



# Transcriptomics using lung resection material to advance our understanding of COPD and idiopathic pulmonary fibrosis pathogenesis

Copyright © The authors 2024

This version is distributed under the terms of the Creative Commons Attribution Licence 4.0.

Received: 26 Jan 2024  
Accepted: 11 April 2024

To the Editor:

COPD and idiopathic pulmonary fibrosis (IPF) are both chronic lung diseases with distinct pathologies. COPD is characterised by progressive airflow obstruction and encompasses emphysema leading to loss of alveolar structure and chronic bronchitis including inflammation [1]. IPF is a progressive fatal interstitial lung disease without effective treatment, characterised by fibrosis of the lung [2]. Risk factors such as smoking and symptoms such as shortness of breath are common to both. Although the clinical features are well defined and categorised, the differences and similarities between mechanistic pathways have not been fully investigated.

Recent studies have shown that both COPD and IPF show altered lung tissue gene expression compared to controls, with a subset of overlapping genes and disease mechanisms [3, 4].

In this study, we aimed to see whether lung tissue from COPD or IPF patients' tissue resections could be used to identify gene and pathway expression differences between control and COPD or IPF lung tissue, and provide new mechanistic insight.

Lung tissue was obtained from the Royal Papworth Hospital research tissue bank. Informed consent was obtained and ethical approval was granted by the Royal Papworth Hospital Tissue Bank (Research Ethics Committee references 08/H0304/56+5 and 18/EE/0269). Tissue was taken from adjacent normal tissue from post-lobar resection or whole/partial-lung transplant procedures. Clinical and demographic information was obtained where possible. The tissue was rinsed in Tyrode's buffer and a piece for RNA extraction was placed in 1 mL RNAlater overnight at 4°C after which the RNAlater was removed, and the tissue frozen at -80°C. RNA was extracted using the Qiagen AllPrep DNA/RNA Mini Kit and sent to Oxford Genomics for RNA sequencing. All samples had RNA integrity numbers  $\geq 8$ .

Gene expression was determined through stranded paired-end polyA RNA sequencing using Illumina version 1.5 chemistry and the NovaSeq6000 at a minimum of 25 million reads. Raw data were trimmed using Trimmomatic and aligned to the Genome Reference Consortium human build 37 assembly using STAR. All data processing and analysis was conducted as previously described [5]. Results were corrected for donor age, sex, smoking history and RNA extraction batch, and adjusted using the Benjamini-Hochberg procedure (false discovery rate 5%,  $p < 0.05$ ).

The Ingenuity Pathway Analysis (IPA) tool (Qiagen) was used to determine disease, pathway and upstream regulator effect associations for the differentially expressed genes as previously described [5].

Comparison of clinical characteristics between the control, COPD and IPF groups identified significant differences in male/female sex ratios, smoking history and lung function, as anticipated (figure 1a).

Gene expression analysis between the control and COPD groups identified a total of 135 differentially expressed genes. Genes such as interleukin receptor subunit  $\alpha 2$  (*IL13RA2*), matrix proteins (*COMP*,

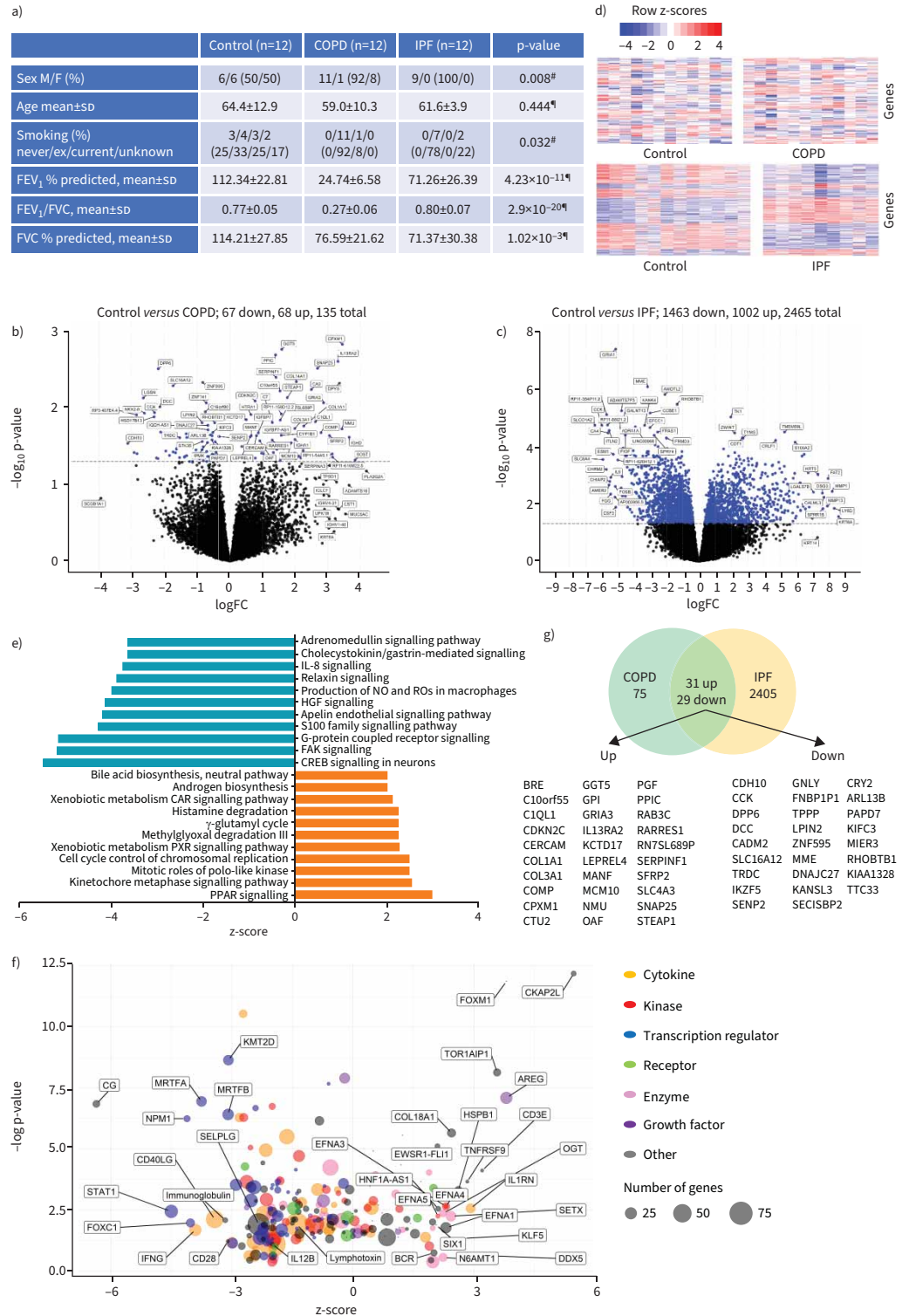


Shareable abstract (@ERSpublications)

Genes involved in cell death, inflammation and viral infection are common to both COPD and IPF. A link to rheumatic disease is unique to COPD, and IPF-specific analyses showed increases in gene expression of keratins, collagens, mucins and MMPs. <https://bit.ly/3JoW73H>

Cite this article as: Rakkar K, Thakker D, Portelli MA, et al. Transcriptomics using lung resection material to advance our understanding of COPD and idiopathic pulmonary fibrosis pathogenesis. *ERJ Open Res* 2024; 10: 00061-2024 [DOI: 10.1183/23120541.00061-2024].





**FIGURE 1** Donor information, differential gene expression and pathway analysis changes between control, COPD and idiopathic pulmonary fibrosis (IPF) groups. **a)** Tissue donor demographic information and analysis. Volcano plots and corresponding heatmaps for differential gene expression analysis between **b)** control and COPD and **c)** control and IPF groups with **d)** corresponding heatmaps. Results were corrected for donor age, sex, smoking history and RNA extraction batch, and adjusted using the Benjamini–Hochberg procedure (false discovery rate 5%;  $p < 0.05$ ). **e)** Ingenuity pathway analysis (IPA) for the control versus IPF differential gene expression dataset showing top activated (z-score  $> 2$ , orange) and inhibited (z-score  $< -2$ , blue) pathways. **f)** IPA for the control versus IPF differential gene expression dataset showing activated (z-score  $> 2$ ) and inhibited

(z-score <2) upstream regulators ( $p < 0.05$ ). g) Overlapping genes between the differential expression analysis for the control versus COPD and control versus IPF datasets. FEV<sub>1</sub>: forced expiratory volume in 1 s; FVC: forced vital capacity; FC: fold change; #: Chi-squared test; \*: one-way ANOVA.

*COL1A1* and *COL14A1*), growth regulators (*EGFL6*, *SOST* and *SFRP2*) and G-proteins (*RAB3C* and *GPR17*) showed logarithmically transformed fold changes (FC) all >2; whereas genes such as cadherin (*CDH10*) and those related to molecule processing (*GSTA3* and *HSD17B13*), phospholipid binding (*SYTL5*), potassium channels (*DPP6*) and cytoskeletal proteins (*LGSN*) all showed log FC < -2. (Figure 1b).

Gene expression analysis between the control and IPF groups resulted in a total of 2465 differentially expressed genes. The most notable overexpressed genes were related to keratins (*KRT5*, *KRT16*, *KRT15* and *KRT13*), collagen (*COL17A1*), mucin (*MUC5AC*) and matrix metalloproteinases (MMPs) (*MMP1*, *MMP13*, *MMP7* and *MMP8*), which all had log FC >5. Genes that were underexpressed were more diverse, such as protein/ion transport (*SLCO1A2* and *SLC6A4*), cytokines (*IL6*, *CSF3*), muscarinic receptor (*CHRM2*) and angiogenesis (*ESM1*), which all showed log FC < -5 (figure 1c).

IPA results from the control versus COPD gene expression dataset showed a small number of associations possibly due to the small number of differentially expressed genes and the individual subject heterogeneity observed in the heatmap (figure 1d). Only the wound healing signalling pathway and upstream regulators *SORL1* and *CG* were seen to be activated with a z-score >2 (data not shown). No pathways or regulators were found to be inhibited and there were no strongly linked diseases apart from those related to cancer.

IPA results from the control versus IPF dataset showed a greater number of associations. Pathways such as peroxisome proliferator-activated signalling, xenobiotic metabolism and cell cycle-related mechanisms were activated, and all had z-scores >2 (figure 1e, only top 10 pathways shown). There were a larger number of inhibited pathways such as those involved cytokine signalling (*IL8*, *IL17* and *IL6*), nitric oxide signalling, white blood cell signalling (natural killer cells and macrophages, which all had z-scores <2) (figure 1e, only top 10 pathways shown). Furthermore, there were also associations with disease, such as IPF ( $p = 9.23 \times 10^{-18}$ ), chronic pulmonary disease ( $p = 3.71 \times 10^{07}$ ), fibrosis ( $p = 6.54 \times 10^{-19}$ ) and cardiac hypertrophy ( $p = 1.15 \times 10^{-4}$ ). >60% of disease associations were cancer related (data not shown).

In the IPF dataset, upstream regulators or factors such as *COL18A1* and *IL1RN* were predicted to be active compared to factors such as *STAT1*, *IFNG* and *IL12B*, which were inhibited in the IPF tissue (figure 1f). Most inhibited factors tended to be transcription regulators or cytokines.

Investigation of the overlap between the COPD and IPF gene expression datasets showed that there were 60 common genes representing 44% of the total differentially expressed genes from the control versus COPD comparison and 2.4% of the total differentially expressed genes in the control versus IPF comparison (figure 1g). Of these, 31 genes were upregulated in both diseases such as *COL31A1*, *COL1A1* and *IL13RA2* and 29 genes were down regulated in both diseases such as *MME*, *DNAJC27* and *KIFC3*. There were no genes which were common to both datasets but showed gene expression in opposite directions.

IPA on overlapping genes showed processes such as “apoptosis” ( $p = 0.002$ ), “extracellular matrix organisation” ( $p = 3.89 \times 10^{-4}$ ) and “necrosis” ( $p = 0.001$ ) were associated with the common upregulated genes, and processes such as “infection by RNA virus” ( $p = 0.024$ ) and “cell junction organisation” ( $p = 0.005$ ) were associated with the common downregulated genes. “Chronic inflammatory disorder” ( $p = 0.015$ ) was associated with both down- and upregulated genes, suggesting these may be common pathways/mechanisms to both diseases.

Interestingly, pathways involved in rheumatic disease ( $p = 0.015$ ) appeared to be associated only with genes differentially upregulated in the control versus COPD dataset (*CYP7B1*, *DPYS*, *HTRA1*, *IGFBP7*, *IGHD*, *SNAI2* and *TMEM266*), indicating genes in this pathway may be unique to COPD lung tissue and/or that rheumatic disease and COPD are strongly linked comorbidities. Over 100 common upstream regulators were identified, of which >70% had either *COL1A1* or *COL3A1* as their target molecule, indicating collagen is a key factor in both diseases, supporting evidence from previous studies [6, 7]. There were no upstream regulators identified by IPA that were unique to the COPD or IPF gene dataset.

This study aimed to further our mechanistic understanding of COPD and IPF *via* transcriptomics using lung resection tissue. Importantly, we show that there is a subset of genes that are differentially expressed in the same direction in both IPF and COPD, suggesting common causal pathways with associated opportunities for therapeutic intervention. In the COPD-specific analyses, we could not provide significant insight due to the small number of differentially expressed genes and inconclusive IPA, although a general increase in matrix-associated protein and growth regulator gene expression was seen and there was a strong link to rheumatic disease. This link has been extensively reported [8] and the genes highlighted in our study may provide novel avenues of exploration. The IPF-specific analyses supported previous literature with regards to increases in expression levels of genes implicated in IPF pathogenesis (*e.g.* elevated levels of keratins, collagens, mucins and MMPs) and, importantly, identified other genes for additional study [9–12].

Although this study would benefit from a larger discovery and a replication cohort due to its modest sampling, and while donor heterogeneity, especially in the COPD tissue, and incomplete comorbidity and medication data limit our ability to fully define common/different pathways between COPD and IPF, this study highlights some potential mechanistic pathways that are involved in both COPD and IPF, such as apoptosis, necrosis, chronic inflammation and viral infection. Similarly, the IPF analyses identified genes involved in transcriptional regulation and inflammation, supporting previously associated factors/pathways, such as MMPs and fibrosis. This study also provides confirmation of pathways and genes unique to COPD, which are involved in other diseases such as rheumatic disease. This study provides a foundation for further in-depth study of distinct and overlapping pathways contribution to these difficult to treat diseases.

Kamini Rakkar <sup>1</sup>, Dhrama Thakker <sup>1</sup>, Michael A. Portelli <sup>1</sup>, Ian Hall<sup>1</sup>, Holger Schlüter<sup>2</sup> and Ian Sayers <sup>1</sup>

<sup>1</sup>Centre for Respiratory Research, NIHR Nottingham Biomedical Research Centre, School of Medicine, Biodiscovery Institute, University of Nottingham, Nottingham, UK. <sup>2</sup>Immunology and Respiratory Department, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany.

Corresponding author: Kamini Rakkar ([kamini.rakkar2@nottingham.ac.uk](mailto:kamini.rakkar2@nottingham.ac.uk))

Provenance: Submitted article, peer reviewed.

Ethics statement: This study received approval from the Royal Papworth Hospital Research Tissue Bank Research Ethics Committee (ref. 08/H0304/56+5 and 18/EE/0269).

Conflict of interest: D. Thakker has nothing to disclose. K. Rakkar reports support for the present manuscript from Boehringer Ingelheim. M.A. Portelli reports support for the present manuscript from Boehringer Ingelheim; funding from the University of Nottingham UNICAS fund, outside the submitted work; and payment for lectures, presentations, speakers' bureaus, manuscript writing or educational events from the University of Malta, outside the submitted work. I. Hall reports support for the present manuscript from Boehringer Ingelheim; grants or contracts from the NIHR and Wellcome Trust, outside the submitted work; and is vice chair of the Trustees of Asthma+Lung UK (unpaid), disclosure made outside the submitted work. H. Schlüter reports support for the present manuscript from Boehringer Ingelheim and is employed by Boehringer Ingelheim. I. Sayers reports support for the present manuscript from Boehringer Ingelheim; and research grants received from GlaxoSmithKline, the Biotechnology and Biological Sciences Research Council, a Wellcome Trust Discovery Award and the Nottingham NIHR Biomedical Research Centre, outside the submitted work.

Support statement: This study was supported by Boehringer Ingelheim. Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- 1 Singh D, Agusti A, Anzueto A, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019. *Eur Respir J* 2019; 53: 1900164.
- 2 Martinez FJ, Collard HR, Pardo A, *et al.* Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017; 3: 17074.
- 3 Ghosh AJ, Hobbs BD, Yun JH, *et al.* Lung tissue shows divergent gene expression between chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Respir Res* 2022; 23: 97.
- 4 Kim S, Herazo-Maya JD, Kang DD, *et al.* Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* 2015; 16: 924.

- 5 Rajasekar P, Hall RJ, Binaya KC, *et al.* Nepalese indoor cookstove smoke extracts alter human airway epithelial gene expression, DNA methylation and hydroxymethylation. *Environ Pollut* 2023; 337: 122561.
- 6 Kranenburg AR, Willems-Widyastuti A, Moori WJ, *et al.* Enhanced bronchial expression of extracellular matrix proteins in chronic obstructive pulmonary disease. *Am J Clin Pathol* 2006; 126: 725–735.
- 7 Upagupta C, Shimbori C, Alsilmi R, *et al.* Matrix abnormalities in pulmonary fibrosis. *Eur Respir Rev* 2018; 27: 180033.
- 8 Gergianaki I, Tsiligiani I. Chronic obstructive pulmonary disease and rheumatic diseases: A systematic review on a neglected comorbidity. *J Comorb* 2019; 9: 2235042X18820209.
- 9 Ballester B, Milara J, Cortijo J. Mucins as a new frontier in pulmonary fibrosis. *J Clin Med* 2019; 8: 1447.
- 10 Todd JL, Vinisko R, Liu Y, *et al.* Circulating matrix metalloproteinases and tissue metalloproteinase inhibitors in patients with idiopathic pulmonary fibrosis in the multicenter IPF-PRO Registry cohort. *BMC Pulm Med* 2020; 20: 64.
- 11 Hewitt RJ, Puttur F, Gaboriau DCA, *et al.* Lung extracellular matrix modulates KRT5<sup>+</sup> basal cell activity in pulmonary fibrosis. *Nat Commun* 2023; 14: 6039.
- 12 Vukmirovic M, Herazo-Maya JD, Blackmon J, *et al.* Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis. *BMC Pulm Med* 2017; 17: 15.