BMC Medical Genomics

RESEARCH

Open Access

Unravelling the transcriptomic characteristics of bronchoalveolar lavage in post-covid pulmonary fibrosis

Mohammad Shadab Ali[1], Vijay Hadda[1*], Sonia Verma[2], Anita Chopra[3], Saurabh Mittal[1], Karan Madan[1], Pawan Tiwari[1], Tejas Menon Suri[1] and Anant Mohan[1]

Abstract

Background Post-Covid Pulmonary Fibrosis (PCPF) has emerged as a significant global issue associated with a poor quality of life and significant morbidity. Currently, our understanding of the molecular pathways of PCPF is limited. Hence, in this study, we performed whole transcriptome sequencing of the RNA isolated from the bronchoalveolar lavage (BAL) samples of PCPF and compared it with idiopathic pulmonary fibrosis (IPF) and non-ILD (Interstitial Lung Disease) control to understand the gene expression profile and associated pathways.

Methods BAL samples from PCPF ($n=3$), IPF ($n=3$), and non-ILD Control ($n=3$) (individuals with apparent healthy lung without interstitial lung disease) groups were obtained and RNA were isolated for whole transcriptomic sequencing. Differentially Expressed Genes (DEGs) were determined followed by functional enrichment analysis and qPCR validation.

Results A panel of differentially expressed genes were identified in bronchoalveolar lavage fluid cells (BALF) of PCPF as compare to control and IPF. Our analysis revealed dysregulated pathways associated with cell cycle regulation, immune responses, and neuroinflammatory processes. Real-time validation further supported these findings. The PPI network and module analysis shed light on potential biomarkers and underscore the complex interplay of molecular mechanisms in PCPF. The comparison of PCPF and IPF identified a significant downregulation of pathways that were more prominent in IPF.

Conclusion This investigation provides crucial insights into the molecular mechanism of PCPF and also outlines avenues for prospective research and the development of therapeutic approaches.

Keywords PCPF, IPF, BAL, RNA-Seq, GSEA, CIBERSORTX

*Correspondence:
Vijay Hadda
vijayhadda@yahoo.com; vijayhadda@aiims.edu
[1]Department of Pulmonary, Critical Care, & Sleep Medicine, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India
[2]Division of Neuroscience and Ageing Biology, CSIR-Central Drug Research Institute, Lucknow, UP 226031, India
[3]Lab Oncology, Dr. BRAIRCH All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India

Background

In 2019, COVID-19 emerged as global health emergency affecting millions of people across the globe [1]; among survivors of COVID-19, approximately 70 to 80% of patients experienced a range of short- or long-term post-infectious sequelae, particularly among those who had severe disease [2, 3]. Among these sequelae, post-COVID pulmonary fibrosis (PCPF) has emerged as a significant health concern and has been reported even after mild or asymptomatic infection also [4–6]. The PCPF manifest as fibrosis-like alterations (in some cases true fibrosis) on computed tomographic scan of thorax and severe reductions in diffusing capacity for carbon monoxide (DLCO) on pulmonary function testing [7]. The affected individuals experience a diminished quality of life and necessitate extra medical attention [8]. Currently, the best therapeutic approach of PCPF is not known. Hence, molecular pathway analysis of PCPF is imperative to understand the potential therapeutic targets or strategies.

The molecular mechanism underlying some of the fibrotic lung diseases such as idiopathic pulmonary fibrosis (IPF), silica-induced pulmonary fibrosis, and hypersensitivity pneumonitis has been studied to some extent [9–12]. The understanding of these molecular mechanisms led to the discovery of specific antifibrotic drugs, such as pirfenidone, and nintedanib for IPF and other fibrotic lung diseases. Superficially, there are few similarities between IPF and PCPF such as both exhibit a diminished DLCO, disrupted alveolar-capillary integrity, imbalanced renin-angiotensin-aldosterone system (RAAS), and heightened oxidative stress [13, 14]. However, there are glaring differences also; for example - IPF is progressive disease whereas PCPF is reversible in many patients, PCPF is usually manifest as NSIP or NSIP/OP pattern on HRCT whereas IPF present as UIP. Nonetheless, our knowledge of the molecular mechanisms governing PCPF remains limited. The significance of delving into PCPF lies in our potential to not only elucidate the molecular landscape of PCPF but also to discern potential commonalities between PCPF and IPF. Furthermore, it offers an avenue to explore the functional attributes of these molecular entities.

The main objective of this pilot study was to identify the DEGs in BAL of PCPF compared to non-ILD control with apparently unaffected lung and IPF. We employed Gene Ontology (GO)/Kyoto Encyclopedia of Genes and Genomics (KEGG) enrichment analysis and Gene Set Enrichment analysis (GSEA) to investigate the functionalities and pathogenic pathways of various DEGs. The study aims to elucidate the association and causality of these genes to the pathogenesis of PCPF. Also, we aim to explore the commonality between PCPF and IPF. For this, we used next generation sequencing technology for transcriptomics analysis and further validated our findings using real time PCR.

Materials and methods

Ethical statement

This study was approved by the institutional review board of All India Institute of Medical Sciences, New Delhi (IEC-388/02.07.2021). Only patients who provided a written informed consent were included in the study.

Sample collection

The inclusion criteria for PCPF were as follow: Adults with a known history of SARS-CoV-2 infection confirmed by real-time PCR, having radiologically proven pulmonary fibrosis, and recovered from acute COVID-19 infection. All the recruited subjects had moderate to severe infection and required hospitalization.

A BAL sample of approx 10–15 mL were collected from PCPF patients ($n=3$), controls ($n=3$), and IPF ($n=3$) as per the standard guidelines [15, 16]. The sample were collected and stored in a sterile container for further processing.

RNA isolation, QC, RNA-sequencing, and data analysis

Following the sample collection, the cells from the BAL were pelleted down and the RNA was isolated using TRIzol™ reagent (Invitrogen, Thermo Fisher Scientific, Massachusetts, USA) as per the manufacturer's recommendation. The RNA samples were subjected to a quality check using a bioanalyzer (Agilent Technologies). We kept only those samples possessing an RNA integrity (RIN) score of more than 7 (Supplementary Fig. 1). Strand-specific libraries were prepared as per manufacture's recommendation (NEB labs™, Massachusetts, US). Paired-end whole transcriptomic sequencing was performed on Illumina NovaSeq 6000 platform.

The raw sequencing reads underwent processing to determine expression profiles. The reads were aligned to the human genome hg38. Identification of Differentially Expressed Genes (DEGs) including lncRNAs were carried out using iDEP. iDEP integrates data from Ensembl and other comprehensive annotation databases to distinguish lncRNAs from protein-coding genes. For differential expression analysis, a threshold of$|\log2FC| \geq 1$ and an FDR cut-off of 0.1 were applied (refer supplementary Fig. 2).

Functional enrichment analysis

To determines the specific GO (Gene Ontology) terms i.e., Biological Processes (BP), Molecular Functions (MF) and Cellular Components (CC) as well as KEGG pathways (Kyoto Encyclopedia of Genes and Genomics) we employed SRplot. We also used GSEA (Gene set

enrichment analysis) [17] for functional enrichment analysis using preranked gene list file.

Ingenuity Pathway Analysis (IPA)

We used IPA software for comprehensive pathway determination of differential gene expressed in PCPF as compare to control samples. The DEGs list was subjected to analysis using the IPA core expression analysis tool with a focus on utilizing the tool's canonical pathways and upstream functions.

Network analysis and module extraction

The interactions between the DEGs were identified using string-db database and Protein-Protein Interaction (PPI) network was constructed using cytoscape. Additionally, we used MCODE plugin for module extraction from the network and cytohubba to identify the hub genes using default parameters.

Validation

We employed SYBR-green qPCR to validate expression of 6 DEGs in PCPF patients, including 3 samples from RNA-seq. Target genes (FN1, FBN2, CCL13, CCL7, TIMD4, CCL24) were compared against controls, using GAPDH for normalization. Primers were designed with Primer3 software (Supplementary Table 1). Additionally, we compared gene expression in PCPF patients to whole blood samples from 3 healthy volunteers.

Cell-type composition characterization

To determine the cell composition of RNA-seq data of PCPF and IPF patients we used the CIBERSORTX algorithm. The raw count data of RNA-seq was processed and converted to CPM (count per million). Subsequently, the data was uploaded and run on the cibersortx platform with 100 permutation and LM22 was used as signature matrix.

Statistical analysis

Statistical analysis for real time PCR as well as cibersortx results were conducted using GraphPad Prism software (version 8.0.2). The comparison of gene expression between the PCPF and control groups was performed using the non-parametric Mann–Whitney U test. A p-value of less than 0.05 was considered statistically significant.

Table 1 Patient's demographic characterstics

	PCPF	IPF	Control	*p*-Value
Age (median)	53	66	35	0.4643
Gender				
Male	3	3	1	0.25
Female	0	0	2	0.25
Race	Indian	Indian	Indian	
Smoking exposure	1 (33.3%)	1 (33.3%)	0	> 0.9999

Results

The median age of participants was 53, 66, and 35 years for PCPF, IPF, and control groups, respectively, with no significant difference ($p = 0.4643$). 1 subject from each PCPF and IPF were exposed to smoking, in both PCPF, and IPF all subjects were males, while in control group there was a solitary male ($p = 0.25$) (Table 1).

Clustering analysis and identification of DEGs in PCPF vs. Control

The RNA-seq analysis was conducted on samples derived from 3 individuals with PCPF, and 3 control subjects. The t-SNE plot of the DEGs showed that the samples clustered into PCPF ($n = 3$) and controls ($n = 3$). The sample tree linkage analysis confirmed the results of the t-SNE plot (Fig. 1a and b).

Based on $|\log2FC| \geq 1$, and FDR cutoff 0.1, the analysis between PCPF and control groups revealed a total of 4374 DEGs, including both protein-coding and non-coding genes (volcano plot and heatmap given in Fig. 1c and d). Using filter of a p-value of ≤ 0.05 and removal of the non-coding RNA, we got a total of 2076 (upregulated 370 and down regulated 1706) DEGs (supplementary data). The top 5 highly upregulated genes include - CCL13, CCL24, TIMD4, MAMDC2, and FBN2. Whereas, top 5 downregulated genes were – HAS2, KRT24, ADAMTS19, DIPK2B, and TRPM5 (Table 2).

LncRNA expression profile in PCPF compared to control

We found a total of 993 (73 upregulated and 920 down regulated) differentially expressed lncRNA in PCPF as compared to control. As shown in Table 2, the majority of upregulated differentially expressed lncRNAs are novel and are not yet annotated. On the other hand, XIST, TARID, AIRN are among the top downregulated lncRNAs.

Functional enrichment and pathway analysis of PCPF vs. control

The gene set enrichment analysis (GSEA) revealed a number of upregulated pathways associated with cell cycle such as E2F targets, G2M checkpoints, and mitotic spindle function highlighting potential dysregulation of cell division machinery. Interestingly, dysregulation of immune-related pathways, including allograft rejection and inflammatory response indicates potential involvement of immune signaling in PCPF.

Conversely, there was only one pathway downregulated in PCPF i.e., KRAS signaling pathway (Fig. 2), indicating suppression of KRAS signaling pathways in PCPF. The

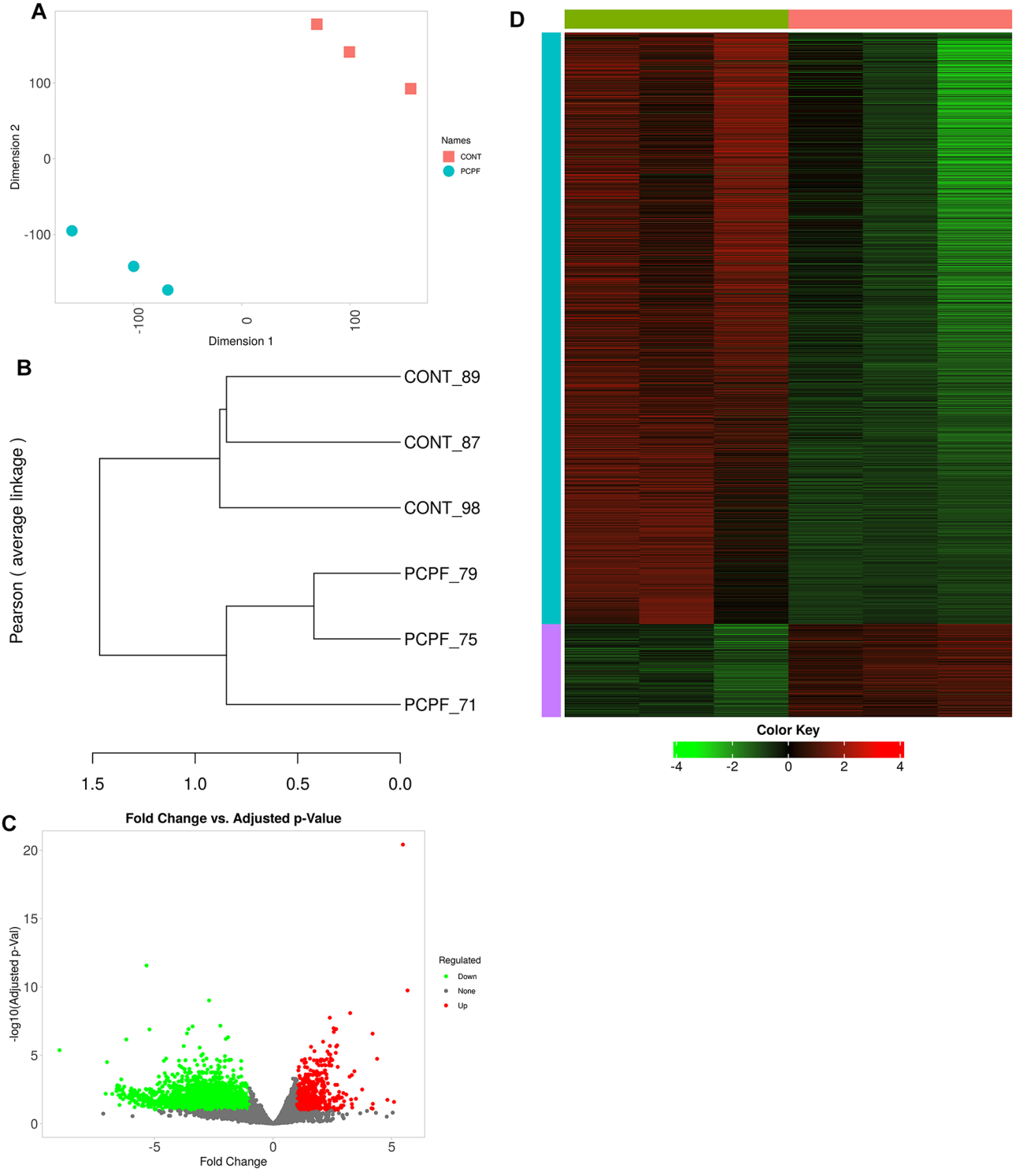

Fig. 1 Summarizes the RNA-Seq analysis of PCPF and control samples. **a** shows a t-SNE plot highlighting the distinct separation of PCPF (red) and control (blue) samples based on their gene expression profiles. **b** presents a hierarchical clustering dendrogram, demonstrating clear grouping of PCPF and control samples. **c** is a volcano plot showing differentially expressed genes (DEGs) between the two groups, with upregulated genes in red, downregulated genes in green, and non-significant genes in gray. Finally, **d** is a heatmap of DEGs, with a color gradient representing gene expression levels (red for upregulation and green for downregulation

Table 2 Top 10 DEGs (protein-coding & lncRNA) in PCPF as compared to control and IPF

PCPF vs. Control				
Regulation	Ensembl ID	log2 Fold Change	Adj.Pval	Symbol
Top 10 DEGs (protein-coding)				
Up	ENSG00000181374	5.667076576	1.82E-10	CCL13
Up	ENSG00000106178	5.474349103	3.87E-21	CCL24
Up	ENSG00000145850	4.385833287	1.81E-05	TIMD4
Up	ENSG00000165072	4.195600396	2.63E-07	MAMDC2
Up	ENSG00000138829	3.755964263	3.20E-03	FBN2
Down	ENSG00000170961	-7.055256938	6.49E-03	HAS2
Down	ENSG00000167916	-6.510791933	8.16E-03	KRT24
Down	ENSG00000145808	-6.485634646	1.79E-03	ADAMTS19
Down	ENSG00000147113	-6.453179274	3.55E-03	DIPK2B
Down	ENSG00000070985	-6.39532631	5.79E-04	TRPM5
Top 10 DElncRNA				
Up	ENSG00000229019	4.818954866	1.83E-02	
Up	ENSG00000259094	3.322362934	2.80E-04	
Up	ENSG00000235843	2.994813936	2.31E-02	
Up	ENSG00000224532	2.965714319	4.66E-02	
Up	ENSG00000280382	2.739452742	9.67E-02	
Down	ENSG00000229807	-8.9995467	4.22E-06	XIST
Down	ENSG00000227954	-6.77518376	6.73E-03	TARID
Down	ENSG00000268257	-6.595220455	4.18E-03	AIRN
Down	ENSG00000285569	-6.573342122	1.71E-03	
Down	ENSG00000253452	-6.561947339	1.55E-03	
PCPF vs. IPF				
Regulation	**Ensembl ID**	**log2 Fold Change**	**Adj.Pval**	**Symbol**
Top 10 DEGs (protein-coding)				
Up	ENSG00000134184	9.743764403	1.44E-03	GSTM1
Up	ENSG00000016490	6.840135496	4.60E-02	CLCA1
Up	ENSG00000269711	6.62481711	6.91E-04	
Up	ENSG00000152467	6.237656979	5.42E-02	ZSCAN1
Up	ENSG00000178473	6.229234454	6.20E-03	UCN3
Down	ENSG00000229894	-7.676491296	6.52E-04	GK3P
Down	ENSG00000196946	-6.802460602	2.96E-03	ZNF705A
Down	ENSG00000196611	-6.772002245	3.18E-08	MMP1
Down	ENSG00000121742	-6.613044034	5.91E-05	GJB6
Down	ENSG00000142319	-6.566829032	2.06E-02	SLC6A3
Top 10 DElncRNA				
Up	ENSG00000251611	6.110480546	1.91E-02	FAM160A1-DT
Up	ENSG00000224057	5.990518818	3.08E-02	EGFR-AS1
Up	ENSG00000234665	5.907536472	1.36E-02	LERFS
Up	ENSG00000231720	5.646051044	9.48E-02	
Up	ENSG00000248673	5.134530065	2.93E-02	LINC01331
Down	ENSG00000234389	-10.50956925	1.67E-03	
Down	ENSG00000249086	-9.094032834	7.42E-05	
Down	ENSG00000259342	-7.985121381	1.86E-04	
Down	ENSG00000237927	-7.926862277	1.76E-07	ENSG00000237927
Down	ENSG00000228318	-7.795711785	1.08E-04	

findings of these analysis are given in the supplementary Table 2.

We also performed the enrichment analysis with SRplot that uses clusterprofiler and pathview tools to determine the GO terms and KEGG pathways in the given datasets. The biological processes mainly included the microtubule movement, formation, membrane potential regulation, cilium movement etc., while molecular function included various activities related to gated channel, ion channel, cation channel, etc. While cAMP signaling,

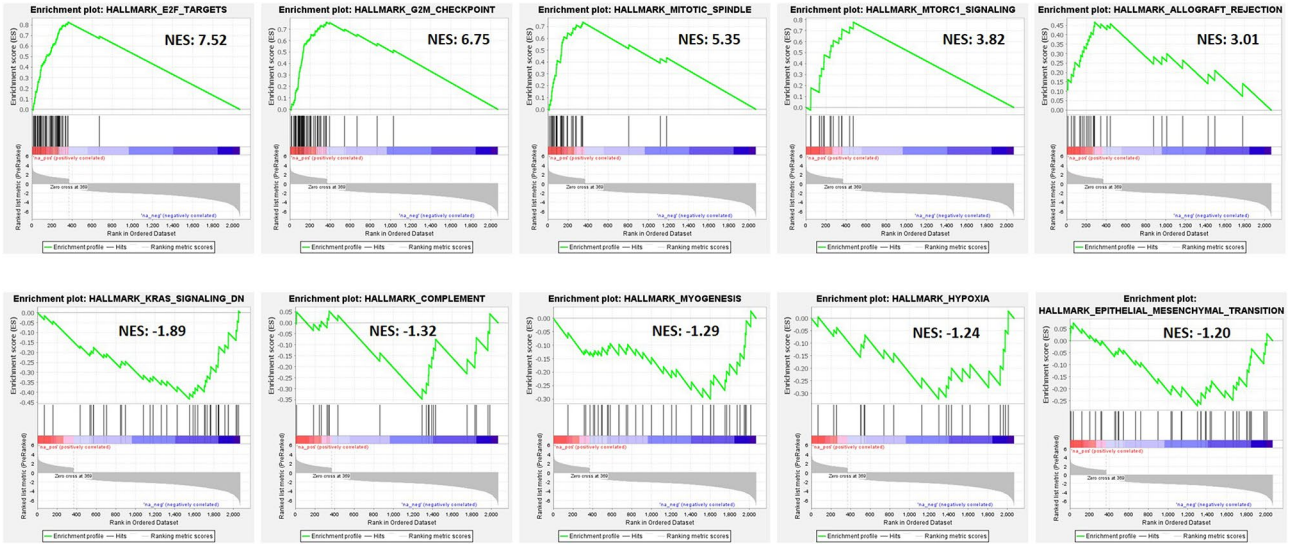

Fig. 2 The figure displays enrichment plots from Gene Set Enrichment Analysis (GSEA) highlighting key pathways differentially enriched between experimental groups. E2F Targets, G2M Checkpoint, Mitotic Spindle, MTORC1 Signaling, and Allograft Rejection are positively enriched, indicating their activation in the PCPF. Conversely, pathways such as KRAS Signaling Down, Complement, Myogenesis, Hypoxia, and Epithelial Mesenchymal Transition are negatively enriched, suggesting their suppression

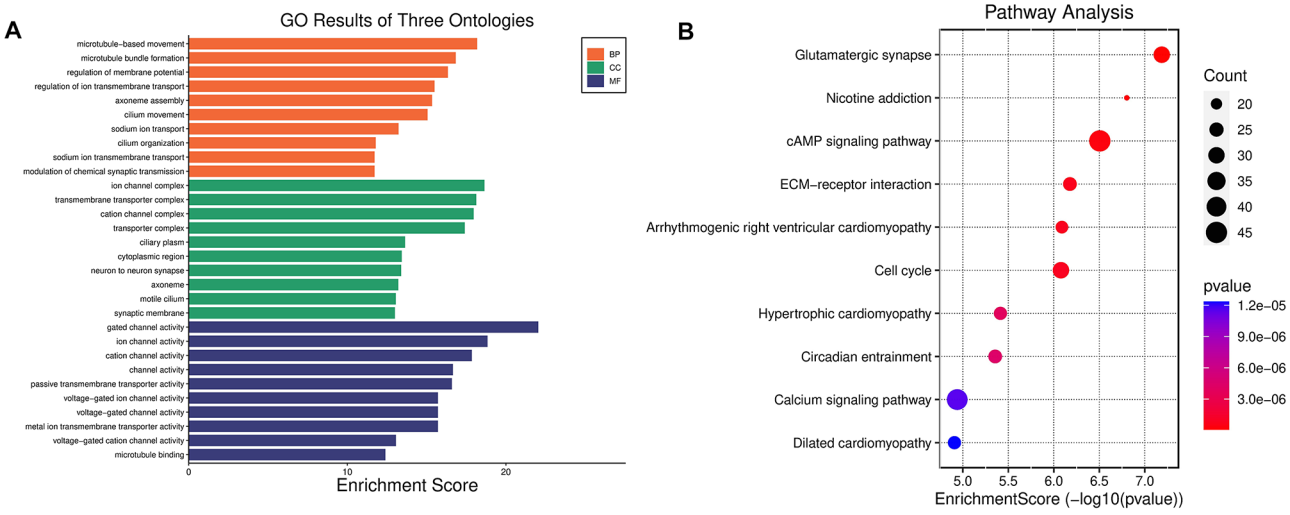

Fig. 3 Top Gene Ontology (BP, CC, and MF) terms and KEGG pathways in PCPF compared to control

ECM-receptor interaction, cell cycle were the top altered KEGG pathways in PCPF (Fig. 3).

To further explore the canonical pathways and upstream regulators, we performed the IPA analysis, wherein all DEGs were mapped to the curated pathways within the IPA database. The Glutaminergic Receptor Signaling Pathway emerged as the top altered canonical pathway. Additionally, we identified multiple altered pathways that includes S100 family signaling pathway, pulmonary fibrosis idiopathic signaling pathway, GPCR signaling pathway, Calcium signaling etc. A comprehensive list of these significant pathways is presented in supplementary Fig. 3. The most prominent upstream regulator as well as causal gene identified was CKAP2L (supplementary Fig. 4).

Validation by qPCR

To validate our RNA-seq data, we selected 6 highly DEGs with potential biological relevance to determine their expression by quantitative real time PCR. CCL13, CCL24, TIMD4, FBN2, FN1, CCL7 were selected for validation. The validation was performed in 6 PCPF patients. We observed a marked increase in expression of all the selected genes (except CCL24) in PCPF group as compared to control (Fig. 4a).

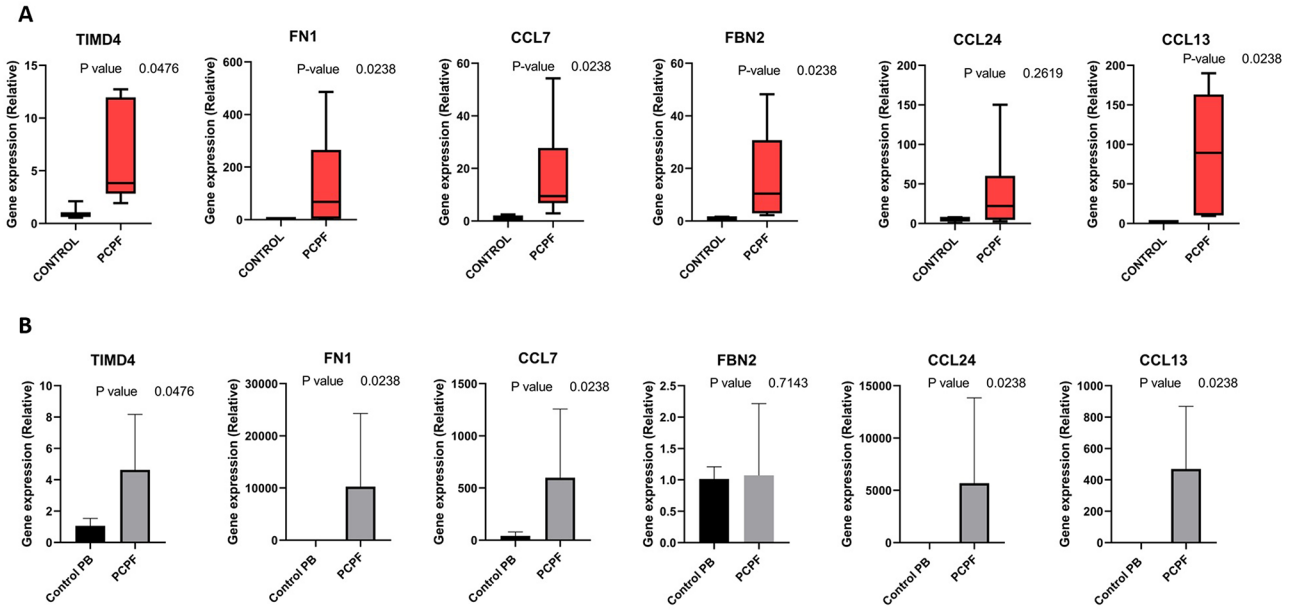

Fig. 4 **a**. qRT-PCR validation of top DEG in PCPF vs. Non-ILD Control (BAL sample), **b**. qRT-PCR validation of top DEG in PCPF vs. Healthy Control (PB sample)

Furthermore, we conducted a comparison of gene expression for all the aforementioned genes with blood samples obtained from healthy volunteers ($n = 3$). we found consistent outcomes as observed with BAL control samples. Except for the FBN2 gene, all other genes exhibited significantly elevated expression in PCPF patients (Fig. 4b).

PPI network construction and module analysis

We generated a PPI network from the top 500 DEGs with default settings. Then, we applied the MCODE plugin to pinpoint highly interconnected networks and selected the top two modules with scores exceeding 10. As determined by the ClueGO plugin for functional enrichment analysis, the majority of genes within these modules were associated with cell cycle activities, regulation, and immunological processes (supplementary Fig. 5a, & 5b). Additionally, we used the Cytohubba plugin to identify the top 10 genes based on their degree rank, the results of the same are shown in supplementary Fig. 6 and supplementary Table 3.

Principle component and differential gene expression analysis between PCPF and IPF

We used PCA plot to show the clustering of samples. Normalized gene expression was used as input. We observed a distinct clustering of both PCPF and IPF samples. Though the distance between the samples were rather prominent but both the samples were clustered differently. Tree linkage analysis also showed that the PCPF and IPF samples are differently clustered (Fig. 5).

We have used the same pipeline for DEG analysis. Our analysis revealed a total of 5143 differentially expressed genes (3203 upregulated and 1940 downregulated) including both protein coding as well as non-coding. After filtering out the non-coding genes the number of differentially expressed protein coding genes was found to be 3230 (upregulated 1687 and downregulated 1543) in our data. Some of the top upregulated gene in PCPF includes GSTM1, CLCA1, UCN3, PRB3, HIF3A. on the other hand, GK3P, ZNF705A, MMP1, GJB6, SLC6A3 are among the top downregulated genes in PCPF as compare to IPF (Table 2).

Unique LncRNA signatures distinguishing PCPF from IPF

We have found a total of 1035 differentially expressed lncRNA in PCPF as compared to IPF. A total of 127 lncRNAs were upregulated while 908 lncRNAs were downregulated in PCPF. FAM160A1-DT, EGFR-AS1, LERFS, ENSG00000231720, LINC01331 are the most significantly upregulated lncRNAs in PCPF. While most of the top downregulated lncRNAs are novel for example ENSG00000234389, ENSG00000249086, ENSG00000259342, and ENSG00000237927 (Table 2).

Functional enrichment analysis of PCPF compared to IPF

In gene ontology the BP terms associated with the PCPF were mainly related to immune cell dysfunction including neutrophil degranulation, T-cell activation, T-cell differentiation etc. while NAD + nucleosidase, NAD(P) + nucleosidase activity, TLR binding, cytokine receptor activity were among the prominent molecular functions. On the other hand, NF-kappa B signaling pathway, TNF signaling

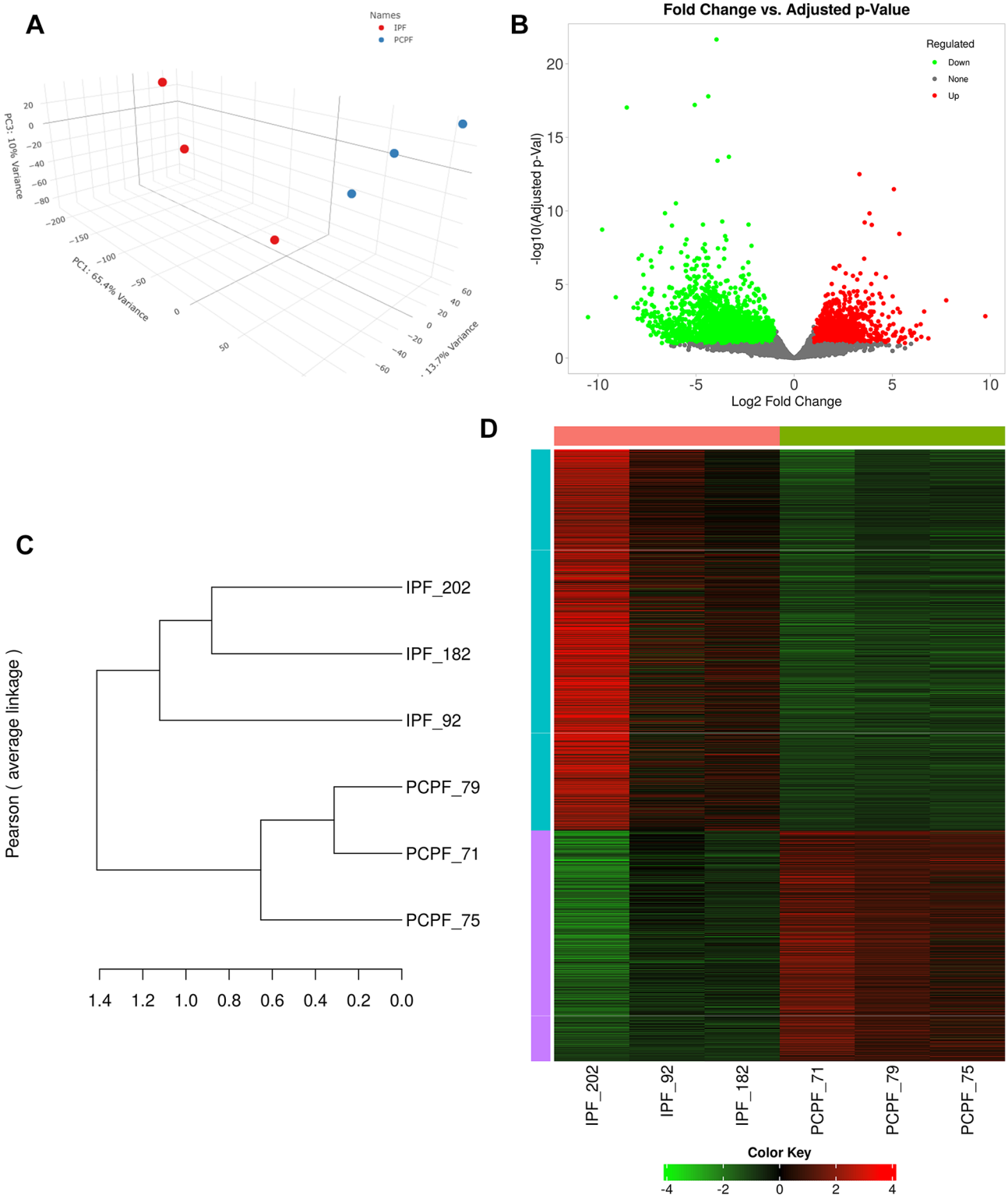

Fig. 5 **a** Shows a PCA plot with clear separation between IPF (red) and PCPF (blue). **b** volcano plot highlighting significantly upregulated (red) and downregulated (green) genes in IPF. **c** presents a hierarchical clustering dendrogram showing distinct sample groupings by expression similarity. **d** displays a heatmap of differentially expressed genes, with red and green indicating upregulation and downregulation, respectively, emphasizing distinct molecular signatures in IPF

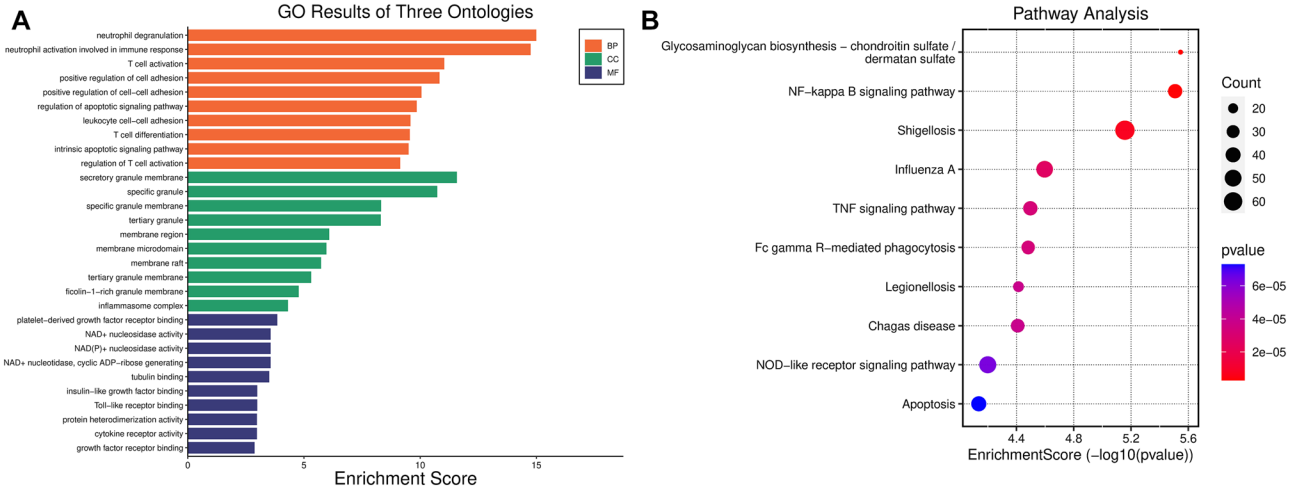

Fig. 6 Top Gene Ontology (BP, CC, and MF) terms and KEGG pathways in PCPF compared to IPF

Table 3 GSEA hallmark pathways downregulated in PCPF as compare to IPF

Name	SIZE	ES	NES	NOM *p*-val
HALLMARK_TNFA_SIGNALING_VIA_NFKB	108	-0.62962	-3.93488	0
HALLMARK_INFLAMMATORY_RESPONSE	84	-0.55236	-3.34064	0
HALLMARK_INTERFERON_GAMMA_RESPONSE	71	-0.41462	-2.38424	0
HALLMARK_COMPLEMENT	60	-0.41	-2.25602	0
HALLMARK_HYPOXIA	55	-0.40864	-2.21896	0
HALLMARK_IL6_JAK_STAT3_SIGNALING	22	-0.53735	-2.16349	0
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	43	-0.40913	-2.0622	0.001667
HALLMARK_IL2_STAT5_SIGNALING	53	-0.3766	-1.99473	0
HALLMARK_KRAS_SIGNALING_UP	44	-0.38679	-1.92633	0
HALLMARK_INTERFERON_ALPHA_RESPONSE	37	-0.39574	-1.92295	0.0033

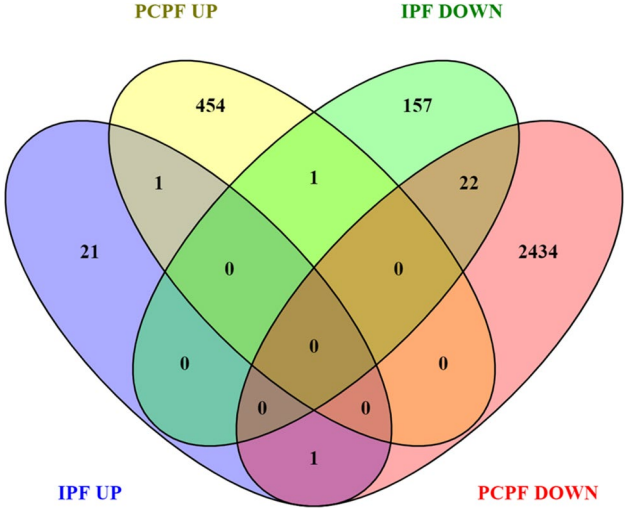

Fig. 7 Venn diagram of number of DEGs common in IPF and PCPF

pathway, NOD-like receptor signaling, and apoptotic pathways were among the top altered KEGG pathways in PCPF as compared to IPF (Fig. 6).

We have used GSEA to analyse the functional enrichment of protein coding genes, we did not observe any significant upregulated hallmark pathway in PCPF as compare to IPF. However, in PCPF a number of hallmark pathways were downregulated that includes; TNFA signaling via NFKB, inflammatory response, IFN gamma response, complement activation, hypoxia, EMT among the most prominent pathways as given in the Table 3.

Commonalities between PCPF and IPF

To identify shared features between PCPF and IPF, we utilized the DEG from both cohorts compared to the control groups. We employed a Venn diagram to identify genes commonly upregulated and downregulated. Subsequently, we conducted functional enrichment analysis on these commonly identified genes.

We have found a total of 139 commonly upregulated genes and 1886 commonly downregulated genes between PCPF and IPF (Fig. 7). Functional enrichment of upregulated genes revealed the involvement of the cellular processes that includes: mitotic nuclear division, chromosome segregation, nuclear division, cell division etc. For MF, Plus-end-directed microtubule motor activity, Microtubule motor activity, ATPase activity, Microtubule binding, were the main functions. On the other hand,

KEGG pathway revealed the upregulation of the following pathways: Cell cycle, cellular senescence, p53 signaling, oocyte meiosis pathway (supplementary Fig. 7a).

Functional enrichment of commonly downregulated genes revealed following GO terms; BP: Cilium movement, axoneme assembly, mictubule movement, and bundle formation. MF: various cellular channel activities including; ion channel, gated channel, cation channel, sodium channel etc. while KEGG pathways includes; glutamatergic, cAMP signaling, nictotine addiction, calcium signaling as the top pathways (supplementary Fig. 8).

Cell type characterization between PCPF and IPF

We used CIBERSORTX algorithm to characterize the cellular composition in the two groups, and determined the fractions of 22 immune cell types. As expected, the macrophage was the most prominent cell type in our data as the BAL sample is mostly consists of macrophage.

The expression of B-naïve cells, T cells CD4 memory resting, NK cells resting, monocytes, macrophage M0, M1, and M3, and neutrophils were quite prominent in both PCPF as well as IPF but statistically they were not significantly differed from each other. Likewise, there were obvious difference observe among the following cell types; T cells CD8, plasma cells, dendritic cells resting, dendritic cells activated but were again no significant difference was observed between IPF and PCPF.

Discussion

Although our sample size is small, but according to Conesa A, et al., even with three replicates per group, when considering a fold change of 2, the statistical power to detect Differentially Expressed Genes (DEGs) becomes well over 85%. Simultaneously, the sequencing depth also plays a key role. For example, a sequencing depth of 15 million has a statistical power of 38% to detect DEGs [18]. In our case the sequencing depth is 100 million which makes it more efficient in detecting DEGs. Although some previous studies have even lesser depth with 3 replicates per group [19, 20].

In this study, we report a distinct gene signature in PCPF compared to the control group that are primarily associated with immunological and extracellular matrix-related genes. Among the top upregulated genes in PCPF, we observed the inclusion of CCL13, CCL24, TIMD4, MAMDC2, and FBN2 [21]. Moreover, from our finding pathways related to inflammatory response, TNF-α, IL2-STAT5 as well cell cycles related pathways were the most important pathways suggesting that immune response and inflammation are key hallmarks of PCPF. Our findings align with reports that SARS-CoV-2-induced cytokine storm exacerbates inflammation, especially in the elderly, contributing to pulmonary fibrosis [22, 23].

The top upregulated signature genes in PCPF patients are mainly inflammatory and immune related genes. Both CCL13, and CCL24 are proinflammatory genes and are upregulated in PCPF. Although CCL13 was also reported in IPF but its expression was not found significant in our dataset. Despite its variable expression across different studies these genes acts as chemokine signature in these conditions and are responsible for fibrogenesis [24]. TIMD4 also called T-cell membrane protein 4 and is mainly expressed on antigen presenting cells including macrophage. Though the role of TIMD4 is not well explored in any fibrotic lung disease but its high expression is observed in CHP [25].

Gene Set Enrichment Analysis (GSEA) provides valuable insights into molecular pathway dysregulation in post-COVID pulmonary fibrosis. It reveals upregulated pathways related to cell cycle regulation, cellular growth, and metabolic processes, alongside immune response pathways. The prominence of cell cycle regulation, highlighted by both GSEA and KEGG pathway analyses, underscores its pivotal role in PCPF molecular changes. Elevated E2F TARGETS suggest enhanced cell proliferation contributing to tissue remodeling, while upregulated G2M checkpoint pathway implies increased cell division aiding tissue repair. These dysregulated pathways observed in PCPF, shared with other IPF and pulmonary fibrosis model, emphasize their central role in pulmonary fibrosis [26, 27]. In our study KRAS signaling pathway was the solitary downregulated pathway involved in the PCPF. Studies suggest its association with lung inflammation and tumor growth, with KRAS mutations implicated in interstitial pneumonia-related lung adenocarcinoma. Its downregulation in PCPF may signify loss of control over cell growth and survival pathways, yet also indicates a complex regulatory response to the disease [28, 29].

The KEGG and the IPA both revealed the Glutamatergic Synapse and glutaminergic receptor signaling pathway respectively as the top pathways altered in PCPF. Hamanaka et al., has demonstrated that the biosynthesis of amino acids derived from glutamine and glutamate is essential for both myofibroblast differentiation and the production of collagen proteins in human lung fibroblasts [30]. These results suggest that the genes involved in these pathways are important for the pathogenesis of pulmonary fibrosis and can be targeted for therapeutic purposes.

cAMP (cyclic adenosine monophosphate) is a key intracellular messenger involved in various cellular processes. Elevated levels of cAMP have been shown to hinder fibroblast proliferation and extracellular matrix (ECM) synthesis, demonstrating anti-fibrotic effects both in laboratory settings and within living organisms [31]. Dysregulation of this pathway could affect signaling and

cellular responses in lung tissue, potentially contributing to fibrosis in PCPF.

The extracellular matrix (ECM) offers structural support to tissues and participates in processes like wound healing and tissue remodeling. Given that ECM deposition and interactions with ECM receptors have historically been significant factors driving pulmonary fibrosis [32, 33], it is entirely expected to observe alterations in this pathway in PCPF as well.

IPA analysis revealed changes in neuroinflammatory pathways, implicating COVID-19 with long-term neurological symptoms and psychological disorders [34]. The presence of these pathways in PCPF patients' molecular profiles also indicates a possible connection between neurological and synaptic processes and the development or progression of pulmonary fibrosis. A previous report have suggested that many of the proinflammatory molecules can lead to CNS inflammation that can cause respiratory center in brain to dysfunction which can cause the number of lung problems [35].

The results of the Protein-Protein Interaction (PPI) network analysis using the MCODE and cytohubba plugin have identified the top modules and hub genes with high centrality and potential significance in the context of the study. Almost all the hub genes we identified were associated with cell cycle and cell division. For instance, TOP2A gene encodes for DNA topoisomerase and involve in chromosome condensation and is associated with a number of pathways including cell cycle, and mitosis. TOP2A is highly expressed in many solid tissue cancers and is also a chemotherapeutic targets in many cancers [36, 37]. AURKB is involve in cell cycle regulation and interact with each other. Targeting this gene can be handy in treatment of pulmonary fibrosis. A previous study has also shown that by inhibiting the AURKB pulmonary fibrosis can be attenuated [38].

We observed a significant downregulation of many pathways in PCPF as compare to IPF. The results obtained from the GSEA analysis comparing PCPF with IPF reveal intriguing insights into the nature of PCPF progression and potential reversibility. Firstly, the absence of significant upregulated hallmark pathways in PCPF compared to IPF suggests that the molecular mechanisms driving fibrosis in PCPF may differ from those in IPF. This distinction could imply that PCPF may have a less aggressive or slower progression compared to IPF, which is known for its relentless and often fatal course [39].

Secondly, the observation of downregulated hallmark pathways in PCPF, such as TNFA signaling via NFKB, inflammatory response, IFN gamma response, complement activation, hypoxia, and epithelial-to-mesenchymal transition (EMT), suggests a potential avenue for reversibility. Downregulation of these pathways could indicate a dampening of pro-inflammatory and fibrotic processes, possibly due to resolution of the underlying COVID-19 infection or a different inflammatory response pattern compared to IPF.

The less pronounced activation of inflammatory and fibrotic pathways in PCPF, as evidenced by the downregulation of hallmark pathways, may contribute to its observed reversibility in many cases. Additionally, the potential ability of PCPF to resolve or stabilize could be influenced by factors such as the timing and severity of the initial COVID-19 infection, individual immune responses, and treatment interventions. Further research is warranted to elucidate the precise mechanisms underlying the differences between PCPF and IPF and to explore strategies for optimizing treatment outcomes in PCPF.

It is crucial to emphasize that the specific mechanisms underpinning these associations necessitate further investigation. Currently, there are no studies available that have examined these pathways in the context of post-COVID pulmonary fibrosis. This study on PCPF represents a pioneering effort that not only uncovers the molecular landscape of the condition but also lays the groundwork for transformative advancements in diagnosis, prognosis, and therapeutic interventions. Our study has some limitation first one is the sample size, second one is the use of BAL sample, if we have used all three samples' types i.e., blood, BAL, and tissue that would have given as a more comprehensive picture.

Conclusion

This study explores Post Covid Pulmonary Fibrosis at molecular level. It emphasizes the scarcity of information on PCPF's molecular pathogenesis and how it is different from IPF. Transcriptome analysis reveals distinct gene expression patterns, identified key genes and pathways involved in PCPF. While we have confirmed our results through real-time PCR, it's important to note that the sample size was limited. Nevertheless, these preliminary findings offer valuable insights for guiding future research in this area.

Abbreviations

BAL	Bronchoalveolar Lavage
BP	Biological Process
CC	Cellular Component
CIBERSORTX	Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts
DEG	Differentially Expressed Genes
DLCO	Diffusing Capacity of the Lungs for Carbon Monoxide
GSEA	Gene Set Enrichment Analysis
GO	Gene Ontology
HRCT	High-Resolution Computed Tomography
ILD	Interstitial Lung Disease
IPA	Ingenuity Pathway Analysis
IPF	Idiopathic Pulmonary Fibrosis
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	Long Non-Coding RNA
MCODE	Molecular Complex Detection

MF	Molecular Function
OP	Organized Pneumonia
NSIP	Non-Specific Interstitial Pneumonia
OP	Organizing Pneumonia
PCPF	Post-Covid Pulmonary Fibrosis
PPI	Protein-Protein Interaction
qPCR	Quantitative Polymerase Chain Reaction
RAAS	Renin-Angiotensin-Aldosterone System
RNA	Ribonucleic Acid
RIN	RNA Integrity Number
UIP	Usual Interstitial Pneumonia

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-025-02110-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements

MSA acknowledge ICMR for senior research fellowship (3/1/3/JRF-2020/HRD-001) and Harsh Goyal for his contribution in qPCR experiments.

Author contributions

MSA: Concept, design, experiments, data analysis, results interpretation, manuscript writing, VH: concept, design, result interpretation, manuscript writing, SV: concept, design, result interpretation, AC: provided platform for RNA-seq analysis, concept, SM, KM, PT, TMS, AM helps in bronchoscopy procedure and manuscript revision. All authors read and approved the manuscript.

Funding

This study is supported by intramural grant of AIIMS, New Delhi (Grant ID: A-COVID-105).

Data availability

The datasets generated and/or analyzed for this study are available in the NCBI repository with BioProject number: PRJNA1105214.

Declarations

Ethics approval and consent to participate

This study was performed in accordance with the Declaration of Helsinki. This human study was approved by the Institute Ethics Committee, AIIMS, New Delhi - approval: IEC-388/02.07.2021. All adult participants provided written informed consent to participate in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 20 April 2024 / Accepted: 20 February 2025
Published online: 17 March 2025

References

1. Hussain A, Yadav S, Hadda V, Suri TM, Tiwari P, Mittal S et al. Covid-19: a comprehensive review of a formidable foe and the road ahead. Expert Rev Respir Med [Internet]. 2020;14(9):869–79. Available from: https://www.tandfonline.com/doi/full/https://doi.org/10.1080/17476348.2020.1782198
2. Carfì A, Bernabei R, Landi F. Persistent symptoms in patients after acute COVID-19. JAMA [Internet]. 2020;324(6):603. Available from: https://jamanetwork.com/journals/jama/fullarticle/2768351
3. Ahmed OF, Kakamad FH, Hama Amin BJ, Abdullah BA, Hassan MN, Salih RQ et al. Post COVID-19 pulmonary complications; a single center experience. Ann Med Surg [Internet]. 2021;72:103052. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2049080121010025
4. Polak SB, Van Gool IC, Cohen D, von der Thüsen JH, van Paassen J. A systematic review of pathological findings in COVID-19: a pathophysiological timeline and possible mechanisms of disease progression. Mod Pathol [Internet]. 2020;33(11):2128–38. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0893395222004367
5. Udwadia ZF, Koul PA, Richeldi L. Post-COVID lung fibrosis. Lung India [Internet]. 2021;38(Suppl 1):S41–7. Available from: https://journals.lww.com/10.4103/lungindia.lungindia_818_20
6. Antony T, Vishak Acharya K, Unnikrishnan B, Keerthi NS. A silent march-Post covid fibrosis in asymptomatics – A cause for concern? Indian J Tuberc [Internet]. 2023;70(2):249–52. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0019570722000725
7. Bazdyrev E, Rusina P, Panova M, Novikov F, Grishagin I, Nebolsin V. Lung Fibrosis after COVID-19: Treatment Prospects. Pharmaceuticals [Internet]. 2021;14(8):807. Available from: https://www.mdpi.com/1424-8247/14/8/807
8. Raimundo K, Chang E, Broder MS, Alexander K, Zazzali J, Swigris JJ. Clinical and economic burden of idiopathic pulmonary fibrosis: a retrospective cohort study. BMC Pulm Med [Internet]. 2016;16(1):2. Available from: http://www.biomedcentral.com/1471-2466/16/2
9. Vukmirovic M, Kaminski N. Impact of transcriptomics on our understanding of pulmonary fibrosis. Front Med [Internet]. 2018;5. Available from: http://journal.frontiersin.org/article/https://doi.org/10.3389/fmed.2018.00087/full
10. Bo C, Zhang J, Sai L, Du Z, Yu G, Li C et al. Integrative transcriptomic and proteomic analysis reveals mechanisms of silica-induced pulmonary fibrosis in rats. BMC Pulm Med [Internet]. 2022;22(1):13. Available from: https://bmcpulmmed.biomedcentral.com/articles/https://doi.org/10.1186/s12890-021-01807-w
11. Mishra S, Shah MI, Udhaya Kumar S, Thirumal Kumar D, Gopalakrishnan C, Al-Subaie AM et al. Network analysis of transcriptomics data for the prediction and prioritization of membrane-associated biomarkers for idiopathic pulmonary fibrosis (IPF) by bioinformatics approach. In: Advances in Protein Chemistry and Structural Biology [Internet]. 2021;241–73. Available from: https://linkinghub.elsevier.com/retrieve/pii/S187616232030081X
12. Ibarra G, Herazo-Maya JD, Kaminski N. Evolving Genomics of Pulmonary Fibrosis. In. 2019;207–39. Available from: http://link.springer.com/https://doi.org/10.1007/978-3-319-99975-3_9
13. Yoon HY, Uh ST. Post–coronavirus disease 2019 pulmonary fibrosis: wait or needs intervention. Tuberc Respir Dis (Seoul) [Internet]. 2022;85(4):320–31. Available from: http://e-trd.org/journal/view.php?doi=10.4046/trd.2022.0053
14. Shchudro O, Bielosludtseva K, Pertseva T, Konopkina L, Krykhtina M, Konopkina L. Carbon monoxide diffusing capacity (DLCO) in COVID-19 survivors versus idiopathic pulmonary fibrosis (IPF): the pathogenetic features. In: Rehabilitation and chronic care [Internet]. European Respiratory Society; 2021;PA3900. Available from: http://publications.ersnet.org/lookup/doi/https://doi.org/10.1183/13993003.congress-2021.PA3900
15. Mohan A, Madan K, Hadda V, Tiwari P, Mittal S, Guleria R et al. Guidelines for diagnostic flexible bronchoscopy in adults: Joint Indian Chest Society/National College of chest physicians (I)/Indian association for bronchology recommendations. Lung India [Internet]. 2019;36(8):37. Available from: https://journals.lww.com/https://doi.org/10.4103/lungindia.lungindia_108_19
16. Meyer KC, Raghu G, Baughman RP, Brown KK, Costabel U, du Bois RM et al. An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. Am J Respir Crit Care Med [Internet]. 2012;185(9):1004–14. Available from: https://www.atsjournals.org/doi/https://doi.org/10.1164/rccm.201202-0320ST
17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci [Internet]. 2005;102(43):15545–50. Available from: https://doi.org/10.1073/pnas.0506580102
18. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A et al. A survey of best practices for RNA-seq data analysis. Genome Biol [Internet]. 2016;17(1):13. Available from: https://genomebiology.biomedcentral.com/articles/https://doi.org/10.1186/s13059-016-0881-8

19. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect [Internet]. 2020;9(1):761–70. Available from: https://www.tandfonline.com/doi/full/https://doi.org/10.1080/22221751.2020.1747363
20. Verma S, Kumar A, Narang R, Bisoi AK, Mitra DK. Signature transcriptome analysis of stage specific atherosclerotic plaques of patients. BMC Med Genomics [Internet]. 2022;15(1):99. Available from: https://bmcmedgenomics.biomedcentral.com/articles/https://doi.org/10.1186/s12920-022-01250-8
21. Cameron MJ, Kelvin DJ. Cytokines and chemokines—their receptors and their genes: an overview. In: Advances in Experimental Medicine and Biology [Internet]. 2003;8–32. Available from: http://link.springer.com/https://doi.org/10.1007/978-1-4615-0171-8_2
22. Pelaia C, Tinello C, Vatrella A, De Sarro G, Pelaia G. Lung under attack by COVID-19-induced cytokine storm: pathogenic mechanisms and therapeutic implications. Ther Adv Respir Dis [Internet]. 2020;14. Available from: https://journals.sagepub.com/doi/https://doi.org/10.1177/1753466620933508
23. Napoli C, Tritto I, Benincasa G, Mansueto G, Ambrosio G. Cardiovascular involvement during COVID-19 and clinical implications in elderly patients. A review. Ann Med Surg [Internet]. 2020;57:236–43. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2049080120302375
24. Russo RC, Quesniaux VFJ, Ryffel B. Homeostatic chemokines as putative therapeutic targets in idiopathic pulmonary fibrosis. Trends Immunol [Internet]. 2023;44(12):1014–30. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1471490623002107
25. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS et al. Chronic hypersensitivity pneumonitis, an interstitial lung disease with distinct molecular signatures. Am J Respir Crit Care Med [Internet]. 2020;202(10):1430–44. Available from: https://www.atsjournals.org/doi/https://doi.org/10.1164/rccm.202001-0134OC
26. Emblom-Callahan MC, Chhina MK, Shlobin OA, Ahmad S, Reese ES, Iyer EPR et al. Genomic phenotype of non-cultured pulmonary fibroblasts in idiopathic pulmonary fibrosis. Genomics [Internet]. 2010;96(3):134–45. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0888754310001138
27. Vukmirovic M, Yan X, Gibson KF, Gulati M, Schupp JC, Deluliis G et al. Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. Eur Respir J [Internet]. 2021;58(6):2002950. Available from: http://publications.ersnet.org/lookup/doi/https://doi.org/10.1183/13993003.02950-2020
28. Sriwastva MK, Teng Y, Mu J, Xu F, Kumar A, Sundaram K et al. An extracellular vesicular mutant KRAS-associated protein complex promotes lung inflammation and tumor growth. J Extracell Vesicles [Internet]. 2023;12(2). Available from: https://onlinelibrary.wiley.com/doi/https://doi.org/10.1002/jev2.12307
29. Kataoka T, Okudela K, Matsumura M, Baba T, Kitamura H, Arai H et al. Significant accumulation of KRAS mutations in bronchiolar metaplasia–associated honeycomb lesions of interstitial pneumonia. Oncol Lett [Internet]. 2022;24(1):225. Available from: http://www.spandidos-publications.com/https://doi.org/10.3892/ol.2022.13346
30. Hamanaka RB, O'Leary EM, Witt LJ, Tian Y, Gökalp GA, Meliton AY et al. Glutamine metabolism is required for collagen protein synthesis in lung fibroblasts. Am J Respir Cell Mol Biol [Internet]. 2019;61(5):597–606. Available from: https://www.atsjournals.org/doi/https://doi.org/10.1165/rcmb.2019-0008OC
31. Liu Y, Xu H, Geng Y, Xu D, Zhang L, Yang Y et al. Dibutyryl-cAMP attenuates pulmonary fibrosis by blocking myofibroblast differentiation via PKA/CREB/CBP signaling in rats with silicosis. Respir Res [Internet]. 2017;18(1):38. Available from: http://respiratory-research.biomedcentral.com/articles/https://doi.org/10.1186/s12931-017-0523-z
32. 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.
33. Thannickal VJ, Henke CA, Horowitz JC, Noble PW, Roman J, Sime PJ et al. Matrix biology of idiopathic pulmonary fibrosis. Am J Pathol [Internet]. 2014;184(6):1643–51. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002944014001564
34. Zhao S, Toniolo S, Hampshire A, Husain M. Effects of COVID-19 on cognition and brain health. Trends Cogn Sci [Internet]. 2023;27(11):1053–67. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1364661323002048
35. Dey J, Alam MT, Chandra S, Gupta J, Ray U, Srivastava AK et al. Neuroinvasion of SARS-CoV-2 may play a role in the breakdown of the respiratory center of the brain. J Med Virol [Internet]. 2021;93(3):1296–303. Available from: https://onlinelibrary.wiley.com/doi/https://doi.org/10.1002/jmv.26521
36. Meng J, Wei Y, Deng Q, Li L, Li X. Study on the expression of TOP2A in hepatocellular carcinoma and its relationship with patient prognosis. Cancer Cell Int [Internet]. 2022;22(1):29. Available from: https://cancerci.biomedcentral.com/articles/https://doi.org/10.1186/s12935-021-02439-0
37. Uusküla-Reimand L, Wilson MD. Untangling the roles of TOP2A and TOP2B in transcription and cancer. Sci Adv [Internet]. 2022;8(44). Available from: https://www.science.org/doi/https://doi.org/10.1126/sciadv.add4920
38. Kasam RK, Ghandikota S, Soundararajan D, Reddy GB, Huang SK, Jegga AG et al. Inhibition of Aurora Kinase B attenuates fibroblast activation and pulmonary fibrosis. EMBO Mol Med [Internet]. 2020;12(9). Available from: https://www.embopress.org/doi/https://doi.org/10.15252/emmm.202012131
39. Puglisi S, Torrisi S, Giuliano R, Vindigni V, Vancheri C. What we know about the pathogenesis of idiopathic pulmonary fibrosis. Semin Respir Crit Care Med [Internet]. 2016;37(03):358–67. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0036-1580693

Publisher's note