STM JOURNALS

Research & Reviews: A Journal of
Bioinformatics

ISSN: 2393-8722
Volume 13, Issue 1, 2026
January–April
DOI (Journal): 10.37591/RRJoBI

https://journals.stmjournals.com/rrjobi

Research RRJoBI

Identification of Hub Genes and Enriched Gene Ontology & Pathways in Idiopathic Pulmonary Fibrosis Through Bioinformatics Approaches

Elakkiya M.*

Abstract

Idiopathic Pulmonary Fibrosis (IPF) is a progressive interstitial lung disease marked by aberrant remodeling of lung tissue and excessive extracellular matrix deposition, ultimately leading to respiratory failure. Despite ongoing research, the molecular mechanisms underlying IPF remain incompletely understood. This research aims to uncover differentially expressed genes (DEGs) and related biological pathways through an integrated analysis of microarray data. Two publicly available datasets, GSE110147 and GSE53845, were obtained from the Gene Expression Omnibus (GEO) database. DEGs were determined using the GEO2R tool with selection criteria of log2 fold change > 1 and adjusted p-value < 0.05. Functional annotation and pathway enrichment analyses were conducted via the STRING database to explore key biological processes associated with IPF. Additionally, protein–protein interaction (PPI) networks were generated using STRING to identify potential hub genes. The analysis revealed consistent dysregulation in genes related to extracellular matrix organization, immune response, and fibrosis-associated pathways. Integration of both datasets enhanced the robustness and reliability of DEG identification. The results provide insight into the molecular framework of IPF and suggest potential biomarkers and therapeutic targets. Further experimental validation is warranted to confirm the role of identified genes and pathways, contributing to improved diagnosis and treatment strategies for IPF.

Keywords: Bioinformatics, differentially expressed genes, hub genes, idiopathic pulmonary fibrosis, protein–protein interaction, STRING

INTRODUCTION

Idiopathic Pulmonary Fibrosis (IPF) is a long-term, progressive, and ultimately deadly interstitial lung disorder characterized by excessive scarring and fibrosis of the lung tissue. It involves injury to lung cells, accumulation of fibroblasts and myofibroblasts, and abnormal repair processes, resulting in lung stiffening and impaired function. Normal lung tissue is replaced by an altered extracellular matrix, leading to the destruction of alveolar structures, reduced lung elasticity, impaired gas exchange, and, eventually, respiratory failure and death [1]. The disease predominantly affects individuals over the age of 50 and is associated with poor prognosis. The 5-year survival rate of IPF patients ranges between 20% and 40%, and the median survival is 2 to 3 years, which is worse than many cancers [2]. One of the biggest challenges in treating IPF is that clinicians do not fully understand what causes the disease or how exactly it progresses. After many drugs fail in clinical trials, two therapeutic agents, pirfenidone, and nintedanib, are confirmed to aid in slowing the progression of IPF and thus reducing the dysfunction of the lung [3, 4]. Although treatments like antifibrotic drugs are available, they only show slow disease progression and do not offer a cure. Gaining insight into the molecular

***Author for Correspondence**
Elakkiya M.
E-mail: Elakkiyam3103@gmail.com

Researcher, Department of Biotechnology, M S Ramaiah Institute of Technology, Bengaluru, Karnataka, India

Received Date: November 11, 2025
Accepted Date: January 31, 2026
Published Date: March 27, 2026

Citation: Elakkiya M. Identification of Hub Genes and Enriched Gene Ontology & Pathways in Idiopathic Pulmonary Fibrosis Through Bioinformatics Approaches. Research & Reviews: A Journal of Bioinformatics. 2026; 13(1): 1–13p.

mechanisms underlying IPF is crucial for advancing the development of more accurate diagnostic biomarkers and targeted treatment strategies. High-throughput technologies, such as microarray analysis, have emerged as powerful tools to examine genome-wide expression changes in diseased tissues compared to healthy controls.

In recent years, several studies have begun to use bioinformatics technology for identifying biomarkers that can help in detecting the disease earlier, prognosis, and evaluate the responses to treatment for IPF from genome sequence database of patients [4, 5]. Publicly available databases, such as the Gene Expression Omnibus (GEO), provide access to comprehensive transcriptomic datasets that allow for in-depth exploration of gene expression profiles in various disease conditions, including IPF [5, 6]. Techniques like microarray profiling help researchers to identify differentially expressed genes due to dysregulation of genes during IPF and these dysregulated genes can be upregulated genes where genes are highly expressed or downregulated genes where genes are lowly expressed [7]. Gene Ontology (GO) and KEGG pathway enrichment analysis can further interpret the significance of differentially expressed genes involved in disease progression [6, 8]. Identification of hub genes that may serve as potential regulators or therapeutic targets in IPF can be visualized by constructing complex networks like protein–protein interaction (PPI) networks using STRING.

Overall, understanding the genetic and molecular changes involved in Idiopathic Pulmonary Fibrosis is essential for improving diagnosis and treatment. With the help of bioinformatics tools and publicly available gene expression data, researchers can identify key genes and pathways that contribute to the disease. These insights may lead to the discovery of useful biomarkers and potential drug targets, which could support earlier detection and more effective, personalized treatment options for IPF in the future [9].

METHODS AND MATERIALS

Data Retrieval

We retrieved two gene expression datasets from the NCBI Gene Expression Omnibus (GEO) database. These datasets were chosen because they provide comprehensive transcriptomic profiles related to idiopathic pulmonary fibrosis (IPF) disease. Gene expression profile GSE110147 were downloaded from the Gene Expression Omnibus database platform was GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array. This microarray obtained samples from recipients' organs of 22 patients with IPF, 10 with NSIP, and 5 with mixed IPF-NSIP undergoing lung transplantation [5]. GSE53845 includes lung tissue samples from 40 patients diagnosed with IPF and 8 healthy individuals and the data were generated using the platform GPL6480, Agilent-014850 Whole Human Genome Microarray 4x44K G4112F. As with GSE110147, we retrieved only 22 IPF patients and 11 normal controls data for this study as our focus is only on IPF disease [10].

Identification of DEGs

Differentially expressed genes (DEGs) between idiopathic pulmonary fibrosis (IPF) and healthy lung tissues, were identified by using GEO2R, an interactive web-based tool that performs R-based statistical analysis on GEO datasets. GEO2R was applied separately to the GSE53845 and GSE110147 datasets. Genes with an absolute log2 fold change (log2FC) greater than 0 and an adjusted p-value of less than 0.05 were considered significantly differentially expressed. Specifically, genes with $\log2FC > 0$ were classified as upregulated, while those with $\log2FC < 0$ were considered downregulated. The top 1% of both upregulated and downregulated were taken for further analysis [11].

Gene Ontology (GO) Enrichment Analysis

Gene Ontology (GO) analysis was performed using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) tool to understand the roles of DEGs in three GO domains: Molecular Function (MF) which is an enrichment analysis that identifies the molecular activities of gene products, such as binding (e.g., cytokine binding, protein kinase activity) and enzymatic activity. Cellular Component analysis reveals the subcellular localization of gene products, such as extracellular matrix,

cytoplasm, plasma membrane, or nucleus, reflecting the structural and compartmental organization affected in IPF. Biological Process (BP) which is an enrichment in biological processes, such as inflammatory response, extracellular matrix organization, and cell proliferation, is assessed to understand the broader biological impact of DEGs [12–15].

Pathway Enrichment Analysis

To explore the involvement of DEGs in known signaling and metabolic pathways, Reactome or Wiki pathway enrichment analysis was conducted through the STRING platform. Enriched pathways are selected based on p-value < 0.05.

Protein–Protein Interaction (PPI) Network Construction

Information of protein–protein interaction (PPI) of the DEGs was evaluated by STRING (Search Tool for the Retrieval of Interacting Genes) online tool [6], with interactions of a combined score 0.4 considered statistically significant [7].

Identification of Hub Genes

To explore the biological classifications of hub genes, functional, and pathway enrichment analyses were performed by [8] STRING. The DEGs were subjected to protein–protein interaction (PPI) network construction using the STRING database (https://string-db.org/) with a confidence score threshold of ≥0.4 (medium confidence) and the organism set to Homo sapiens. The interaction map among the encoded proteins and provided key network parameters including number of nodes, edges, average node degree, and clustering coefficient can be visualized [16–19].

RESULTS

Identification of DEGs

In the dataset GSE110147, 23513 DEGs met the criteria of adjusted p-value < 0.05 out of which 9920 genes were upregulated, and 13593 genes were downregulated. In the dataset GSE53845, 8517 DEGs met the criteria of adjusted p-value < 0.05 out of which 4610 genes were upregulated, and 5907 genes were downregulated. In Volcano map, DEGs between idiopathic pulmonary fibrosis (IPF) and the control groups are represented where red dots represent upregulated DEGs, the blue dots represent downregulated DEGs and gray dots represent undifferentiated genes. Differentially expressed genes (DEGs) of GSE110147 and GSE53845 dataset are shown in Figures 1 and 2, respectively. The top 1% of both upregulated and downregulated genes from GSE110147 (234 genes) and GSE53845 (85 genes), a total of 319 genes were taken for further analysis.

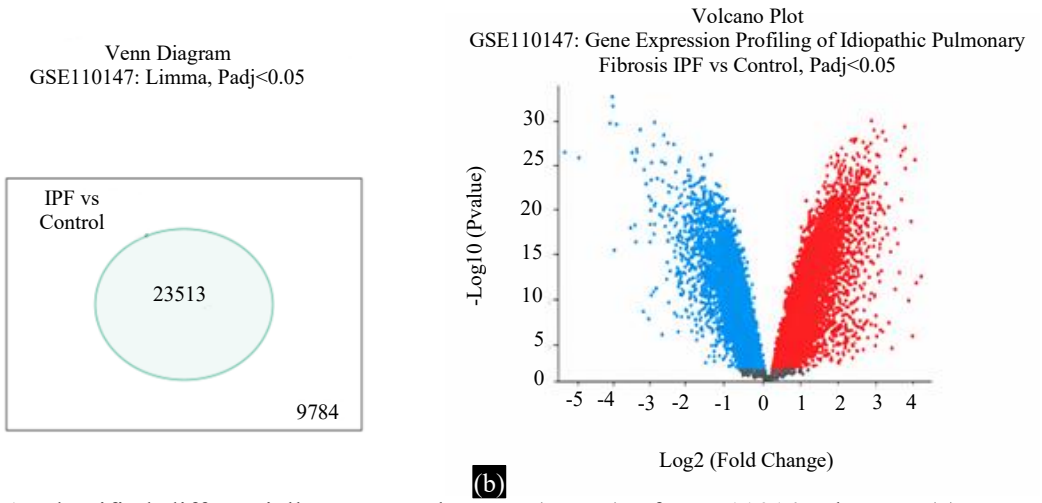


Figure 1. Identified differentially expressed genes (DEGs) of GSE110147 dataset. (a) presents the number of DEGs that met the criteria of p-adj < 0.05 in the form Venn diagram. (b) presents all the upregulated and downregulated expressed genes in the form of a volcano plot.

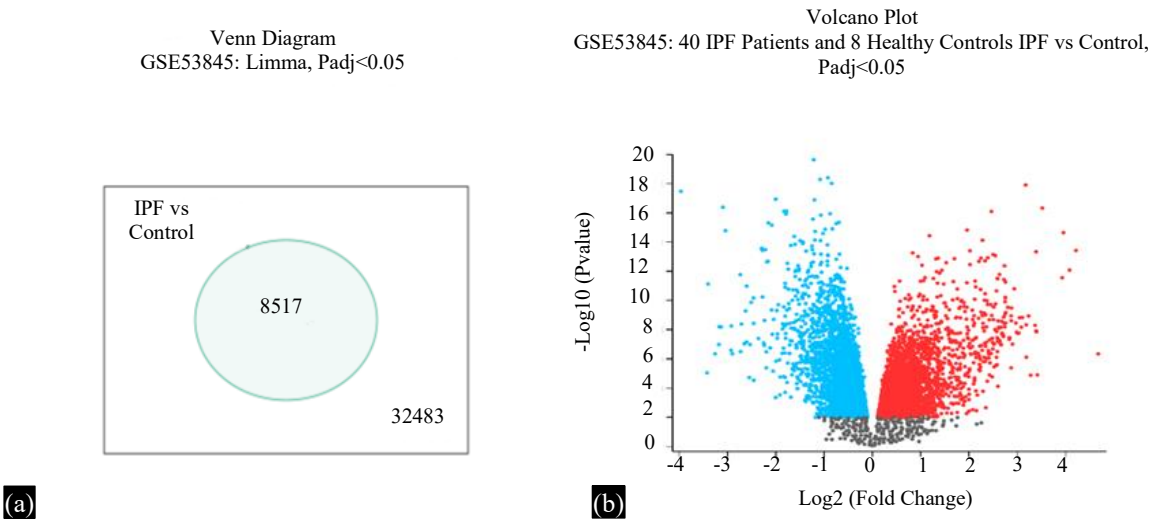


Figure 2. Identified differentially expressed genes (DEGs) of GSE53845 dataset. (a) presents the number of DEGs that met the criteria of p-adj < 0.05 in the form Venn diagram. (b) presents all the upregulated and downregulated expressed genes in the form of a volcano plot.

Gene Ontology (GO) and Pathway Enrichment Analysis

The top 1% of both upregulated and downregulated genes from GSE110147 (234 genes) and GSE53845 (85 genes), a total of 319 genes were Inputted in STRING for enrichment analysis for GO analysis and pathway analysis. The top significant terms of BP, CC, and MF are displayed in Table 1 and top 5 pathway terms are displayed in Table 2. The gene ontology enrichment is shown in Figure 3 and the representation of pathway analysis is shown in Figure 4.

Table 1. Gene ontology analysis of differentially expressed genes (DEGs).

GO Category	Go term ID	Term description	FDR	Matching proteins in network
Biological process	GO:0006959	Humoral immune response	4.20E-05	CCL2, CCL21, C6, FCN3, CXCL10, CCL19, PLA2G1B, DEFA3, IGLL1, CXCL14, BPIFA1, CFH, XCL1, S100A12, POU2AF1, IGLL5, JCHAIN
	GO:0032103	Positive regulation of response to external stimulus	3.96E-05	CPB2, LGALS2, IL1RL1, CCR7, FABP4, CCL21, FCN3, PDCD4, LY96, CXCL10, CCL19, FBN1, HSP90AA1, CXCL14, ARG1, XCL1, S100A12, PGC, AGER, EDN1, DDT, ALOX5AP
	GO:0051707	Response to another organism	7.84E-06	CCL2, CCR7, BPIFB1, FABP4, CCL21, C6, FCN3, PDCD4, GPM6A, LY96, TIMP4, CXCL10, CCL19, PLA2G1B, DEFA3, GKN2, IGLL1, HSP90AA1, CXCL14, CAV1, IFI6, LYAR, BPIFA1, ARG1, CTR9, EPRS1, CFH, IVNS1ABP, XCL1, SLAMF7, S100A12, UBD, EDNRB, EDN1, POU2AF1, IGLL5, JCHAIN, FKBP5, DIO2, TMF1, ABCC8, CCDC186
	GO:0009617	Response to bacterium	9.46E-05	CCL2, CCR7, FABP4, PDCD4, GPM6A, LY96, TIMP4, CXCL10, PLA2G1B, DEFA3, GKN2, IGLL1, CAV1, BPIFA1, ARG1, CTR9, S100A12, EDNRB, EDN1, IGLL5, JCHAIN, FKBP5, DIO2, TMF1, ABCC8, CCDC186
	GO:0009605	Response to external stimulus	5.20E-06	PDK4, CYP24A1, RP1, CCL2, COL1A1, CCR7, KRT5, BPIFB1, FABP4, CHL1, CCL21, SLC6A4, C6, FGF7, FCN3, SFRP2, PDCD4, GPM6A, LY96, TIMP4, AGL, CXCL10, CCL19, VPS41, PLA2G1B, FBN1, DEFA3, GKN2, IGLL1, ALDH1A3, HSP90AA1, CXCL14, CAV1, IFI6, LYAR, BPIFA1, ARG1, CTR9, EPRS1, CFH, IVNS1ABP, XCL1, SLAMF7, S100A12, UBD, EDNRB, EDN1, POSTN,

				POU2AF1, SPP1, FAM107A, IGLL5, NUCB2, JCHAIN, SESN3, FKBP5, DIO2, TTN, TMF1, ABCC8, CCDC186
	GO:0044419	Biological process involved in interspecies interaction between organisms	3.43E-05	CCL2, SELPLG, CCR7, BPIFB1, FABP4, CCL21, C6, FCN3, PDCD4, GPM6A, LY96, TIMP4, CXCL10, CCL19, PLA2G1B, DEFA3, GKN2, IGLL1, HSP90AA1, CXCL14, CAV1, IFI6, LYAR, BPIFA1, ARG1, CTR9, EPRS1, CFH, IVNS1ABP, XCL1, SLAMF7, S100A12, UBD, EDNRB, EDN1, POU2AF1, IGLL5, JCHAIN, FKBP5, DIO2, TMF1, ABCC8, CCDC186
	GO:0032101	Regulation of response to external stimulus	9.58E-05	CPB2, LGALS2, CCL2, IL1RL1, CCR7, FABP4, CCL21, BIRC3, FCN3, PDCD4, LY96, CXCL10, CCL19, FBN1, HSP90AA1, CXCL14, CAV1, LYAR, ARG1, MYOZ1, CFH, XCL1, S100A12, PGC, AGER, EDNRB, EDN1, SPP1, DDT, NUCB2, ALOX5AP, ABCC8
	GO:0070887	Cellular response to chemical stimulus	7.29E-06	PDK4, CPB2, CYP24A1, COMP, CYP3A5, CCL2, COL1A1, SELPLG, IL1RL1, CCR7, DLL4, RAMP2, FMO5, FABP4, CCL21, SLC6A4, BIRC3, EPAS1, DNAJC10, FGF7, SFRP2, PDCD4, LY96, COL1A2, COL3A1, CXCL10, CCL19, PLA2G1B, SPOCK2, FBN1, EIF4A2, MT1E, DEFA3, IL1R2, MT1F, CXCL14, CYP3A7, CAV1, USP9Y, ARG1, CTR9, EPRS1, XCL1, S100A12, IL13RA2, ATRX, AGER,ID1, EDNRB, EDN1, POSTN, GPX3, SPP1, LRRC32, DERL3, FAM107A, SESN3,BDKRB2, TMEM100, ALOX5AP, TMF1,HBB, ABCC8, CCDC186
	GO:0071560	Cellular response to transforming growth factor beta stimulus	0.0019	COL1A1, COL1A2, COL3A1, FBN1, CAV1, USP9Y, ARG1, XCL1, ID1, EDN1, LRRC32
	GO:0042310	Vasoconstriction	0.0038	COMP, SLC6A4, CAV1, EDNRB, EDN1, BDKRB2
Molecular function	GO:0030020	Extracellular matrix structural constituent conferring tensile strength	0.0089	COL1A1,COL1A2,COL14A1,COL3A1,COL10A1,COL17A1
	GO:0051959	Dynein light intermediate chain binding	0.045	DNAH5, DNAH12, CCDC88A, DNAH6, DYNC2H1
	GO:0005515	Protein binding	0.0474	RP1, COMP, CCL2, COL1A1, SELPLG, IL1RL1, CCR7, GNG11, DLL4, KRT5, FABP4, CHL1, CCL21, SLC6A4, EPAS1, DNAJC10, TOP2B, CP, DNAH5, FGF7, FCN3, SFRP2, LY96, TIMP4, AGL, HHIP, COL1A2, ANGPTL4, LMAN2, COL3A1, CXCL10, ADRA1B, GCC2, CCL19, VPS41, PLA2G1B, DNAH12, STXBP6, FBN1, DEFA3, IGLL1, BTNL9, CTHRC1, IL1R2, ALDH1A3, HSP90AA1, GLDN, CXCL14, ZBTB16, CCDC88A, CAV1, SAXO2, USP9Y, LYAR, ARG1, FXR1, CD163, MYOZ1, EMP2, DSC3, KIF21A, CTR9, DZIP3, EPRS1, CFH, STX11, XCL1, SLAMF7, S100A3, S100A12, ABCD3, IL13RA2, ATRX, AGER, ID1, EDNRB, HIF3A, RAD50, ATL2, EDN1, POSTN, GPX3, DNAH6, SPP1, KTN1, DDT, USP47, LRRC32, DERL3, CEBPD, NECAB1,

				TARS1, ICAM2, FAM107A, PROM1, IGLL5, NUCB2, JCHAIN, FKBP5, CEP290, BDKRB2, DIO2, TTN, SBNO1, ALOX5AP, TMF1, HBB, ABCC8, DYNC2H1, CCDC186, S100A2
	GO:0008009	Chemokine activity	0.0474	CCL2, CCL21, CXCL10, CCL19, CXCL14, XCL1
Cellular component	GO:0005576	Extracellular region	4.43E-08	KRT14, CPB2, ESF1, COMP, CCL2, COL1A1, IL1RL1, RRAS, KRT5, BPIFB1, FABP4, PEBP4, CHL1, CCL21, PI15, MMP7, C6, CP, DNAH5, FGF7, FCN3, SFRP2, GPM6A, LY96, TIMP4, AGL, HHIP, COL1A2, COL14A1, ARMC3, CA4, ANGPTL4, LMAN2, COL3A1, CXCL10, CCL19, PLA2G1B, PRSS8, SPOCK2, MMP1, FBN1, COL10A1, DEFA3, GKN2, IGLL1, CTHRC1, IL1R2, ALDH1A3, HSP90AA1, GLDN, CXCL14, BPIFA1, ARG1, CFAP43, SUSD2, CD163, DSC3, CFH, XCL1, ITLN2, S100A12, CRTAC1, RBP4, IL13RA2, ADIRF, PGC, CD52, AGER, ASPN, LGALS7, EDN1, POSTN, GPX3, SPP1, DDT, LRRC32, GDPD3, TARS1, ATP6V1F, PROM1, IGLL5, CFAP54, RPS13, NUCB2, JCHAIN, FKBP5, CEP290, TTN, HBB, DYNC2H1, COL17A1
	GO:0005615	Extracellular space	4.43E-08	KRT14, CPB2, ESF1, COMP, CCL2, COL1A1, RRAS, KRT5, BPIFB1, FABP4, PEBP4, CHL1, CCL21, PI15, MMP7, C6, CP, FGF7, FCN3, SFRP2, GPM6A, LY96, TIMP4, COL1A2, COL14A1, ARMC3, CA4, ANGPTL4, LMAN2, COL3A1, CXCL10, CCL19, PLA2G1B, PRSS8, FBN1, COL10A1, DEFA3, GKN2, IGLL1, CTHRC1, ALDH1A3, HSP90AA1, GLDN, CXCL14, BPIFA1, ARG1, SUSD2, CFH, XCL1, ITLN2, S100A12, CRTAC1, RBP4, IL13RA2, ADIRF, PGC, ASPN, LGALS7, EDN1, POSTN, GPX3, SPP1, DDT, LRRC32, GDPD3, TARS1, ATP6V1F, PROM1, IGLL5, RPS13, NUCB2, JCHAIN, FKBP5, TTN, HBB, DYNC2H1, COL17A1
	GO:0005581	Collagen trimer	0.0016	COL1A1, FCN3, COL1A2, COL14A1, COL3A1, COL10A1, CTHRC1, GLDN, COL17A1
	GO:0031012	Extracellular matrix	0.0033	COMP, COL1A1, MMP7, FCN3, SFRP2, TIMP4, COL1A2, COL14A1, ANGPTL4, COL3A1, SPOCK2, MMP1, FBN1, COL10A1, CTHRC1, HSP90AA1, ASPN, POSTN, LRRC32, COL17A1
	GO:0098644	Complex of collagen trimers	0.0266	COL1A1, COL1A2, COL3A1, COL10A1

Table 2. Pathway analysis of differentially expressed genes (DEGs).

GO Category	Term ID with description	FDR	Matching proteins in network
Reactome Pathway	HSA-2173782– Binding and Uptake of Ligands by Scavenger Receptors	0.0029	COL1A1, COL1A2, COL3A1, HSP90AA1, CD163, JCHAIN, HBB
	HSA-1442490– Collagen degradation	0.0029	COL1A1, MMP7, COL1A2, COL14A1, COL3A1, MMP1, COL10A1, COL17A1
	HSA-2022090– Assembly of collagen fibrils and other multimeric structures	0.0053	COL1A1, MMP7, COL1A2, COL14A1, COL3A1, COL10A1, COL17A1
	HSA-216083– Integrin cell surface interactions	0.0053	COMP, COL1A1, COL1A2, COL3A1, FBN1, COL10A1, SPP1, ICAM2
	HSA-8948216– Collagen chain trimerization	0.007	COL1A1, COL1A2, COL14A1, COL3A1, COL10A1, COL17A1

Wiki Pathway	WP5115– Network map of SARS-CoV-2 signaling pathway	0.0083	CCL2, RRAS, CCL21, CXCL10, VPS41, EIF4A2, IGLL1, IL1R2, IFI6, CD163, CFH, HBB
	WP5055– Burn wound healing	0.0107	CCL2, COL1A1, SFRP2, LY96, COL1A2, MMP1, FBN1, CXCL14
	WP2882– Nuclear receptors meta-pathway	0.0192	PDK4, CYP3A5, CCL2, SLC6A4, BIRC3, ANGPTL4, HSP90AA1, CYP3A7, AGER, GPX3, KTN1, FKBP5, ALOX5AP
	WP5095– Overview of proinflammatory and profibrotic mediators	0.0248	CCL2, CCL21, CXCL10, CCL19, MMP1, CXCL14, XCL1, SPP1
	WP2880– Glucocorticoid receptor pathway	0.0312	CCL2, BIRC3, ANGPTL4, HSP90AA1, KTN1, ALOX5AP

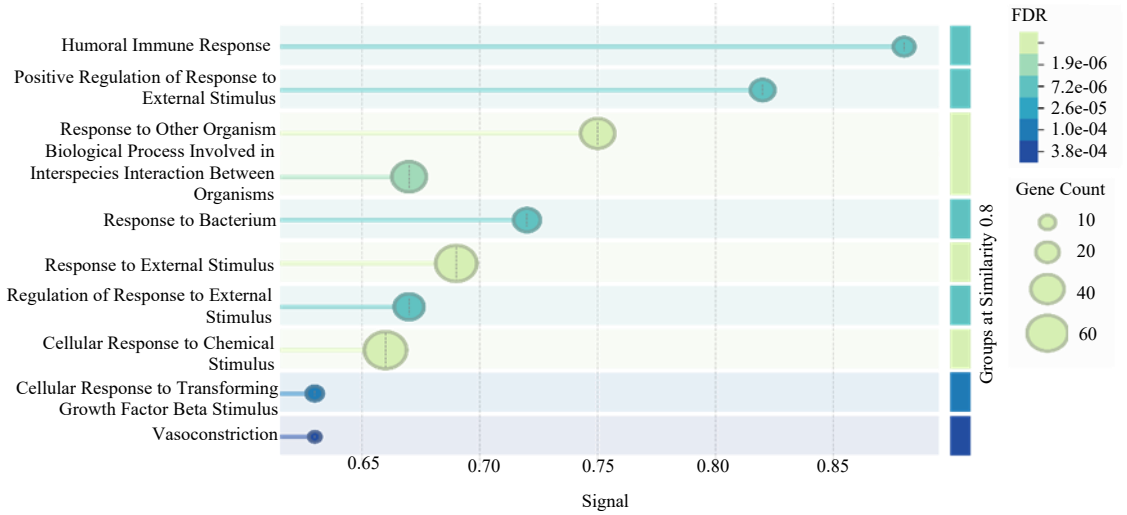


Figure 3. Representation of biological process (gene ontology) enrichment.

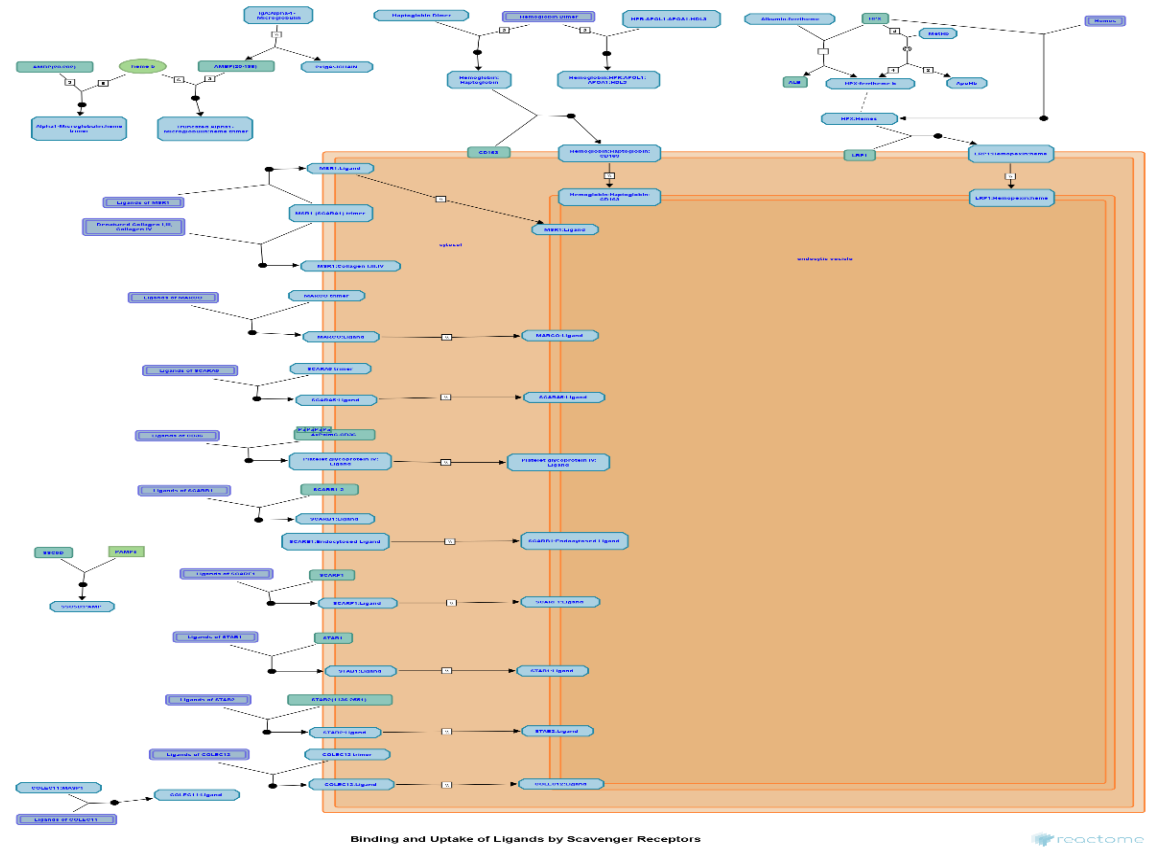
Protein–Protein Network

To explore the functional associations among the differentially expressed genes (DEGs) identified from datasets GSE110147 and GSE53845, a protein–protein interaction (PPI) network was constructed using the STRING database. The network was composed of 221 nodes and 450 edges, representing the interactions between protein products of the DEGs. The expected number of edges was 262, and the PPI enrichment p-value was < 1.0e-16, indicating that the observed interactions were significantly greater than what would be expected by chance. This suggests a high level of interconnectivity and biological relevance among the identified genes.

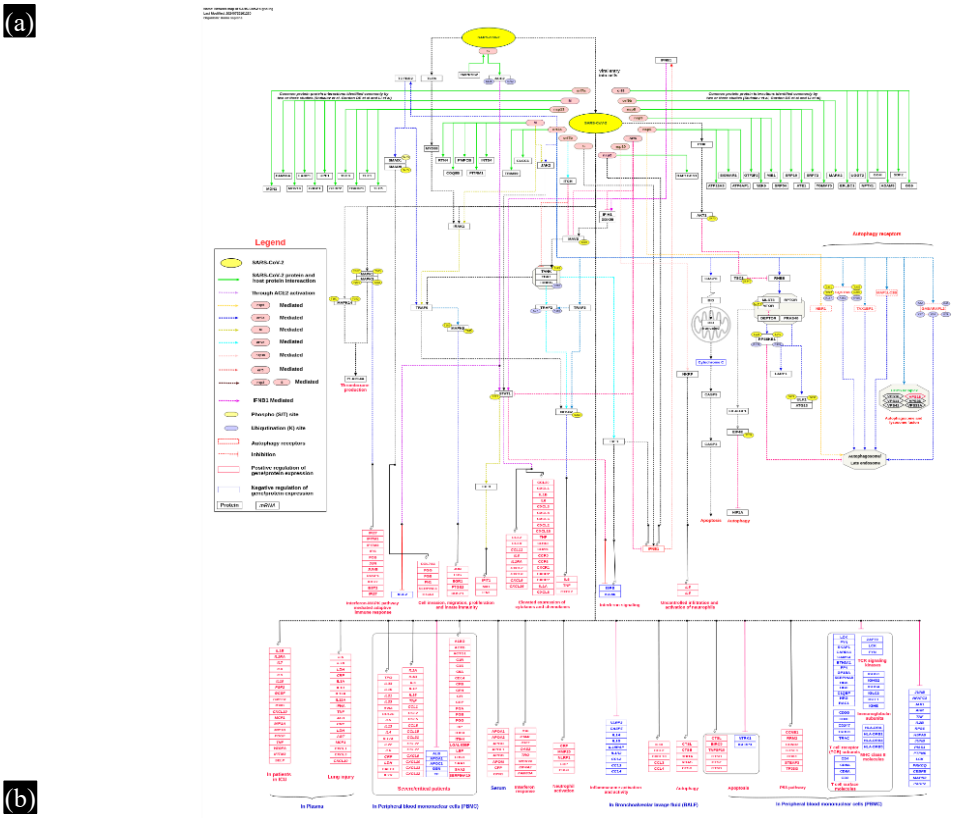
The average node degree was calculated to be 4.07, suggesting that, on average, each protein interacted with approximately four others within the network. Furthermore, the average local clustering coefficient was 0.457, reflecting a moderate tendency of the proteins to form tightly interconnected clusters or functional modules. These findings imply the presence of highly coordinated biological processes among the DEGs involved in Idiopathic Pulmonary Fibrosis (IPF). The Protein–Protein interaction network is shown in Figure 5.

Identification of Hub Genes

Following the construction of the PPI network using the STRING database, hub genes were identified based on the topological characteristics of the network. The most prominent hub genes identified include CCL19, IGLL1, JCHAIN, DEFA3, FCN3, CXCL10, CCL21, S100A12, CCL2, XCL1, CXCL14, CFH, and PLA2G1B. Functional annotation through Gene Ontology (GO) enrichment analysis revealed that these genes are significantly associated with biological processes, such as the humoral immune response, molecular function, such as protein binding, and cellular component, such as extracellular space.



(a)



(b)

Figure 4. Representation of pathway analysis. (a) Reactome pathway of binding and uptake of ligands by scavenger receptors (HSA-2173782). (b) wiki pathway of network map of SARS-CoV-2 signaling pathway (WP5115).

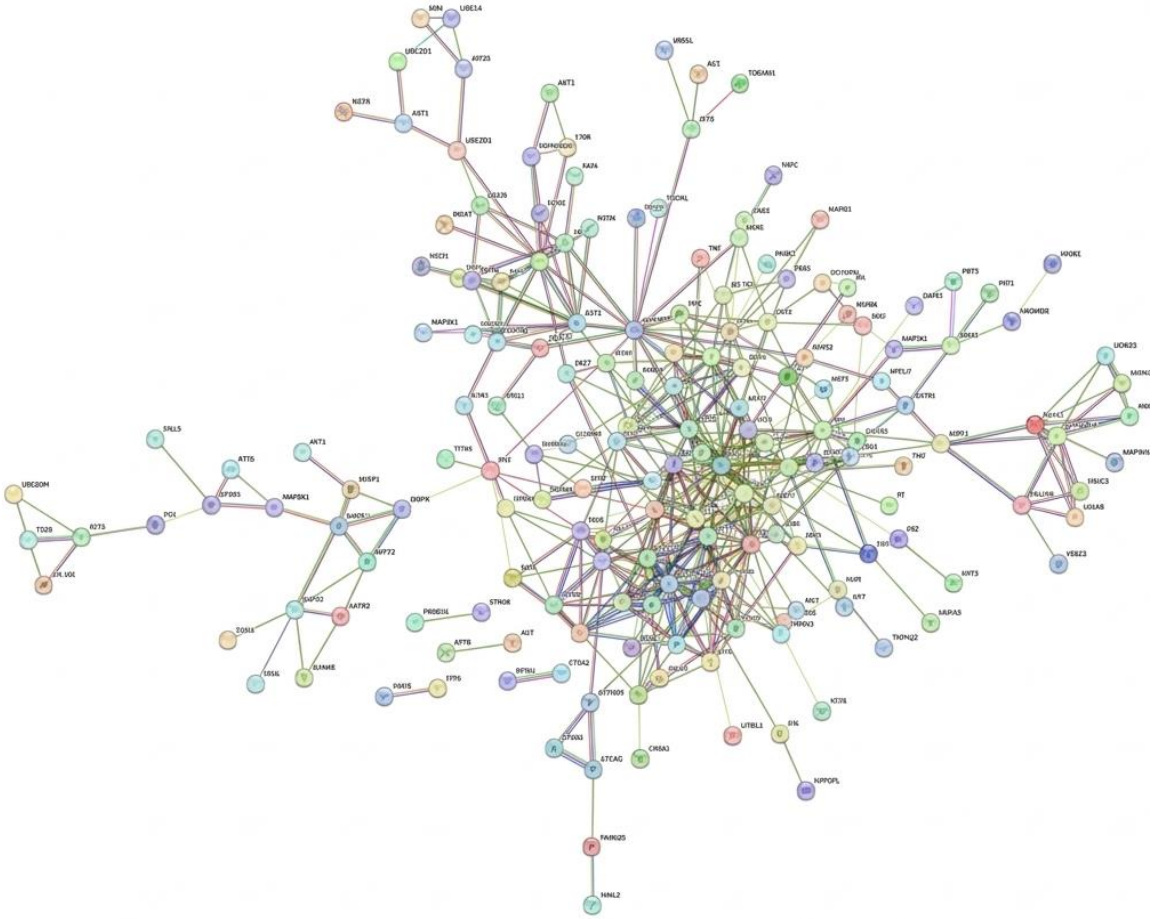


Figure 5. Protein–Protein interaction (PPI) network of differentially expressed genes (DEGs).

Multiple chemokines among the hub genes including CCL2, CCL19, CCL21, CXCL10, CXCL14, and XCL1 are involved in immune cell recruitment and inflammatory signaling, highlighting the importance of immune dysregulation in IPF. JCHAIN and IGLL1, which are critical components of immunoglobulin assembly, further support the role of altered humoral immunity. Additionally, genes like S100A12 and DEFA3, which are linked to antimicrobial responses and neutrophil activation, reflect the contribution of innate immunity to lung injury and fibrosis. These hub genes were also enriched in the extracellular space, suggesting a direct role in matrix interactions, inflammatory signaling, or tissue remodeling.

Among these, a refined subset of four genes that are CCL21, IGLL1, CXCL10, and CFH were further found to be associated with the SARS-CoV-2 signaling pathway in the pathway enrichment results. These genes play dual roles in IPF and viral-mediated lung injury, suggesting a shared immune-inflammatory axis that may exacerbate fibrotic responses during or after viral infections such as COVID-19. For instance, CXCL10, and CCL21 are chemokines induced during viral infection and have been implicated in the cytokine storm and immune cell migration observed in severe SARS-CoV-2 cases. CFH, a regulator of the complement system, contributes to controlling inflammation, while IGLL1 supports humoral immunity.

The convergence of these genes in both fibrotic lung disease and viral infection pathways supports the hypothesis that viral exposure, such as to SARS-CoV-2, may influence or aggravate fibrotic progression. These hub genes, therefore, represent potential biomarkers for early detection and promising therapeutic targets for managing IPF and its complications related to viral infections.

DISCUSSION

From the present analysis, a total of 319 top DEGs (from the top 1% of up- and downregulated genes) were identified and selected for downstream analysis. The large number of upregulated and downregulated genes across both dataset points to widespread transcriptional reprogramming in IPF. GO enrichment analysis offered insights into the biological roles of these DEGs across three domains. In the Biological Process (BP) category, top enriched terms, such as humoral immune response, response to other organisms, response to external stimulus, and regulation of response to chemical stimulus, reflect active immune signaling in the IPF lung microenvironment. This is consistent with the idea that chronic, aberrant immune activity contributes to tissue injury and fibrosis. The upregulation of cytokine and chemokine genes indicates heightened leukocyte recruitment and systemic immune involvement. Under Molecular Function (MF), terms, such as protein binding and chemokine activity, were highly enriched. These indicate that many DEGs encode molecules involved in cell signaling, inflammation, and cell–cell interactions. Chemokines like CCL2, CCL21, CXCL10, and CXCL14 not only bind to receptors to attract immune cells but also stimulate fibroblast activity and vascular remodeling – processes integral to fibrotic progression. In the Cellular Component (CC) domain, enrichment in extracellular space, extracellular region, and extracellular matrix reflects the localization of many of these gene products outside the cell, where they directly interact with tissue structure, matrix components, and immune effectors. The consistent presence of DEGs in extracellular locations supports their role in IPF pathogenesis, which is characterized by abnormal deposition of extracellular matrix proteins and disrupted alveolar architecture.

Reactome and WikiPathways enrichment analysis further contextualized these findings. Reactome pathways, such as collagen degradation, assembly of collagen fibrils, and integrin cell surface interactions, were highly enriched. These pathways are well established in fibrotic diseases and correlate with structural remodeling of the lung parenchyma. Collagens, matrix metalloproteinases (MMPs), and integrins coordinate to drive tissue stiffness and dysfunction. The enrichment of binding and uptake of ligands by scavenger receptors suggests a role for dysregulated innate immunity and tissue clearance mechanisms. WikiPathways enrichment analysis revealed a significant association with the SARS-CoV-2 signaling pathway. This finding is particularly compelling given recent clinical observations of post COVID-19 lung fibrosis. The shared molecular features suggest that immune and matrix responses in IPF may overlap with post-viral fibrotic pathways, reinforcing the idea that SARS-CoV-2 infection could mimic fibrotic processes.

A central aspect of this study was the identification of hub genes within the PPI network. A total of 14 hub genes were identified based on their high connectivity and involvement in key functional categories: CCL19, IGLL1, JCHAIN, DEFA3, FCN3, CXCL10, CCL21, S100A12, CCL2, XCL1, CXCL14, CFH, and PLA2G1B. These genes were commonly enriched in humoral immune response, protein binding, and extracellular space, indicating their integral role in inflammation, immune regulation, and tissue remodeling. Many of these hub genes encode chemokines (e.g., CCL2, CCL19, CCL21, CXCL10, CXCL14) that mediate immune cell recruitment, contributing to the persistent inflammation seen in IPF. Others, such as JCHAIN and IGLL1, are essential for immunoglobulin assembly and B-cell receptor signaling, highlighting a less-explored yet critical role of humoral immunity in lung fibrosis. Genes like S100A12 and DEFA3, associated with neutrophil activity and antimicrobial defense, suggest an interplay between infection-like immune responses and fibrotic tissue damage.

Among the 14 hub genes, a refined subset of four genes CCL21, IGLL1, CXCL10, and CFH was found to overlap with the SARS-CoV-2 signaling pathway. This convergence implies a molecular link between IPF pathogenesis and virally induced lung injury. CXCL10 and CCL21 are viral response chemokines that contribute to immune cell infiltration and cytokine storm phenomena. Several studies have indicated that the incidence and progression of IPF were closely linked to cytokines and chemokines [16, 17]. CFH, a regulator of the complement system, is crucial in modulating inflammation

and preventing immune overactivation. IGLL1, a marker of immature B cells reflects activation of the adaptive immune system in response to viral or tissue damage. This overlap supports growing clinical evidence that post-viral lung injury, particularly from COVID-19, may accelerate or trigger fibrotic pathways, especially in genetically or environmentally affected individuals. The dual involvement of these hub genes in both IPF and SARS-CoV-2-associated pathways highlights their potential as biomarkers for disease progression and targets for therapeutic intervention.

CONCLUSION

In this study, an integrative bioinformatics approach was employed to explore the molecular mechanisms underlying Idiopathic Pulmonary Fibrosis (IPF) by analyzing two publicly available microarray datasets (GSE110147 and GSE53845). A total of 319 highly significant differentially expressed genes (DEGs) were identified and analyzed through functional enrichment, pathway analysis, and protein–protein interaction network construction. The results highlighted key biological processes, such as humoral immune response, protein binding, and extracellular matrix involvement, all of which are central to the progression of IPF. A set of 14 hub genes including chemokines, immunoglobulin components, and inflammation-related genes were identified as potential molecular drivers of disease. Among these, four genes (CCL21, IGLL1, CXCL10, and CFH) were also enriched in the SARS-CoV-2 signaling pathway, suggesting a potential link between viral infection and fibrotic progression. These findings offer valuable insights into the complex interplay between immune dysregulation, extracellular remodeling, and potential environmental triggers in IPF. The identified hub genes not only represent potential biomarkers for early detection but also promising targets for therapeutic intervention. Further experimental validation and clinical research are warranted to confirm these results and support the development of targeted therapies to improve the management and prognosis of IPF patients.

Acknowledgment

I would like to express my heartfelt gratitude to everyone who contributed their knowledge and support towards the successful completion of my internship work and the project.

First and foremost, I would like to extend my sincere thanks to *BioNome* company and their staffs for providing their continuous support and guidance throughout the Bioinformatics data analysis course of this project.

I am deeply grateful to *Ms. Deekshitha M***.**, for her assistance in teaching Bioinformatics tools and molecular docking.

I am grateful to *Ms. Susha Dinesh,* for her assistance in teaching the NGS Analysis, and microarray profiling, which is a part of my project.

Their encouragement, expertise, and consistent support have been crucial in every stage of this project, and I am truly thankful for the opportunity to work under their guidance.

Lastly, I am thankful to Ramaiah institute of technology and department of Biotechnology for their encouragement and giving me the opportunity to carry out this internship and project.

Abbreviations

Abbreviation	Full form
IPF	Idiopathic Pulmonary Fibrosis.
DEG(s)	Differentially Expressed Gene(s).
GEO	Gene Expression Omnibus.
GEO2R	Gene Expression Omnibus to R tool.
GO	Gene Ontology.
BP	Biological Process.

MF	Molecular Function.
CC	Cellular Component.
KEGG	Kyoto Encyclopedia of Genes and Genomes.
PPI	Protein-Protein Interaction.
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins.
NSIP	Non-Specific Interstitial Pneumonia.
FDR	False Discovery Rate.
TGF-β	Transforming Growth Factor Beta.
COVID-19	Coronavirus Disease 2019.
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2.
FC	Fold Change.

REFERENCES

1. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. Lancet. 2017;389(10082):1941–52.
2. Zhang TT, Zhang JY, Li Y. Integrated bioinformatics analysis of key candidate genes and therapeutic drugs for idiopathic pulmonary fibrosis. Asian Toxicol Res. 2020;2(4):151–64.
3. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2083–92.
4. Varone F, Sgalla G, Iovene B, Bruni T, Richeldi L. Nintedanib for the treatment of idiopathic pulmonary fibrosis. Expert Opin Pharmacother. 2018;19(2):167–75.
5. Li N, Qiu L, Zeng C, Fang Z, Chen S, Song X, et al. Bioinformatic analysis of differentially expressed genes and pathways in idiopathic pulmonary fibrosis. Ann Transl Med. 2021;9(18):1459.
6. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(D1):D447–52.
7. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
8. Li D, Liu Y, Wang B. Identification of transcriptomic markers for developing idiopathic pulmonary fibrosis: An integrative analysis of gene expression profiles. Int J Clin Exp Pathol. 2020;13(7):1698–706.
9. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020;181(5):1036–45.
10. George PM, Wells AU, Jenkins RG. Pulmonary fibrosis and COVID-19: The potential role for antifibrotic therapy. Lancet Respir Med. 2020;8(8):807–15.
11. Wynn TA, Ramalingam TR. Mechanisms of fibrosis: Therapeutic translation for fibrotic disease. Nat Med. 2012;18(7):1028–40.
12. Todd NW, Atamas SP, Luzina IG, Galvin JR. Permanent alveolar collapse is the predominant mechanism in idiopathic pulmonary fibrosis. Expert Rev Respir Med. 2015;9(4):411–8.
13. Gross TJ, Hunninghake GW. Idiopathic pulmonary fibrosis. N Engl J Med. 2001;345(7):517–25.
14. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. Nat Rev Dis Primers. 2017;3(1):17074.
15. Wakwaya Y, Brown KK. Idiopathic pulmonary fibrosis: Epidemiology, diagnosis and outcomes. Am J Med Sci. 2019;357(5):359–69.
16. Zhang HP, Zou J, Xie P, Gao F, Mu H. Association of HLA and cytokine gene polymorphisms with idiopathic pulmonary fibrosis. 2015.
17. Alhamad EH, Cal JG, Shakoor Z, Almogren A, AlBoukai AA. Cytokine gene polymorphisms and serum cytokine levels in patients with idiopathic pulmonary fibrosis. BMC Med Genet. 2013;14:66.

18. Xu Z, Mo L, Feng X, Huang M, Li L. Using bioinformatics approach identifies key genes and pathways in idiopathic pulmonary fibrosis. Medicine (Baltimore). 2020;99(36):e22099.
19. Mahmud SH, Al-Mustanjid M, Akter F, Rahman MS, Ahmed K, Rahman MH, et al. Bioinformatics and system biology approach to identify the influences of SARS-CoV-2 infections to idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease patients. Brief Bioinform. 2021;22(5):bbab115.

