to dominate in fHP, which was consistent with results from functional enrichment. These molecular and cellular discrepancies may underpin the observed clinical diversity among ILDs. It is imperative to note that the pathways and biomarkers identified herein warrant further validation through experimental research and clinical trials to confirm their relevance and applicability.

AUTHOR CONTRIBUTIONS

WL and KH contributed to study design, literature search, data acquisition, data analysis, statistical analysis, manuscript preparation, editing and review. XY and PW contributed to study design, data analysis, manuscript preparation editing and review.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This work was supported by Beijing Natural Science Foundation (7222132) and National High Level Hospital Clinical Research Funding (2022-PUMCH-A-130).

REFERENCES

1. Antoniou KM, Margaritopoulos GA, Tomassetti S, Bonella F, Costabel U, Poletti V. Interstitial lung disease. *Eur Respir Rev.* 2014; 23:40–54. <https://doi.org/10.1183/09059180.00009113> PMID:24591661
2. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet.* 2017; 389:1941–52. [https://doi.org/10.1016/S0140-6736\(17\)30866-8](https://doi.org/10.1016/S0140-6736(17)30866-8) PMID:28365056
3. Chanda D, Otoupalova E, Smith SR, Volckaert T, De Langhe SP, Thannickal VJ. Developmental pathways in the pathogenesis of lung fibrosis. *Mol Aspects Med.* 2019; 65:56–69. <https://doi.org/10.1016/j.mam.2018.08.004> PMID:30130563
4. Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. *Annu Rev Pathol.* 2014; 9:157–79. <https://doi.org/10.1146/annurev-pathol-012513-104706> PMID:24050627

5. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, Swigris JJ, Taniguchi H, Wells AU. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers*. 2017; 3:17074.
<https://doi.org/10.1038/nrdp.2017.74> PMID:[29052582](https://pubmed.ncbi.nlm.nih.gov/29052582/)
6. Barnes H, Troy L, Lee CT, Sperling A, Strek M, Glaspole I. Hypersensitivity pneumonitis: Current concepts in pathogenesis, diagnosis, and treatment. *Allergy*. 2022; 77:442–53.
<https://doi.org/10.1111/all.15017> PMID:[34293188](https://pubmed.ncbi.nlm.nih.gov/34293188/)
7. Costabel U, Miyazaki Y, Pardo A, Koschel D, Bonella F, Spagnolo P, Guzman J, Ryerson CJ, Selman M. Hypersensitivity pneumonitis. *Nat Rev Dis Primers*. 2020; 6:65.
<https://doi.org/10.1038/s41572-020-0191-z> PMID:[32764620](https://pubmed.ncbi.nlm.nih.gov/32764620/)
8. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, Bang TJ, Schwarz MI, Brown KK, Kropski JA, Rojas M, Cool CD, Lee JS, et al. Chronic Hypersensitivity Pneumonitis, an Interstitial Lung Disease with Distinct Molecular Signatures. *Am J Respir Crit Care Med*. 2020; 202:1430–44.
<https://doi.org/10.1164/rccm.202001-0134OC> PMID:[32602730](https://pubmed.ncbi.nlm.nih.gov/32602730/)
9. Mathai SC, Danoff SK. Management of interstitial lung disease associated with connective tissue disease. *BMJ*. 2016; 352:h6819.
<https://doi.org/10.1136/bmj.h6819> PMID:[26912511](https://pubmed.ncbi.nlm.nih.gov/26912511/)
10. Jung SM, Park KS, Kim KJ. Integrative analysis of lung molecular signatures reveals key drivers of systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis*. 2022; 81:108–16.
<https://doi.org/10.1136/annrheumdis-2021-220493> PMID:[34380701](https://pubmed.ncbi.nlm.nih.gov/34380701/)
11. Fang D, Chen B, Lescoat A, Khanna D, Mu R. Immune cell dysregulation as a mediator of fibrosis in systemic sclerosis. *Nat Rev Rheumatol*. 2022; 18:683–93.
<https://doi.org/10.1038/s41584-022-00864-7> PMID:[36352098](https://pubmed.ncbi.nlm.nih.gov/36352098/)
12. Rosendahl AH, Schönborn K, Krieg T. Pathophysiology of systemic sclerosis (scleroderma). *Kaohsiung J Med Sci*. 2022; 38:187–95.
<https://doi.org/10.1002/kjm2.12505> PMID:[35234358](https://pubmed.ncbi.nlm.nih.gov/35234358/)
13. Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW, Kim DS, Kaminski N. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2009; 180:167–75.
<https://doi.org/10.1164/rccm.200810-1596OC> PMID:[19363140](https://pubmed.ncbi.nlm.nih.gov/19363140/)
14. Christmann RB, Sampaio-Barros P, Stifano G, Borges CL, de Carvalho CR, Kairalla R, Parra ER, Spira A, Simms R, Capellozzi VL, Lafyatis R. Association of Interferon- and transforming growth factor β -regulated genes and macrophage activation with systemic sclerosis-related progressive lung fibrosis. *Arthritis Rheumatol*. 2014; 66:714–25.
<https://doi.org/10.1002/art.38288> PMID:[24574232](https://pubmed.ncbi.nlm.nih.gov/24574232/)
15. Cecchini MJ, Hosein K, Howlett CJ, Joseph M, Mura N. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. *Respir Res*. 2018; 19:153.
<https://doi.org/10.1186/s12931-018-0857-1> PMID:[30111332](https://pubmed.ncbi.nlm.nih.gov/30111332/)
16. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, et al, and ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. 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.
<https://doi.org/10.1164/rccm.2009-040GL> PMID:[21471066](https://pubmed.ncbi.nlm.nih.gov/21471066/)
17. Morisset J, Johannson KA, Jones KD, Wolters PJ, Collard HR, Walsh SLF, Ley B, and HP Delphi Collaborators. Identification of Diagnostic Criteria for Chronic Hypersensitivity Pneumonitis: An International Modified Delphi Survey. *Am J Respir Crit Care Med*. 2018; 197:1036–44.
<https://doi.org/10.1164/rccm.201710-1986OC> PMID:[29172641](https://pubmed.ncbi.nlm.nih.gov/29172641/)
18. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, Combe B, Costenbader KH, Dougados M, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis*. 2010; 69:1580–8.
<https://doi.org/10.1136/ard.2010.138461> PMID:[20699241](https://pubmed.ncbi.nlm.nih.gov/20699241/)
19. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum*. 1980; 23:581–90.
<https://doi.org/10.1002/art.1780230510> PMID:[7378088](https://pubmed.ncbi.nlm.nih.gov/7378088/)
20. Borie R, Cardwell J, Konigsberg IR, Moore CM, Zhang W, Sasse SK, Gally F, Dobrinskikh E, Walts A, Powers J,

- Brancato J, Rojas M, Wolters PJ, et al. Colocalization of Gene Expression and DNA Methylation with Genetic Risk Variants Supports Functional Roles of *MUC5B* and *DSP* in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2022; 206:1259–70. <https://doi.org/10.1164/rccm.202110-2308OC> PMID:35816432
21. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. *Arthritis Rheum*. 2011; 63:783–94. <https://doi.org/10.1002/art.30159> PMID:21360508
22. Renaud L, da Silveira WA, Takamura N, Hardiman G, Feghali-Bostwick C. Prominence of IL6, IGF, TLR, and Bioenergetics Pathway Perturbation in Lung Tissues of Scleroderma Patients With Pulmonary Fibrosis. *Front Immunol*. 2020; 11:383. <https://doi.org/10.3389/fimmu.2020.00383> PMID:32210969
23. Wang S, Liu M, Li X, Zhang J, Wang F, Zhang C, Roden A, Ryu JH, Warrington KJ, Sun J, Matteson EL, Tschumperlin DJ, Vassallo R. Canonical and noncanonical regulatory roles for JAK2 in the pathogenesis of rheumatoid arthritis-associated interstitial lung disease and idiopathic pulmonary fibrosis. *FASEB J*. 2022; 36:e22336. <https://doi.org/10.1096/fj.202101436R> PMID:35522243
24. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28:882–3. <https://doi.org/10.1093/bioinformatics/bts034> PMID:22257669
25. 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. <https://doi.org/10.1093/nar/gkv007> PMID:25605792
26. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–7. <https://doi.org/10.1089/omi.2011.0118> PMID:22455463
27. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021; 49:D605–12. <https://doi.org/10.1093/nar/gkaa1074> PMID:33237311
28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504. <https://doi.org/10.1101/gr.1239303> PMID:14597658
29. Herrmann C, Van de Sande B, Potier D, Aerts S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res*. 2012; 40:e114. <https://doi.org/10.1093/nar/gks543> PMID:22718975
30. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017; 18:220. <https://doi.org/10.1186/s13059-017-1349-1> PMID:29141660
31. Sturm G, Finotello F, List M. Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. *Methods Mol Biol*. 2020; 2120:223–32. https://doi.org/10.1007/978-1-0716-0327-7_16 PMID:32124323
32. Renzoni EA, Poletti V, Mackintosh JA. Disease pathology in fibrotic interstitial lung disease: is it all about usual interstitial pneumonia? *Lancet*. 2021; 398:1437–49. [https://doi.org/10.1016/S0140-6736\(21\)01961-9](https://doi.org/10.1016/S0140-6736(21)01961-9) PMID:34499865
33. Kunej T. Integrative Map of *HIF1A* Regulatory Elements and Variations. *Genes (Basel)*. 2021; 12:1526. <https://doi.org/10.3390/genes12101526> PMID:34680921
34. Kobayashi Y, Tata A, Konkimalla A, Katsura H, Lee RF, Ou J, Banovich NE, Kropski JA, Tata PR. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat Cell Biol*. 2020; 22:934–46. <https://doi.org/10.1038/s41556-020-0542-8> PMID:32661339
35. Chilosi M, Poletti V, Murer B, Lestani M, Cancellieri A, Montagna L, Piccoli P, Cangi G, Semenzato G, Doglioni C. Abnormal re-epithelialization and lung remodeling in idiopathic pulmonary fibrosis: the role of deltaN-p63. *Lab Invest*. 2002; 82:1335–45. <https://doi.org/10.1097/01.lab.0000032380.82232.67> PMID:12379768
36. Sun J, Guo Y, Chen T, Jin T, Ma L, Ai L, Guo J, Niu Z, Yang R, Wang Q, Yu X, Gao H, Zhang Y, et al. Systematic

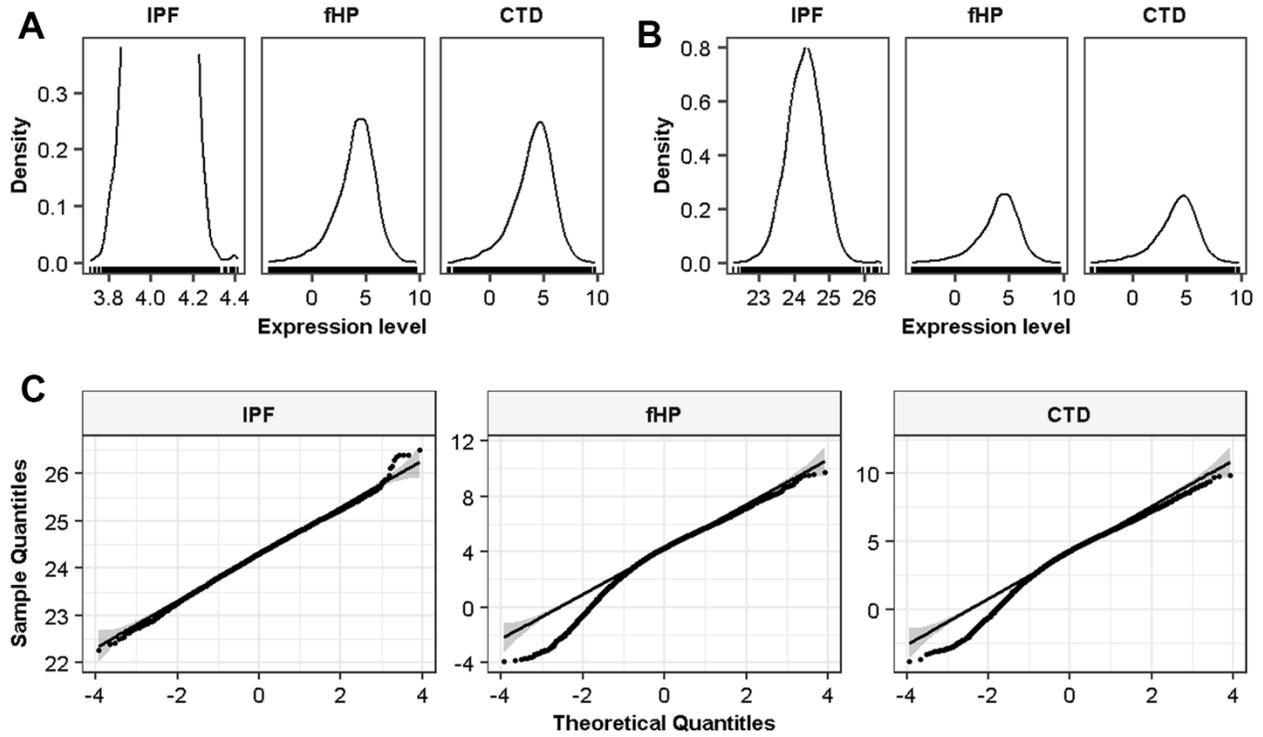
- analyses identify the anti-fibrotic role of lncRNA TP53TG1 in IPF. *Cell Death Dis.* 2022; 13:525.
<https://doi.org/10.1038/s41419-022-04975-7>
PMID:35661695
37. Yao C, Guan X, Carraro G, Parimon T, Liu X, Huang G, Mulay A, Soukiasian HJ, David G, Weigt SS, Belperio JA, Chen P, Jiang D, et al. Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med.* 2021; 203:707–17.
<https://doi.org/10.1164/rccm.202004-1274OC>
PMID:32991815
38. Qu J, Yang SZ, Zhu Y, Guo T, Thannickal VJ, Zhou Y. Targeting mechanosensitive MDM4 promotes lung fibrosis resolution in aged mice. *J Exp Med.* 2021; 218:e20202033.
<https://doi.org/10.1084/jem.20202033>
PMID:33688918
39. Mirakaj V, Rosenberger P. Immunomodulatory Functions of Neuronal Guidance Proteins. *Trends Immunol.* 2017; 38:444–56.
<https://doi.org/10.1016/j.it.2017.03.007>
PMID:28438491
40. Gao R, Peng X, Perry C, Sun H, Ntokou A, Ryu C, Gomez JL, Reeves BC, Walia A, Kaminski N, Neumark N, Ishikawa G, Black KE, et al. Macrophage-derived netrin-1 drives adrenergic nerve-associated lung fibrosis. *J Clin Invest.* 2021; 131:e136542.
<https://doi.org/10.1172/JCI136542>
PMID:33393489
41. Sun H, Zhu Y, Pan H, Chen X, Balestrini JL, Lam TT, Kanyo JE, Eichmann A, Gulati M, Fares WH, Bai H, Feghali-Bostwick CA, Gan Y, et al. Netrin-1 Regulates Fibrocyte Accumulation in the Decellularized Fibrotic Sclerodermatous Lung Microenvironment and in Bleomycin-Induced Pulmonary Fibrosis. *Arthritis Rheumatol.* 2016; 68:1251–61.
<https://doi.org/10.1002/art.39575>
PMID:26749424
42. Barrera L, Mendoza F, Zuñiga J, Estrada A, Zamora AC, Melendro EI, Ramírez R, Pardo A, Selman M. Functional diversity of T-cell subpopulations in subacute and chronic hypersensitivity pneumonitis. *Am J Respir Crit Care Med.* 2008; 177:44–55.
<https://doi.org/10.1164/rccm.200701-093OC>
PMID:17947613
43. De Sadeleer LJ, McDonough JE, Schupp JC, Yan X, Vanstapel A, Van Herck A, Everaerts S, Geudens V, Sacreas A, Goos T, Aelbrecht C, Nawrot TS, Martens DS, et al. Lung Microenvironments and Disease Progression in Fibrotic Hypersensitivity Pneumonitis. *Am J Respir Crit Care Med.* 2022; 205:60–74.
<https://doi.org/10.1164/rccm.202103-0569OC>
PMID:34724391
44. Xiao C, Calado DP, Galler G, Thai TH, Patterson HC, Wang J, Rajewsky N, Bender TP, Rajewsky K. MiR-150 Controls B Cell Differentiation by Targeting the Transcription Factor c-Myb. *Cell.* 2016; 165:1027.
<https://doi.org/10.1016/j.cell.2016.04.056>
PMID:27153500
45. Yoshizawa Y, Ohtsuka M, Noguchi K, Uchida Y, Suko M, Hasegawa S. Hypersensitivity pneumonitis induced by toluene diisocyanate: sequelae of continuous exposure. *Ann Intern Med.* 1989; 110:31–4.
<https://doi.org/10.7326/0003-4819-110-1-31>
PMID:2535687
46. Sato T, Kitajima A, Ohmoto S, Chikuma M, Kado M, Nagai S, Izumi T, Takeyama M, Masada M. Determination of human immunoglobulin A and secretory immunoglobulin A in bronchoalveolar lavage fluids by solid phase enzyme immunoassay. *Clin Chim Acta.* 1993; 220:145–56.
[https://doi.org/10.1016/0009-8981\(93\)90043-4](https://doi.org/10.1016/0009-8981(93)90043-4)
PMID:8111959
47. Johnson KJ, Wilson BS, Till GO, Ward PA. Acute lung injury in rat caused by immunoglobulin A immune complexes. *J Clin Invest.* 1984; 74:358–69.
<https://doi.org/10.1172/JCI111431> PMID:6430958
48. Mokuda S, Miyazaki T, Ubara Y, Kanno M, Sugiyama E, Takasugi K, Masumoto J. CD1a+ survivin+ dendritic cell infiltration in dermal lesions of systemic sclerosis. *Arthritis Res Ther.* 2015; 17:275.
<https://doi.org/10.1186/s13075-015-0785-0>
PMID:26419626
49. Distler O, Assassi S, Cottin V, Cutolo M, Danoff SK, Denton CP, Distler JHW, Hoffmann-Vold AM, Johnson SR, Müller-Ladner U, Smith V, Volkman ER, Maher TM. Predictors of progression in systemic sclerosis patients with interstitial lung disease. *Eur Respir J.* 2020; 55:1902026.
<https://doi.org/10.1183/13993003.02026-2019>
PMID:32079645
50. Akiyama M, Kaneko Y. Pathogenesis, clinical features, and treatment strategy for rheumatoid arthritis-associated interstitial lung disease. *Autoimmun Rev.* 2022; 21:103056.
<https://doi.org/10.1016/j.autrev.2022.103056>
PMID:35121155
51. Dieker J, Tel J, Pieterse E, Thielen A, Rother N, Bakker M, Fransen J, Dijkman HBPM, Berden JH, de Vries JM, Hilbrands LB, van der Vlag J. Circulating Apoptotic Microparticles in Systemic Lupus Erythematosus Patients Drive the Activation of Dendritic Cell Subsets and Prime Neutrophils for NETosis. *Arthritis Rheumatol.* 2016; 68:462–72.
<https://doi.org/10.1002/art.39417>
PMID:26360137

52. Das A, Heesters BA, Bialas A, O'Flynn J, Rifkin IR, Ochando J, Mittereder N, Carlesso G, Herbst R, Carroll MC. Follicular Dendritic Cell Activation by TLR Ligands Promotes Autoreactive B Cell Responses. *Immunity*. 2017; 46:106–19.
<https://doi.org/10.1016/j.immuni.2016.12.014>
PMID:[28099860](https://pubmed.ncbi.nlm.nih.gov/28099860/)

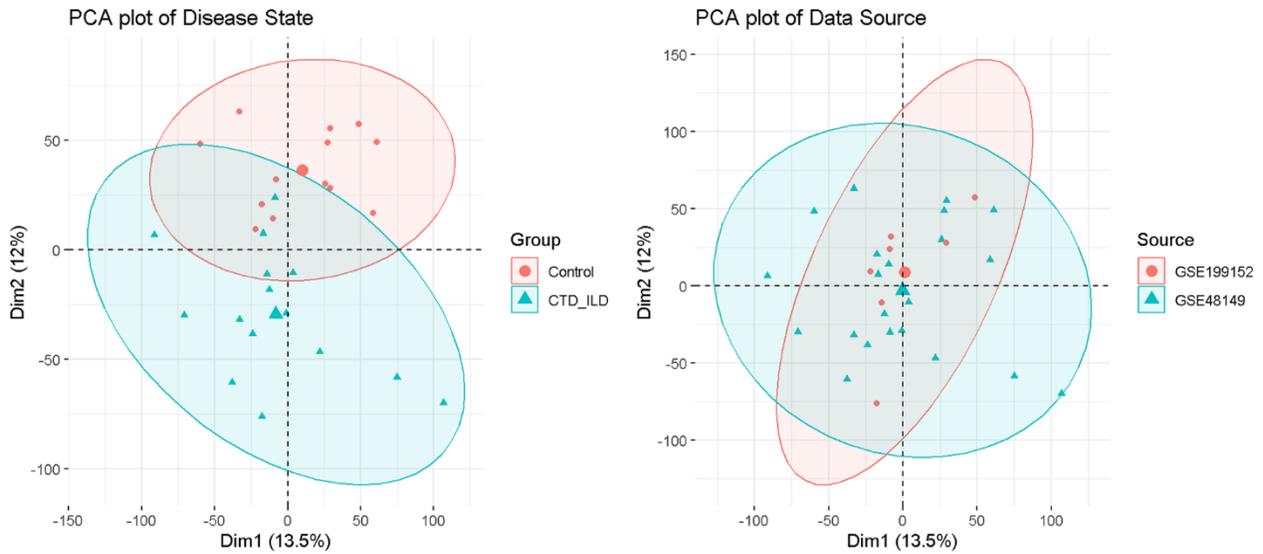
53. Cifuentes-Rius A, Desai A, Yuen D, Johnston APR, Voelcker NH. Inducing immune tolerance with dendritic cell-targeting nanomedicines. *Nat Nanotechnol*. 2021; 16:37–46.
<https://doi.org/10.1038/s41565-020-00810-2>
PMID:[33349685](https://pubmed.ncbi.nlm.nih.gov/33349685/)

SUPPLEMENTARY MATERIALS

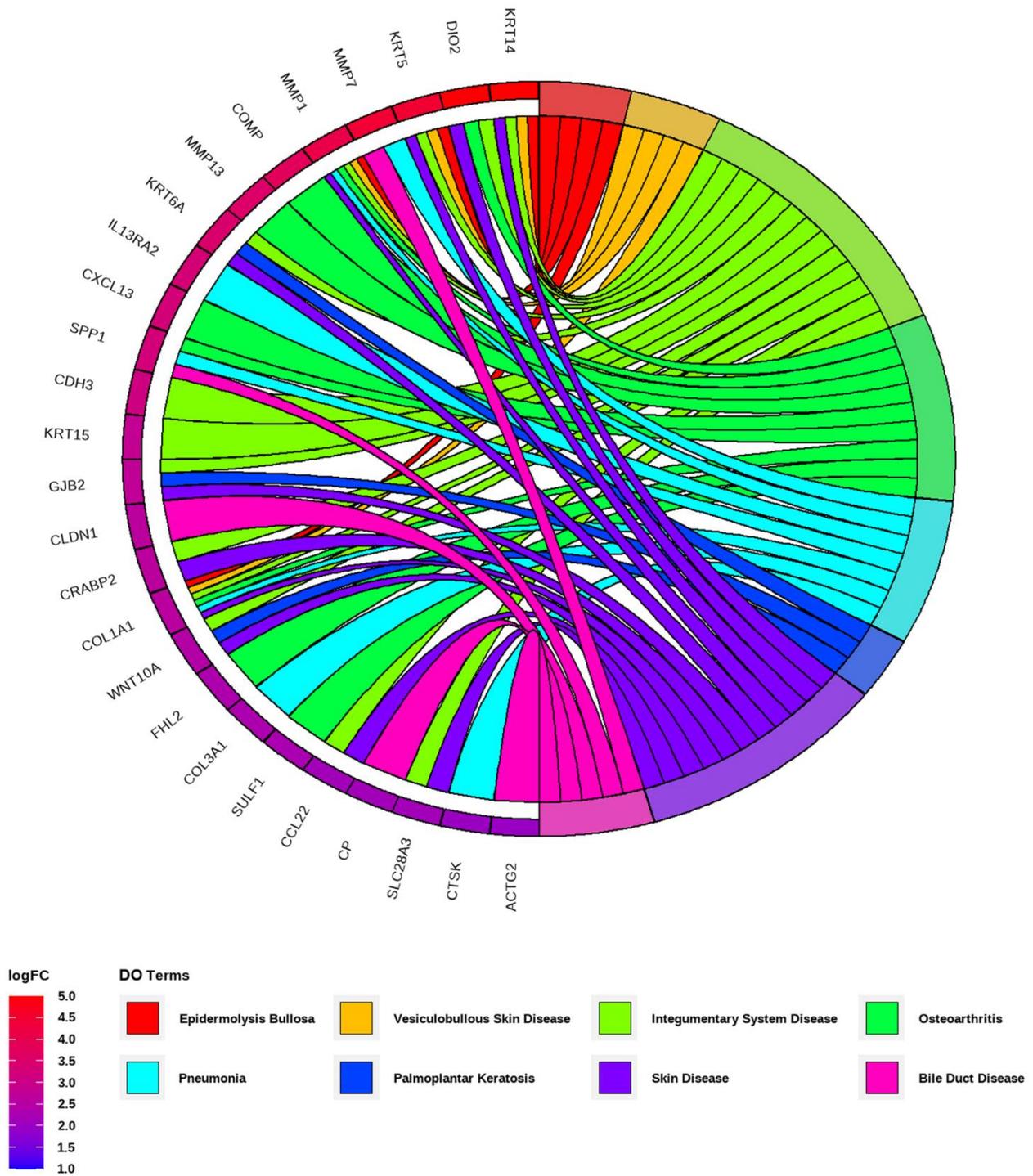
Supplementary Figures



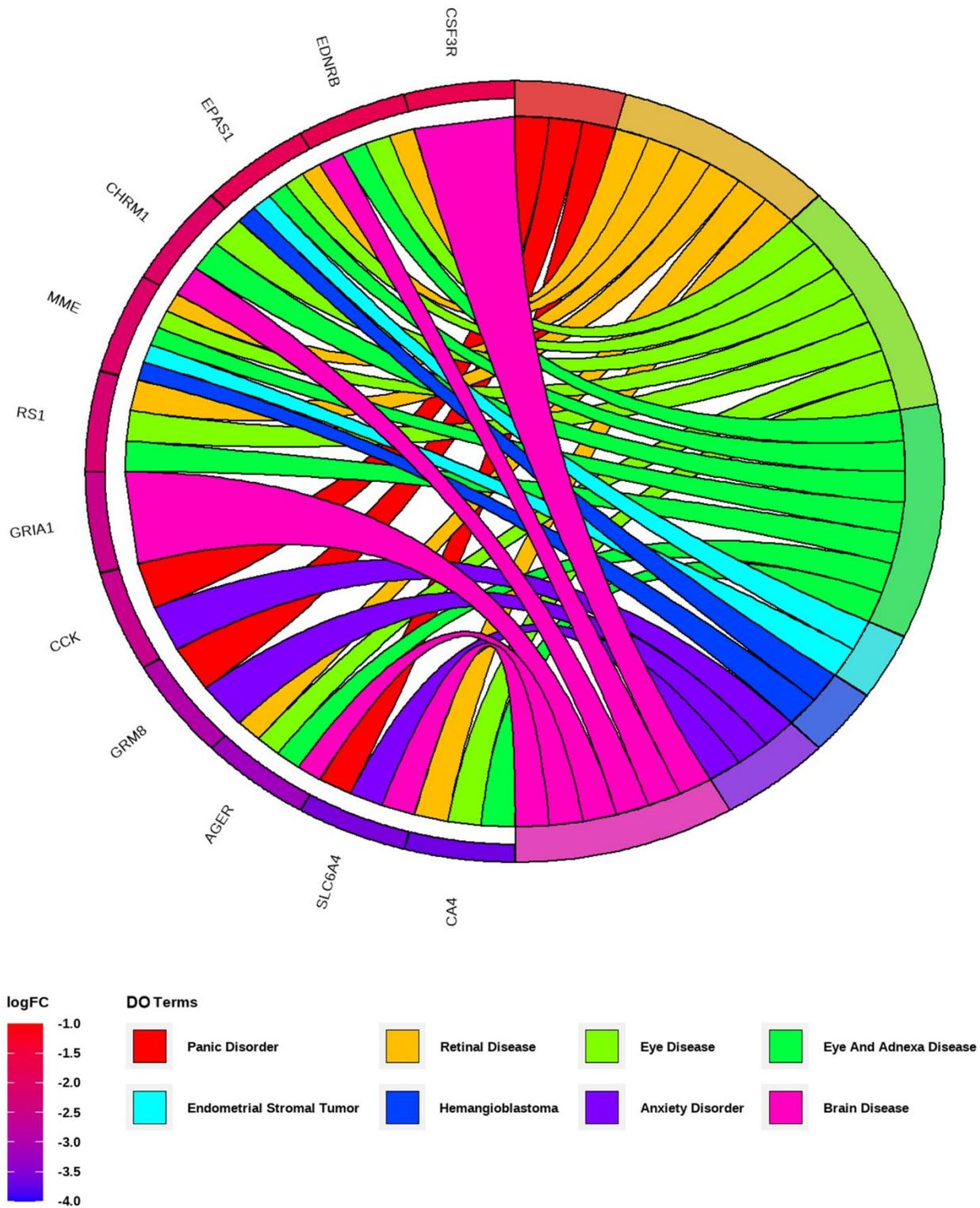
Supplementary Figure 1. The correction of datasets used in this study. (A, B) The density plots of IPF, fHP and CTD-ILD datasets before (A) and after (B) normalization and removing batch-effects. (C) The Q-Q plot of the three datasets after correction.



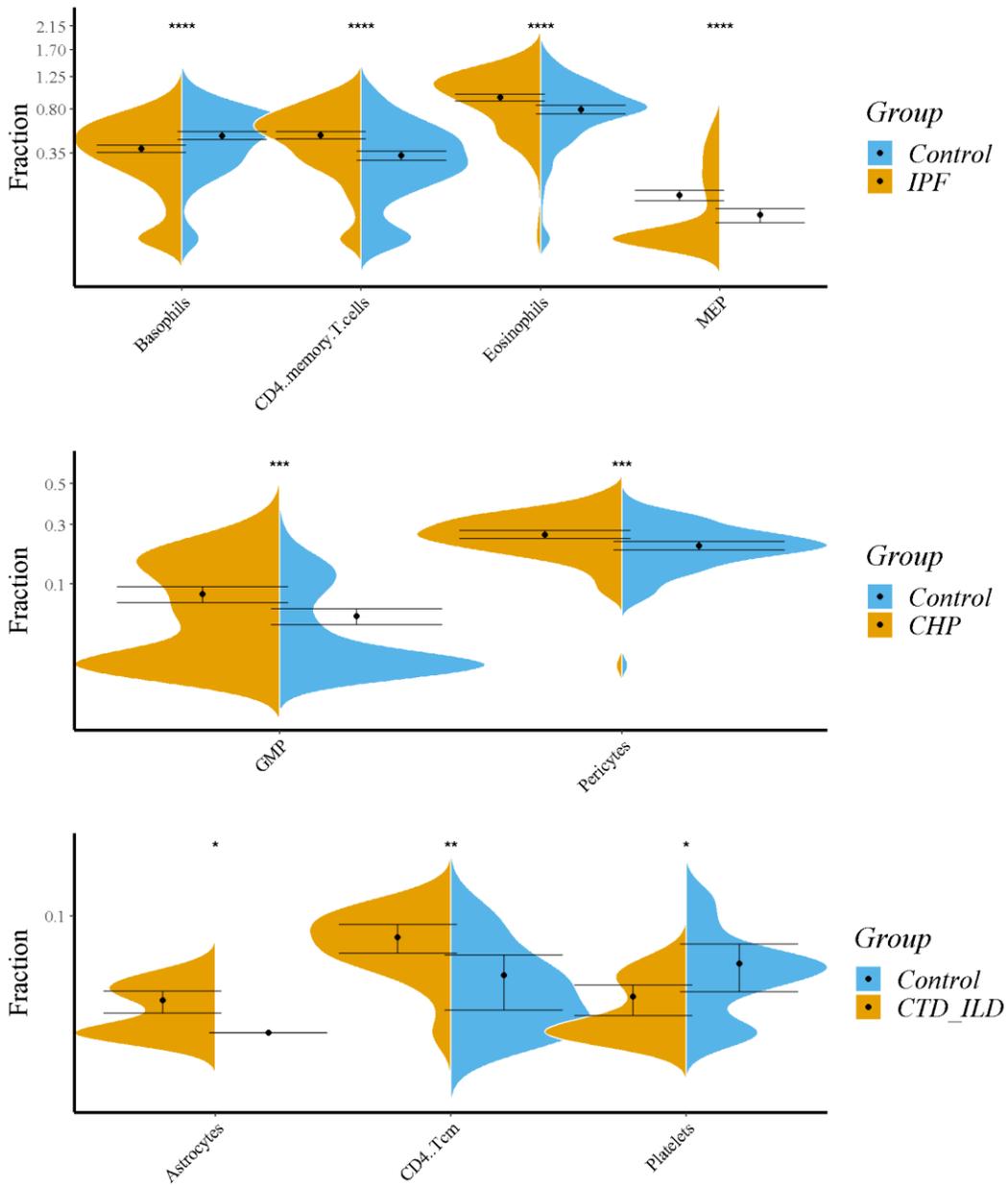
Supplementary Figure 2. The principal component analysis (PCA) plot of the merged dataset by disease state and the original datasets.



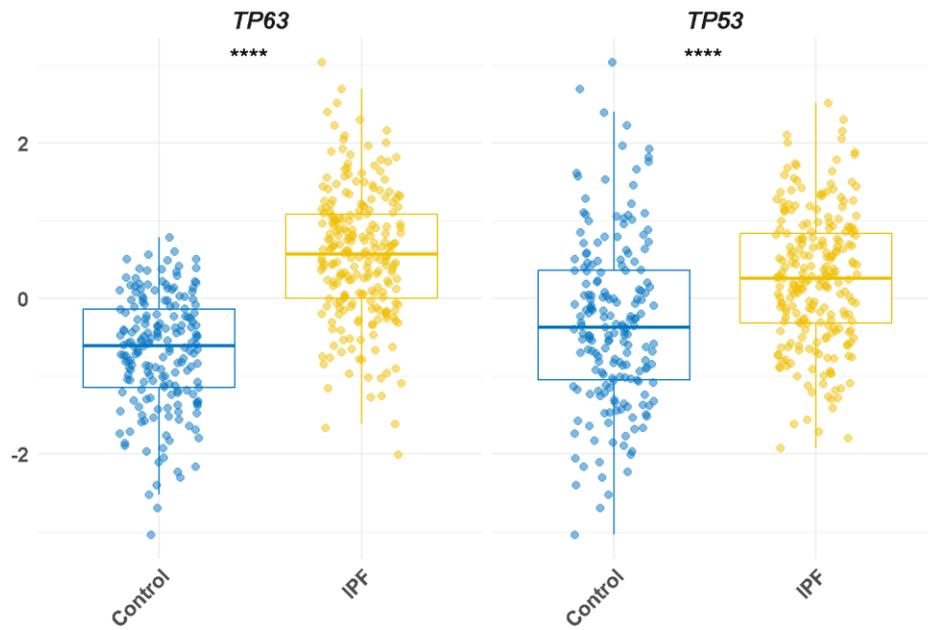
Supplementary Figure 3. DO enrichment analysis. Chord diagram showed the correlation between diseases and common down-regulated DEGs, with different colors corresponding to different DO terms.



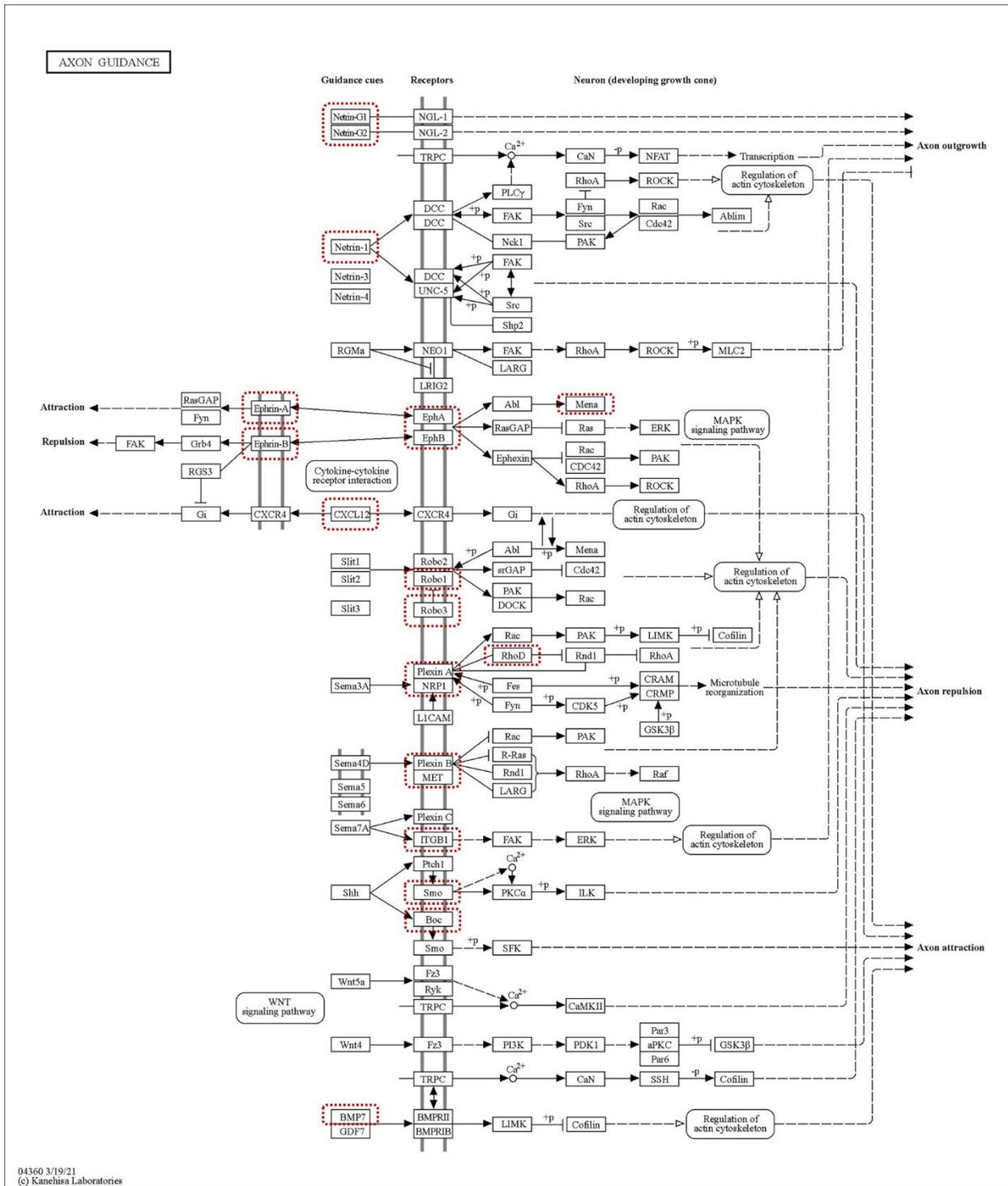
Supplementary Figure 4. Immune infiltration analysis. Violin diagram showing three diseases' unique proportion variation of immune cells obtained using the xCell. *:P < 0.05; **:P < 0.01; ***:P < 0.001; ****:P < 0.0001



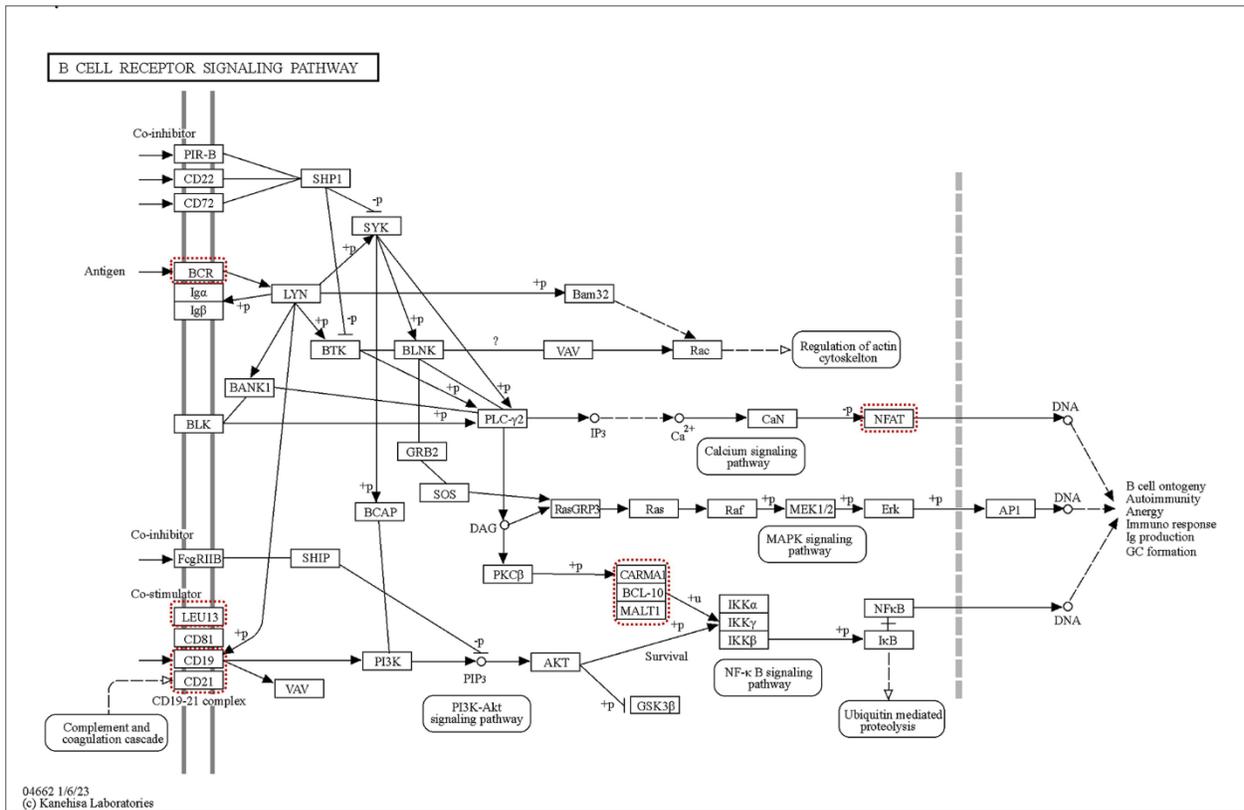
Supplementary Figure 5. Activation of the p53 family in IPF patients. Boxplot showing the up-regulation of p53 family members in patients with idiopathic lung fibrosis. ****: $P < 0.0001$.



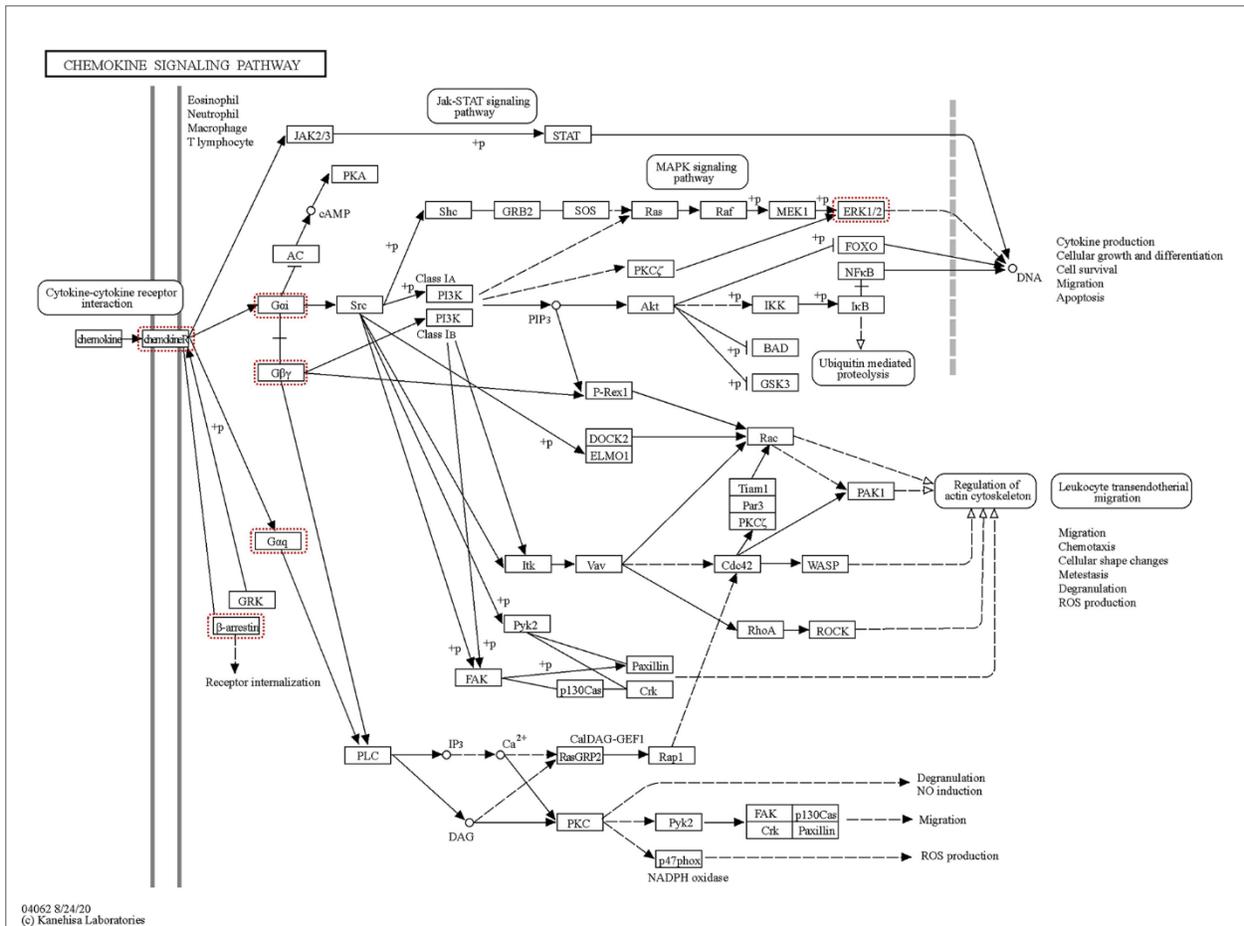
Supplementary Figure 6. Activation of the p53 family in IPF patients. Boxplot showing the up-regulation of p53 family members in patients with idiopathic lung fibrosis. ****: $P < 0.0001$.



Supplementary Figure 8. Pathway view of axon guidance. The IPF-specific DEGs were emphasized with red box.



Supplementary Figure 9. Pathway view of B cell receptor signaling. The CHP-specific DEGs were emphasized with red box.



Supplementary Figure 10. Pathway view of chemokine signaling. The CTD_ILD-specific DEGs were emphasized with red box.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–4, 7–10.

Supplementary Table 1. Differentially expressed genes of fHP.

Supplementary Table 2. Differentially expressed genes of IPF.

Supplementary Table 3. Differentially expressed genes of CTD-ILD.

Supplementary Table 4. All common differentially expressed genes of IPF, fHP and CTD-ILD.

Supplementary Table 5. Up-regulated common differentially expressed genes of IPF, fHP and CTD-ILD.

hgnc_symbol	logFC_IPF	logFC_CHP	logFC_CTD.ILD
ACTG2	1.16867332289099	1.06884319181345	1.57264761554731
ALDH1A3	1.37938443626019	1.203935667529	2.02540445237465
BAAT	1.50486346899843	2.11337420730513	2.15145213982184
CCL22	1.21217326638706	1.32230959545926	1.46059285766171
CD1A	1.28563576066025	1.19629725112951	1.30184435915926
CD207	1.28503231545761	1.32747024212248	1.4483832918451
CD24	1.3567283865497	1.30857017633235	1.35159383429003
CD79A	1.1001129187577	1.60528405537751	1.31164920456388
CDH3	1.55254377511596	1.77422180393457	1.77715247941456
CILP2	1.43081105473093	1.63602244229272	1.08019620698807
CLCA2	1.1889021961	1.35272328038441	1.90153630382498
CLDN1	1.35303152912343	1.26599260803912	2.03637389720905
COL10A1	1.41114064628794	2.15572235346214	2.02614774212009
COL15A1	1.32217306566972	1.16719339312545	2.03491115864745
COL17A1	1.56579578040413	3.11378506173711	2.09905827748447
COL1A1	1.19967078320226	1.36940141859273	2.09054631335168
COL22A1	1.42887166919087	1.08783036913002	1.07928914726629
COL3A1	1.16138217513617	1.19792189825317	1.96960198466899
COMP	1.44021388998733	2.18515988472482	2.29253797286178
CP	1.25006262921802	1.59861653042657	1.15130529617468
CR2	1.12132181659738	4.76726574883564	1.20100749441066
CRABP2	1.46726880643017	1.68475783603254	1.40603848092354
CST1	1.14702917580896	4.67484799147212	2.41315816141039
CST2	1.32344449403933	2.43414705816646	2.09366865964664
CTHRC1	1.40382311258675	1.40711633966439	1.98177353779443
CTSK	1.30885109655958	1.20945785400863	1.27868864164933
CXCL13	1.28943496908849	2.5439019027623	1.65581870981868
CXCL14	1.44220942925456	1.81851882996128	1.46659837028222
CXCL6	1.14205998213226	1.42926109616135	1.51310015479465
CYP24A1	1.38183581653317	2.91870164219946	1.45646381062104
DES	1.15760796003788	1.05027649574864	1.18699183415991
DIO2	1.50085088162593	4.84026100872179	1.7024286569945
DSG3	1.38275286113405	1.92086551148894	1.35386259338207
FAM83A	1.32648290606239	1.00390407052197	1.36762517730964
FHL2	1.45987116429258	1.04584658659112	1.80808977685406
FNDC1	1.44169186308349	2.3074685642745	1.39798967421432
GJB2	1.38939930334731	1.90585410149485	1.44066489248103
HS6ST2	1.53810380412916	1.29069179835172	1.04641986611265
IGFL1	1.4173567987978	4.10096599631326	1.30994741132254
IGFL2	1.50505031248118	4.10850318037433	2.20639209105057
IL13RA2	1.50103761252354	1.65577949471032	2.25078732597388
KRT14	1.48071175889854	4.79832833957068	1.95558847504554
KRT15	1.30519460842457	1.82896466456427	1.68717749254765
KRT16	1.31863430545208	4.60334339417622	1.15553978505549
KRT17	1.45526856223259	2.38141883794919	3.08328692257748
KRT5	1.345186402039	2.36519549370348	3.13204675919996
KRT6A	1.27725887623621	2.34136032196097	2.06608901185009

KRT6B	1.17825127028074	2.79711661913147	1.70233098181525
LRRTM1	1.34460832671252	2.86863997930985	1.42821470996562
MDK	1.28116659256428	1.14996875744481	1.49032602470256
MEOX1	1.2194108341199	2.47679192252652	1.317949319174
MMP1	1.41926330419899	2.36944621973944	2.52411096843166
MMP10	1.25697335482946	1.85519717382516	1.92540471014862
MMP11	1.49021125416502	1.87211172610846	1.88099243180746
MMP13	1.48302296385838	3.44614767062312	1.2528582962855
MMP7	1.51474262024814	2.29674819012599	2.82787276324862
PCP4	1.39770370221466	3.49108552294338	1.91730096143447
PDLIM4	1.42251447811563	1.49837642836773	1.23591530117167
POSTN	1.34102947292959	1.42996786093157	1.71850107261227
POU2AF1	1.2747842380671	1.26539652745947	1.21263194662377
SCG5	1.49102062977927	2.95977608871547	1.4585972748726
SERPINB5	1.36125280521898	2.42067874661796	1.9944708717494
SERPIND1	1.24819123091871	1.53196417422222	1.96175863645607
SFRP2	1.38283433251574	1.55744901541122	2.37094794321419
SLC28A3	1.28073117579086	1.02551876584639	1.74312735750169
SLN	1.20043899312754	1.62301222630393	1.09394416395194
SPP1	1.20157691284556	1.49184373727737	2.97845347776939
SPRR1A	1.337886445773	4.26712471227044	1.13264020888067
SULF1	1.43755354979499	1.07367922897018	1.70103084623035
THBS2	1.18709534646714	1.06521212072047	1.68326877480194
THY1	1.3345079226186	1.86940548944793	2.74004298678487
TRPV6	1.35762436742283	1.15224496082757	1.20221419882409
TUBB3	1.36740853052801	2.53274536700379	1.30448508092816
UCN2	1.33588447389693	1.31905940884364	1.13027254748341
UGT1A6	1.32118085405306	2.82876121457458	1.58714934524663
WNT10A	1.37898853272126	1.45352136155049	1.52635758438604

Supplementary Table 6. Down-regulated common differentially expressed genes of IPF, fHP and CTD-ILD.

hgnc symbol	logFC IPF	logFC CHP	logFC CTD.ILD
AGER	-1.34899997867565	-1.42199951903428	-2.64609928429044
BTNL9	-1.42003716481868	-1.62515713982911	-2.34544821942771
CA4	-1.4501328267098	-1.5884186896536	-2.90308555733666
CCK	-1.120331223529	-2.62561468464483	-1.10951595760675
CHRM1	-1.16493901069145	-1.45895209169062	-1.19574462535598
CRTAC1	-1.22207832130913	-1.1324549012924	-1.66030534025204
CSF3R	-1.23097381642726	-1.15494893358054	-1.0504694882291
EDNRB	-1.30050312761451	-1.1119168035004	-1.07601469183884
EPAS1	-1.28412788703569	-1.26627830723914	-1.02230942444717
FCN3	-1.26412115067964	-1.24178649339451	-2.57438309243446
GALNT13	-1.56703692435974	-1.93826251238563	-1.23444085378394
GRIA1	-1.33586041118294	-1.89998969983665	-1.23841567821961
GRM8	-1.12127198304424	-2.89794990829132	-1.32327719649941
HIF3A	-1.3482116537219	-1.36676416826069	-1.14455075292685
HTR3C	-1.16608921452496	-1.35575762501819	-1.42230123302719
ITLN2	-1.46698772149713	-3.02657947692901	-3.38646612110895
KLRF1	-1.22424721414267	-1.64203657219513	-1.04733916449086
MME	-1.32923029628891	-1.33764545145942	-1.17963698890775
NDRG4	-1.22391023699594	-1.41714821190263	-1.31730113529117
PNMT	-1.1658308453927	-1.0634866396352	-1.53618040595259
PRX	-1.35922208820321	-1.27536241010644	-1.60625736257521
PTPRB	-1.36952057540505	-1.02877211145758	-1.20802965804873
RGS9BP	-1.2561647656702	-1.09809486301492	-1.29897202906918
RS1	-1.33465448139793	-1.05537623233956	-1.72487566322996
SLC6A4	-1.17726120301447	-2.32365869095908	-2.339236094182
SLCO1A2	-1.3625353577266	-3.28361942036096	-1.11892598862302
STXBP6	-1.33747791880127	-1.52101017786342	-1.1840624542589
TMEM100	-1.18507102935744	-1.29489581247938	-2.3646363303669
VIPR1	-1.23025374592456	-1.70707144428038	-2.30090054973427

Supplementary Table 7. Gene ontology enrichment analysis of up-regulated common differentially expressed genes.

Supplementary Table 8. Gene ontology enrichment analysis of down-regulated common differentially expressed genes.

Supplementary Table 9. KEGG enrichment analysis of up-regulated common differentially expressed genes.

Supplementary Table 10. KEGG enrichment analysis of down-regulated common differentially expressed genes.