Interactions between airway epithelial cells and the microenvironment in idiopathic pulmonary fibrosis

Richard James Hewitt

Inflammation, Repair and Development Section
National Heart and Lung Institute
Imperial College London
London SW7 2AZ

A thesis submitted for the degree of Doctor of Philosophy

2021

Declaration of originality

I conducted or directly supervised all the experiments and work described in this thesis, with the following exceptions. Dr Andrew Rogers performed transmission electron microscopy at the Royal Brompton Hospital (Chapter 3). Lorraine Lawrence performed wax- embedding, sectioning, haematoxylin and eosin (H&E) staining and periodic acid-Schiff (PAS) staining of air-liquid interface culture inserts (Chapter 3). Poonam Ghai and Dr Patricia Ogger conducted flow cytometry, and Jessica Rowley of the Imperial Flow Core Facility carried out fluorescence activated cell sorting (Chapter 4). I prepared all RNA sequencing libraries in Anne O'Garra's laboratory, and the Francis Crick Institute's Advanced Sequencing Facility (ASF) performed the sequencing, demultiplexed the data, and provided raw FASTQ files for downstream analysis (Chapter 4). I prepared single-cell suspensions from airway brushings and Dr Jimena Perez-Lloret performed single-cell library construction and sequencing in the ASF (Chapter 5). Single-cell RNA-seq data analysis was performed in collaboration with Probir Chakravarty, who executed all coding for the project and generated figures where indicated in the legends (Chapter 5). Poonam Ghai conducted MUC5B genotyping (Chapter 5). Professor Anand Devaraj (Radiology Department, Royal Brompton Hospital) performed computed tomography scan scoring (Chapter 5). Professor Andrew Nicholson and Dr Alexandra Rice (Histopathology Department, Royal Brompton Hospital) provided IPF and control lung tissue sections (Chapter 6). I stained the lung tissue sections and identified regions for further evaluation. Dr Fred Fercoq acquired spectral and second harmonic generation imaging of the tissue and performed initial image analysis at the Cancer Research UK Beatson Institute in Glasgow (Chapter 6). I performed further analysis and interpretation of the data in collaboration with Dr Fred Fercoq and Dr Leo Carlin (Chapter 6). Dr David Gaboriau provided direct assistance with microscopy and live tracking of basal cell migration through cell-derived matrices (CDMs) (Chapter 6). I prepared cell-derived extracellular matrix fractions from lung fibroblasts. Maryline Fresquet performed sample preparation for mass spectrometry and pre-processed the raw data at the Wellcome Trust Centre for Cell-Matrix Research in Manchester. I performed downstream analysis of the data and generated figures in collaboration with Maryline Fresquet and Professor Rachel Lennon (Chapter 6).

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

First and foremost, I would like to thank my supervisors Professor Clare Lloyd, Professor Toby Maher and Dr Adam Byrne for their continual inspiration, guidance and unwavering support throughout my PhD. I am immensely grateful to Professor Lloyd for the encouragement to pursue such ambitious research in her laboratory, and our hugely enjoyable and rewarding weekly supervisor meetings. I would also like to thank Professor Anne O'Garra for hosting me in her laboratory, imparting her expertise in transcriptomics and for her critical reading of Chapters 4 and 5.

I am indebted to friends and colleagues in the IRD Section at Imperial who have provided invaluable advice and assistance throughout my PhD. A special thanks goes to Dr Franz Puttur, whose approachability, training and mentorship have been instrumental to the success of the project. A further special mention goes to Dr David Gaboriau whose enthusiasm and microscopy skill brought Chapter 6 to life. My experience in the lab was transformed by the generosity of past and present laboratory members; Simone Walker, Dr Sarah Allden, Dr James Cook, Dr Patricia Ogger, Dr Rachele Invernizzi, Dr Dhiren Patel, Dr James Harker, Dr Maria Zarcone, Dr Lewis Entwistle, Poonam Ghai and Dr Lisa Gregory. I would also like to express gratitude to members of the O'Garra lab at the Francis Crick Institute, for welcoming me as a satellite member, Christine Graham for her calm supervision during preparation of the RNA sequencing libraries, and Probir Chakravarty for his single- cell bioinformatic wizardry and tutelage.

I would like to thank the ILD consultants at the Royal Brompton Hospital who provided expert clinical insight during patient recruitment and supervised bronchoscopy sessions. I am very grateful to Dr Philip Molyneaux and Dr Samuel Kemp for supervising the healthy volunteer bronchoscopies.

Further thanks to my collaborators whose interactions have been transformative to the project; Dr Fred Fercoq and Dr Leo Carlin (Cancer Research UK Beatson Institute, Glasgow, UK), Maryline Fresquet and Professor Rachel Lennon (Wellcome Trust Centre for Cell-Matrix Research, Manchester, UK), and Professor Martijn Nawijn (University of Groningen, Groningen, The Netherlands).

I am sincerely grateful for the generous funding provided by The Imperial College Clinician-Investigator Scholarship that enabled me to pursue this critically important research into IPF.

This project was only possible through the participation of IPF patients and healthy volunteers. I am beholden to each and every one for their contribution to this research and I dedicate this work to them.

Finally, I thank my friends and family, especially my wife Affy, for her tireless support, encouragement and understanding over the last three years. Without her help navigating the demands of the PhD with that of parenthood, it would have been impossible to complete the project and for that I am forever grateful.

Abstract

Idiopathic pulmonary fibrosis (IPF) is a progressive, scarring lung disease driven by alveolar epithelial cell injury, dysregulated fibroblast activity and excessive extracellular matrix (ECM) deposition. Through unknown mechanisms, basal cells, normally restricted to the airway epithelium, are expanded in remodeled fibrotic lung. The aim of this thesis was to understand how airway epithelial cell phenotype and function is shaped through interactions with the local immune and tissue microenvironment in IPF, and how this contributes to disease pathogenesis.

Functional characterization of primary bronchial epithelial cells from newly diagnosed IPF patients and healthy controls was performed *in vitro*. There were no significant differences in proliferation, wound healing capacity, barrier function or response to viral challenge. Human precision- cut lung slices demonstrated a close spatial relationship between airway epithelial cells and $CD206^{+}$ airway macrophages (AMs), therefore their molecular phenotype was explored with bulk RNA-sequencing. AMs purified from bronchoalveolar lavage displayed a unique transcriptomic profile with upregulation of genes involved in cell movement and matrix remodelling. Bulk RNA-seq of bronchial brushings exposed transcripts indicative of immune cells, therefore single-cell RNA-seq was utilised to illuminate cellular heterogeneity in the airway mucosa. This revealed expansion in AM subsets in IPF and increased interactions between epithelial secretory cells and AMs.

In distal fibrotic lung tissue, second harmonic generation imaging demonstrated regional differences in collagen organisation. A positive association was identified between the number of $KRT5^{+}$ basal cells and collagen density and fibre orientation. Basal cell migration was modulated by components of the ECM *in vitro*. Cell-derived matrices (CDMs) generated from IPF fibroblasts had a distinct matrisome determined by mass spectrometry- based proteomics and were more prohibitive to basal cell migration than control CDMs.

Collectively, these findings demonstrate that interactions between airway epithelial cells and the microenvironment in the airway and distal lung, are integral to IPF pathogenesis.

Contents

Declaration of originality **2**

Copyright declaration **3**

Acknowledgements **4**

Abstract **6**

Contents **7**

List of figures **12**

List of tables **17**

List of abbreviations **18**

1 Introduction **22**

1.1 Idiopathic pulmonary fibrosis (IPF) – a clinical overview **22**
1.1.1 Epidemiology 22
1.1.2 Clinical features 23
1.1.3 Diagnosis 23
1.1.4 Treatment 26

1.2 Pathogenesis of IPF **28**
1.2.1 Predisposing factors 28
1.2.1.1 Environmental exposures 28
1.2.1.2 Genetic risk 29
1.2.2 Alveolar epithelial injury and dysfunction 31
1.2.3 Aberrant wound repair by activated fibroblasts 36
1.2.4 Changes in the extracellular matrix (ECM) 37
1.2.5 Profibrotic macrophages 38
1.2.6 A role for airway epithelial cells in the pathogenesis of IPF 41
1.2.6.1 Contemporary views of the airway epithelial cell landscape 41
1.2.6.2 Airway epithelial cell abnormalities in IPF 45
1.2.7 Limitations of animal models of pulmonary fibrosis 47

1.3 Summary **48**

1.4 Hypothesis **49**

1.5 Aims **49**

2 Material and methods **51**

2.1 Subject recruitment **51**
2.1.1 IPF patients 51
2.1.2 Healthy controls 51

2.2 Human samples **52**
2.2.1 Flexible bronchoscopy with bronchoalveolar lavage (BAL) and bronchial brushings 52
2.2.2 Blood sampling 53

2.2.3 Fresh human lung tissue specimens 53
2.2.4 Archived formalin fixed paraffin embedded (FFPE) human lung tissue 54

2.3 Cell culture 54
2.3.1 Characteristics of primary bronchial epithelial cell (PBEC) donors 54
2.3.2 Isolation and culture of PBECs 56
2.3.3 Air-liquid interface (ALI) culture 56
2.3.4 Isolation and culture of primary human lung fibroblasts (HLFs) 57
2.3.5 Epithelial-fibroblast co-culture 57

2.4 Immunofluorescence staining and imaging 58
2.4.1 Cytospins of bronchial brushes 58
2.4.2 Basal cells in submerged culture 60
2.4.3 Basal cells on ECM proteins 60
2.4.4 Differentiated ALI cultures 61
2.4.5 Precision Cut Lung Slices (PCLS) 61
2.4.6 Cell-derived matrices (CDMs) from primary HLFs 62
2.4.7 Paraffin embedded lung tissue sections 63

2.5 Histology processing of ALI Transwell inserts 64

2.6 Transmission Electron Microscopy (TEM) 64

2.7 Functional assessment of PBECs in submerged culture 65
2.7.1 Proliferation assay 65
2.7.2 Wound closure assay 65

2.8 Functional assessment of PBECs in ALI culture 66
2.8.1 Transepithelial electrical resistance (TEER) 66
2.8.2 Ciliary beat frequency (CBF) 66

2.9 Influenza virus infection of differentiated epithelial cells 66
2.9.1 Influenza A virus infection 66
2.9.2 Lactate dehydrogenase (LDH) assay 67
2.9.3 LEGENDplex™ cytometric bead array 67

2.10 Flow cytometry of epithelial cells 68
2.10.1 Airway brushings 68
2.10.2 Epithelial cells in submerged culture 70

2.11 Airway macrophage enrichment from BAL for bulk RNA-sequencing 71
2.11.1 BAL processing 71
2.11.2 Cell counting 71
2.11.3 Fluorescence- activated cell sorting (FACS) 72
2.11.4 Magnetic-Activated Cell Sorting (MACS) 73
2.11.5 Flow cytometry to assess sample purity 74

2.12 RNA extraction and quality control 76

2.13 Wound healing PCR array 76

2.14 Bulk RNA-seq of human airway brush and airway macrophages 77
2.14.1 Preparation of cDNA libraries 77
2.14.2 Sequencing 78

2.15 Bulk RNA-seq data analysis 78
2.15.1 Quality control 78
2.15.2 Alignment 78
2.15.3 Quantification and normalisation 79
2.15.4 Filtering for low counts 79
2.15.5 Principal components analysis (PCA) 79
2.15.6 Hierarchical clustering 79

2.15.7 Differential expression analysis 80
2.15.8 Gene ontology (GO) Enrichment Analysis 80
2.15.9 Ingenuity Pathway Analysis (IPA) 80
2.15.10 Gene Set Enrichment Analysis (GSEA) 81
2.15.11 CIBERSORTx 82

2.16 Single-cell RNA-seq of human airway brushes 83
2.16.1 Isolation of single cells from human airway brushes 83
2.16.2 Chromium 10x Genomics library construction and sequencing 84

2.17 Single-cell RNA-seq data analysis 84
2.17.1 Pre-processing 84
2.17.2 Sample integration 85
2.17.3 Cluster annotation 86
2.17.4 Subclustering 86
2.17.5 Differential expression analysis 86
2.17.6 Trajectory inference using Monocle 3 86
2.17.7 Trajectory inference using Slingshot 87
2.17.8 Cell-cell interactions using CellPhoneDB 87

2.18 *MUC5B* genotyping assay 89

2.19 Second harmonic generation imaging and analysis 89
2.19.1 Image acquisition 89
2.19.2 Collagen texture analysis 89

2.20 Cell-derived matrices (CDM) from primary human lung fibroblasts 91

2.21 Basal cell migration experiments 91
2.21.1 Basal cell migration on ECM substrates 91
2.21.2 Basal cell migration through CDMs 92

2.22 Cell tracking analysis using Nikon NIS software 93
2.22.1 Tracking basal cell migration on ECM substrates 93
2.22.2 Tracking basal cell migration through CDM (3D) 94

2.23 MS-based proteomics of cell-derived matrices 95
2.23.1 Isolation of enriched cell-derived ECM and MS sample preparation 95
2.23.2 Western blotting 96
2.23.3 MS data acquisition 96
2.23.4 MS data analysis 97

2.24 Statistical analysis (excluding transcriptomic and proteomic data) 99

3 *In vitro* phenotyping and functional studies of the IPF airway epithelium 102

3.1 Background and rationale 102

3.2 Hypothesis 103

3.3 Aims 104

3.4 Results 104
3.4.1 Phenotyping airway epithelial cells from bronchial brushings using microscopy techniques 104
3.4.2 Culture characteristics of IPF airway epithelial cells 108
3.4.3 Epithelial wound healing assay 109
3.4.4 IPF PBEC air-liquid interface (ALI) culture characteristics 111
3.4.5 Influenza virus challenge model 113
3.4.6 Airway epithelial- fibroblast co-culture model 120

3.4.7 Airway epithelial-macrophage interactions 122
3.5 Discussion 123

4 Bulk transcriptomic profiling of the IPF airway epithelial niche 130

4.1 Background and rationale 130
4.2 Hypothesis 131
4.3 Aims 131
4.4 Results 131
4.4.1 Clinical characterisation of subjects recruited 131
4.4.2 Sample collection and processing for RNA-seq 134
4.4.3 Quality control of RNA 136
4.4.4 Quality control of cDNA libraries and RNA-seq data 137
4.4.5 Gene filtering 140
4.4.6 Transcriptomic analysis of airway brushings 141
4.4.6.1 Highly abundant genes 141
4.4.6.2 Dimensional reduction 142
4.4.6.3 Clustering 143
4.4.6.4 Differential expression analysis 145
4.4.6.5 Immunostaining human samples for KRT13 149
4.4.6.6 Immune cell contaminants 150
4.4.7 Transcriptomic analysis of airway macrophages (AMs) 151
4.4.7.1 Highly abundant genes 151
4.4.7.2 Deconvolution using CIBERSORTx 152
4.4.7.3 Dimensional reduction 153
4.4.7.4 Clustering 154
4.4.7.5 Differentially expressed genes between IPF and healthy controls 157
4.4.7.6 Ingenuity Pathway Analysis (IPA) 160
4.4.7.7 Gene set enrichment analysis (GSEA) 167
4.4.7.8 Comparison with published datasets 169
4.4.7.9 Comparison of differentially expressed genes in paired airway brushings and sorted AMs 170
4.5 Discussion 171
4.5.1 Airway brushings 171
4.5.2 Airway macrophages 174

5 Single-cell profiling of the IPF airway epithelial niche 179

5.1 Background and rationale 179
5.2 Hypothesis 180
5.3 Aims 181
5.4 Results 181
5.4.1 Clinical characterisation of subjects recruited 181
5.4.2 Sample quality control 184
5.4.3 Data quality control 186
5.4.4 Cluster analysis of integrated dataset 190
5.4.5 Cluster analysis by condition 201
5.4.6 Differential expression analysis 209
5.4.7 Basal cell subclustering analysis 212
5.4.8 Trajectory analysis of integrated dataset 225
5.4.9 Cell- cell interactions in the IPF airway 230

5.5 Discussion 237

6 Basal cell - matrix interactions in IPF 247

6.1 Background 247

6.2 Hypothesis 248

6.3 Aims 248

6.4 Results 249

6.4.1 KRT5+ basal cells populate parenchymal lung tissue and co-localise with collagen-rich fibrotic matrix IPF 249

6.4.2 Regional collagen organisation in the healthy and fibrotic lung 251

6.4.3 Extracellular matrix proteins collagen I and collagen III promote increased motility of healthy and IPF airway basal cells. 259

6.4.4 IPF and healthy basal cells exhibit similar migratory capacity through fibrotic cell-derived matrix 266

6.4.5 Changes to matrix microenvironment induce differential expression of wound healing genes in basal cells 269

6.4.6 Basal cell migration is restricted through IPF cell-derived matrix 273

6.4.7 MS-based proteomics of cell-derived matrices 276

6.4.8 Secreted Protein Acidic and Cysteine Rich (SPARC) modulates basal cell migration in vitro 286

6.5 Discussion 288

6.6 Conclusion 295

7 General discussion & future work 298

7.1 Airway epithelial immune homeostasis is disturbed at a molecular level in IPF 299

7.2 Airway macrophage populations are expanded in IPF and exhibit a unique transcriptomic signature 302

7.3 Cues from the fibrotic extracellular matrix dictate basal cell activity in the distal lung 305

7.4 Strengths of the research 307

7.5 Limitations of the research 307

7.6 Future work 308

7.7 Concluding remarks 308

References 310

Appendix 343

Conference abstracts & presentations arising from this thesis 348

List of figures

Figures	Page
Figure 1.1 The classification system of interstitial lung diseases	22
Figure 1.2 Radiological pattern of usual interstitial pneumonia (UIP)	24
Figure 1.3 Histopathological features of UIP	25
Figure 1.4 Current model of IPF pathogenesis	35
Figure 1.5 Contemporary map of the airway epithelial landscape captured by scRNA-Seq	42
Figure 1.6 Summary of PhD hypothesis	49
Figure 2.1 Subject recruitment and sample utilisation for experimental studies	53
Figure 2.2 Gating strategy for epithelial cells using 16-HBE cell line	70
Figure 2.3 Gating strategy for airway macrophages	75
Figure 2.4 Bulk RNA-seq analysis pipeline	83
Figure 2.5 Single-cell RNA-seq pipeline	88
Figure 2.6 Calculation of straightness ratio	94
Figure 2.7 Proteomics workflow for human lung fibroblast- derived matrices	98
Figure 3.1 Transmission electron microscopy (TEM) images of PBECs from a patient with fibrotic lung disease	106
Figure 3.2 Immunofluorescence microscopy of IPF bronchial brush cytospin preparations	107
Figure 3.3 PBEC culture characteristics	108
Figure 3.4 PBEC proliferation assay	109
Figure 3.5 PBEC wound healing assay using JuLI^{TM} Stage	110
Figure 3.6 PBEC ALI transepithelial electrical resistance (TEER)	111
Figure 3.7 Characterisation of IPF PBEC ALI cultures	112
Figure 3.8 Influenza virus challenge model	113
Figure 3.9 Influenza virus infection of IPF and healthy PBEC ALI cultures	114
Figure 3.10 Inflammatory mediators produced by PBECs at 72 hours post-influenza A infection	116
Figure 3.11 Wound Healing PCR Array	118
Figure 3.12 Epithelial - fibroblast interactions following IAV infection	119
Figure 3.13 Transmission electron microscopy of PBEC-HLF co-culture system	121
Figure 3.14 PCLS from human distal lung tissue	122

Figure 4.1 Bulk transcriptomic profiling of airway brushings and sorted airway macrophages from IPF patients and healthy controls	132
Figure 4.2 Sorting airway macrophages (AM) from human bronchoalveolar lavage fluid	135
Figure 4.3 Quality control of RNA samples used for sequencing	136
Figure 4.4 Quality control metrics of raw paired-end sequencing libraries	138
Figure 4.5 Alignment statistics for sequencing files	139
Figure 4.6 Gene filtering by raw read count	140
Figure 4.7 Venn diagram showing overlapping genes between the top 500 most abundant genes in airway brushings from IPF patients and healthy controls	141
Figure 4.8 Principal component analysis (PCA) of bulk RNA-seq of airway brushings	142
Figure 4.9 Unsupervised hierarchical clustering of the most variable genes in airway brushings	144
Figure 4.10 Volcano plot showing differentially expressed genes between IPF and healthy airway brushings	145
Figure 4.11 Hierarchical clustering of differentially expressed genes between IPF and healthy airway brushings	147
Figure 4.12 Correlation between key upregulated chemokines and IPF composite physiological index (CPI)	148
Figure 4.13 KRT13-expressing cells on airway brushings and tissue sections	149
Figure 4.14 Immune cell deconvolution of bulk-RNA seq of airway brushings using CIBERSORTx	150
Figure 4.15 Venn diagram showing overlapping genes between the top 500 most abundant genes in sorted AMs from IPF patients and healthy controls	151
Figure 4.16 Immune cell deconvolution of bulk-RNA seq of sorted AMs using CIBERSORTx	152
Figure 4.17 Principal component analysis (PCA) of bulk RNA-seq of AMs	153
Figure 4.18 Unsupervised hierarchical clustering of the most variable genes in CD206+ AMs	155
Figure 4.19 Gene ontology (GO) enrichment analysis of selected gene modules in hierarchical clustering of the 5% most variant genes	156
Figure 4.20 Volcano plot showing differentially expressed genes between IPF and healthy sorted AMs	157
Figure 4.21 Hierarchical clustering of differentially expressed genes between IPF and healthy sorted AMs	159
Figure 4.22 Most significant IPA canonical pathways in 393 genes upregulated in IPF AMs compared to healthy controls	161

Figure 4.23 Network for top cytokine upstream regulator, *CSF2*	164
Figure 4.24 Gene interaction network: 'Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking'	165
Figure 4.25 Gene interaction network: 'Cancer, Connective Tissue Disorders, Organismal Injury and Abnormalities.'	165
Figure 4.26 Gene interaction network: 'Connective Tissue Disorders, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders.'	166
Figure 4.27 Most significant IPA canonical pathways in 173 genes downregulated in IPF AMs compared to healthy controls	167
Figure 4.28 Gene Set Enrichment Analysis (GSEA) using curated pulmonary fibrosis gene set	168
Figure 4.29 Violin plot showing expression of macrophage subset marker genes	169
Figure 4.30 Venn diagram showing overlapping upregulated genes in IPF airway brushings and airway macrophages	170
Figure 5.1 Single-cell RNA sequencing to study heterogeneous cell populations in airway brushings	182
Figure 5.2 Estimation of data dimensionality	187
Figure 5.3 Quality control metrics for scRNA-seq of airway brushings by cluster	188
Figure 5.4 Integrated UMAP of single-cell data from all healthy and IPF subjects showing predicted doublets	189
Figure 5.5 Clustering tree of integrated dataset	190
Figure 5.6 Top 5 marker genes identifying each cluster in the integrated dataset	192
Figure 5.7 Heatmap of top 5 marker gene expression for each cell in a cluster for the integrated dataset	193
Figure 5.8 Gene set enrichment analysis (GSEA) using cell type specific marker genes from published datasets	196
Figure 5.9 Dot plot of macrophage/ monocyte marker genes from Morse et al	197
Figure 5.10 Integrated UMAP visualisation of airway cells from all subjects	198
Figure 5.11 Gene Ontology (GO) enrichment analysis for selected epithelial cell clusters	199
Figure 5.12 Gene Ontology (GO) enrichment analysis for macrophage clusters	200
Figure 5.13 UMAP visualisation of airway cells by condition	201
Figure 5.14 Total proportion of each cell type identified in airway brushings by condition	202
Figure 5.15 UMAP visualisation of airway cells per individual subject	203

Figure 5.16 Quantification of epithelial cell types in IPF and healthy airways per subject	205
Figure 5.17 Quantification of immune cell types in IPF and healthy airways per subject	206
Figure 5.18 Correlation analysis between macrophage cluster percentage and the extent of disease (composite physiological index)	207
Figure 5.19 KRT13 and KRT4 expressing cells visualised on integrated UMAP plots	208
Figure 5.20 Differentially expressed genes between clusters in IPF and healthy control	209
Figure 5.21 Gene Ontology analysis of clusters with greatest number of DE genes	211
Figure 5.22 Basal cell subclusters	213
Figure 5.23 Heatmap showing expression of top 10 marker genes for each cell in the basal cell subclusters	214
Figure 5.24 UMAP visualisation of basal cell subclusters per individual subject	216
Figure 5.25 Immune cell contamination in basal cell subclustering	217
Figure 5.26 Filtered basal cell subclusters	219
Figure 5.27 Heatmap showing expression of top 10 marker genes for each cell in the filtered basal cell subclusters	220
Figure 5.28 Quantification of basal cell subclusters in IPF and healthy airways	221
Figure 5.29 Dot plot of the expression of genes related to the Notch pathway in each of the basal cell subclusters	225
Figure 5.30 Trajectory analysis of airway epithelial cells using Monocle 3	227
Figure 5.31 Trajectory analysis of airway epithelial cells using Slingshot	229
Figure 5.32 Heatmaps displaying the number of interactions between all clusters in healthy and IPF samples	231
Figure 5.33 Network displaying interacting cell clusters in healthy and IPF samples	232
Figure 5.34 Specific receptor – ligand interactions between secretory cells (cluster 7) and proliferating macrophages (cluster 12)	234
Figure 5.35 Specific receptor – ligand interactions between macrophages (cluster 6) and secretory cells (cluster 7)	235
Figure 5.36 Specific receptor – ligand interactions between macrophages (cluster 6) and proliferating macrophages (cluster 12)	236
Figure 5.37 Summary of key cell-cell interactions in the IPF airway mucosa predicted from scRNA-seq data	243
Figure 6.1 Immunofluorescence microscopy of KRT5+ basal cells in distal lung	250
Figure 6.2 Evaluation of regional collagen organisation in IPF and control lung using second harmonic generation (SHG) imaging	252

Figure 6.3 Quantitative analysis of SHG images of distal lung tissue from controls and IPF patients	254
Figure 6.4 Principal component analysis of quantitative image analysis of IPF and control distal lung	256
Figure 6.5 KRT5+ basal cell distribution correlates with regional collagen organisation	257
Figure 6.6 KRT5+ basal cells co-localise with fibroblasts in the fibrotic niche	258
Figure 6.7 Human airway basal cell phenotyping	260
Figure 6.8 Extracellular matrix proteins modulate migratory dynamics of healthy airway basal cells	262
Figure 6.9 Extracellular matrix proteins modulate migratory dynamics of IPF airway basal cells	264
Figure 6.10 Generation of CDMs from primary human lung fibroblasts	267
Figure 6.11 Healthy and IPF airway basal cell migration through IPF CDM	268
Figure 6.12 Extracellular matrix microenvironment drives gene expression changes in basal cells	271
Figure 6.13 Airway basal cell migration through control and IPF CDMs	274
Figure 6.14 Interactions between basal cells and fibronectin in IPF CDMs	275
Figure 6.15 Isolation and MS analysis of enriched lung fibroblast-derived ECM	277
Figure 6.16 Unsupervised clustering of 154 matrisome proteins identified in lung fibroblast generated CDM by MS	278
Figure 6.17 IPF and control lung fibroblasts have differential ECM composition	280
Figure 6.18 SPARC restricts migration of healthy basal cells at higher concentrations	287
Figure 7.1 Standard curve generated for each analyte in the LEGENDplex™ assay	343

List of tables

Table	Page
Table 1.1 Airway epithelial cell marker genes used for scRNA-seq studies	44
Table 2.1 Characteristics of PBEC donors for functional in vitro studies in Chapter 3	55
Table 2.2 List of antibodies used for confocal microscopy	59
Table 2.3 Antibodies used for PBEC flow cytometry	69
Table 2.4 Antibodies used for airway macrophage sorting	73
Table 2.5 Collagen analysis measurements	90
Table 4.1 Clinical characteristics of recruited subjects for bulk RNA-seq	133
Table 4.2 Top canonical pathways for genes upregulated in IPF AMs	162
Table 4.3 Top 5 activated upstream regulators in IPF	164
Table 4.4 Top canonical pathways for genes downregulated in IPF AMs	167
Table 4.5 Genes associated with pulmonary fibrosis upregulated in IPF AMs	168
Table 5.1 Clinical characteristics of subjects recruited for scRNA-seq	183
Table 5.2 Sample characteristics for scRNA-seq	185
Table 5.3 Predicted singlets and doublets in the integrated dataset of all healthy and IPF subjects	189
Table 5.4 Cluster cell type identity	195
Table 5.5 GO enrichment analysis for basal cell subclusters	223
Table 6.1 Proteins identified in human lung fibroblast generated CDMs	281
Table 7.1 Wound healing gene assayed by QIAGEN RT^2 Profiler PCR Array	344
Table 7.2 Lung tissue samples used in Section 6.4.1	346
Table 7.3 PBECs used in Section 6.4.3	346
Table 7.4 Primary cells used in Section 6.4.4	347
Table 7.5 PBECs used in Section 6.4.5	347
Table 7.6 Primary cells used in Section 6.4.6	347

List of abbreviations

α-SMA	Alpha- smooth muscle actin
AF	Alexa fluor
ALI	Air- liquid interface
AM	Airway macrophage
ANOVA	Analysis of variance
APC	Allophycocyanin
ASM	Angular second moment
AT1	Alveolar epithelial type 1 cells
AT2	Alveolar epithelial type 2 cells
BAL	Bronchoalveolar lavage
BV	Brilliant violet
CBF	Ciliary beat frequency
CDM	Cell-derived matrix
cDNA	Complimentary DNA
CHP	Chronic hypersensitivity pneumonitis
COPD	Chronic obstructive pulmonary disease
CPI	Composite physiologic index
Cy7	Cyanine7
DAPI	4′,6-diamidino-2-phenylindole
DE	Differentially expressed
DLCO	Diffusing capacity for carbon monoxide
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EPCAM	Epithelial cell adhesion molecule
FACS	Fluorescence- activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
FEV1	Forced expiratory volume in 1 second

FFPE	Formalin fixed paraffin embedded
FITC	Fluorescein isothiocyanate
FMO	Fluorescence minus one
FSC	Forward scatter
FVC	Forced vital capacity
GLCM	Grey-level co-occurrence matrix
GO	Gene Ontology
GOI	Gene of interest
GSEA	Gene set enrichment analysis
HBSS	Hank's Balanced Salt Solution
HKG	Housekeeping gene
HLF	Human lung fibroblasts
HRCT	High-resolution computed tomography
IAV	Influenza A virus
IDM	Inverse difference moment
ILD	Interstitial lung disease
IPA	Ingenuity pathway analysis
IPF	Idiopathic pulmonary fibrosis
KRT5	Cytokeratin 5
LDH	Lactate dehydrogenase
MACS	Magnetic- activated cell sorting
MEM	Minimum Essential Media
MHC	Major histocompatibility complex
MMP	Matrix metalloproteinase
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MSD	Mean squared displacement
MUC5B	Mucin 5B
PBEC	Primary bronchial epithelial cell
PCA	Principle component analysis
PCLS	Precision-cut lung slice

PCR	Polymerase chain reaction
PE	Phycoerythrin
PFA	Paraformaldehyde
PFT	Pulmonary function tests
PIC-Seq	Physically interacting cells RNA Sequencing
RBH	Royal Brompton Hospital
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
ROI	Region of interest
RT-PCR	Reverse transcription polymerase chain reaction
scRNA-Seq	Single-cell RNA sequencing
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SHG	Second harmonic generation
SSC	Side scatter
TEAB	Triethylammonium bicarbonate
TEER	Transepithelial electrical resistance
TEM	Transmission electron microscopy
TGF-β	Transforming growth factor beta
TP63	Transformation-related protein 63
UIP	Usual interstitial pneumonia
UMAP	Uniform Manifold Approximation and Projection
UV	Ultraviolet

CHAPTER 1

Introduction

1 Introduction

1.1 Idiopathic pulmonary fibrosis (IPF) – a clinical overview

IPF is a progressive and incurable fibrotic lung disease[1,2]. It is the most common idiopathic interstitial pneumonia (IIP), belonging to a diverse group of interstitial lung diseases (ILDs) (Fig. 1.1) which are characterised by scarring of the lung parenchyma[3,4]. The aetiology of IPF remains unclear, although it is postulated that genetic predisposition, various environmental insults and cellular ageing contribute to the development of the disease[2]. Evidence suggests repetitive injury and activation of a susceptible alveolar epithelium provokes a dysregulated wound repair response by fibroblasts with excessive extracellular matrix (ECM) deposition[5,6]. Relentless and irreversible fibrosis causes destruction of normal alveolar architecture, remodelling and impaired gas exchange leading to respiratory failure and ultimately death.

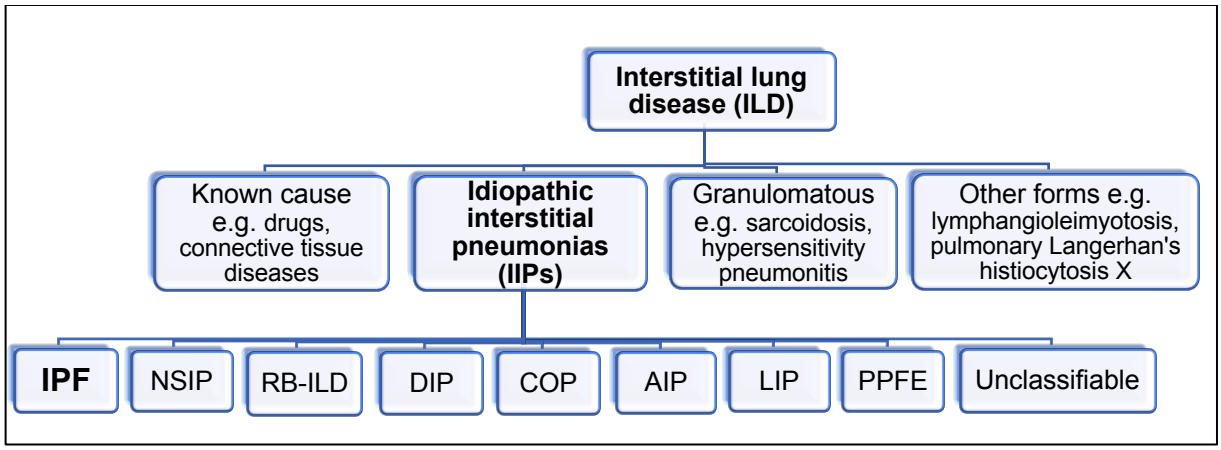


Figure 1.1 The classification system of interstitial lung diseases. Adapted from American Thoracic Society/ European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias (2002, 2013). (NSIP – non-specific interstitial pneumonia, RB-ILD – respiratory bronchiolitis interstitial lung disease, DIP – desquamative interstitial pneumonia, COP – cryptogenic organising pneumonia, AIP – acute interstitial pneumonia, LIP – lymphoid interstitial pneumonia, PPFE – pleuroparenchymal fibroelastosis).

1.1.1 Epidemiology

Epidemiological estimates of the burden of IPF from large population database studies, vary according to the methodology used to identify and define cases[7,8]. Studies suggest the incidence of IPF is increasing[9,10], and there are an estimated 3 - 9 new cases per 100,000 per year in Europe and North America[11]. Each year in the UK there are approximately 6,000 new cases diagnosed[12], and an estimated 5,500

deaths from IPF[13]. Notably the incidence of IPF is higher in males[14], and increases with age with most diagnoses occurring over 60 years of age[15,16].

1.1.2 Clinical features

IPF typically manifests in the sixth or seventh decade with insidious onset exertional dyspnoea and a proportion of patients experience a dry cough[1,15]. Dyspnoea is progressive and debilitating. The presence of cough was reported to independently predict disease progression[17]. Fine, bibasal, inspiratory crackles are classically detected on chest auscultation, and digital clubbing may also be present[1]. As the disease advances, pulmonary function testing demonstrates restrictive lung physiology with reduced forced vital capacity (FVC), and impaired gas exchange with low diffusing capacity for carbon monoxide (DLco). An average decline in FVC of 150 – 200mls per year was demonstrated in placebo-treated IPF patients with mild to moderate lung function impairment at baseline from clinical studies[18]. Measurements for DLco, FVC and FEV1 can be used to calculate the composite physiological index (CPI), which has been shown to correlate with disease extent on CT and prognosis[19]. CPI is particularly valuable if lung function indices are confounded by coexisting emphysema[19]. Oxygen desaturation to 88% or less during a 6-minute walk test has been associated with an increased risk of death in those with biopsy-proven IPF[20]. Disease trajectory varies between patients and is difficult to predict; it may progress slowly, rapidly or be relatively stable but punctuated with periods of acute decline[21].

Untreated, IPF carries a prognosis in line with many types of cancer, with a median survival time of 3 years[16,22]. The cause of death is commonly progression of the disease itself[23]. Gender (G), age (A) and pulmonary physiology (P) indices (FVC, DLco) can be used in a multidimensional staging tool to calculate the ‘GAP’ index, to aid prognostication in IPF patients[24].

1.1.3 Diagnosis

The clinical manifestations of IPF are nonspecific and in one study many patients reported that diagnosis took more than 1 year[25]. Evidence-based international guidelines[1] recommend firstly excluding alternative causes of ILD such as connective tissue diseases, occupational exposures and chronic hypersensitivity pneumonitis

(CHP). This is achieved by a detailed history, clinical examination and serological testing. Morell et al. discovered that 43% of patients diagnosed with IPF according to 2011 international guidelines, had a subsequent diagnosis of CHP upon rigorous re-assessment with specific serum IgG, bronchoalveolar lavage (BAL) cellular analysis and bronchial challenge testing[26].

The identification of a usual interstitial pneumonia (UIP) pattern on high-resolution computed tomography (HRCT) scan (Fig. 1.2, A and B) is highly supportive of a diagnosis of IPF but can also be associated with other conditions. Radiologically, UIP is characterised by honeycombing, reticulation and traction bronchiectasis in a subpleural and basal distribution[1]. The extent to which these typical features are present on HRCT is assessed and in the appropriate clinical context, a confident diagnosis of IPF can be reached on the basis of radiographic findings alone[27]. If all criteria for a UIP pattern are not fulfilled, BAL cellular analysis (to exclude alternative causes) and/or a lung biopsy may be necessary for a conclusive diagnosis.

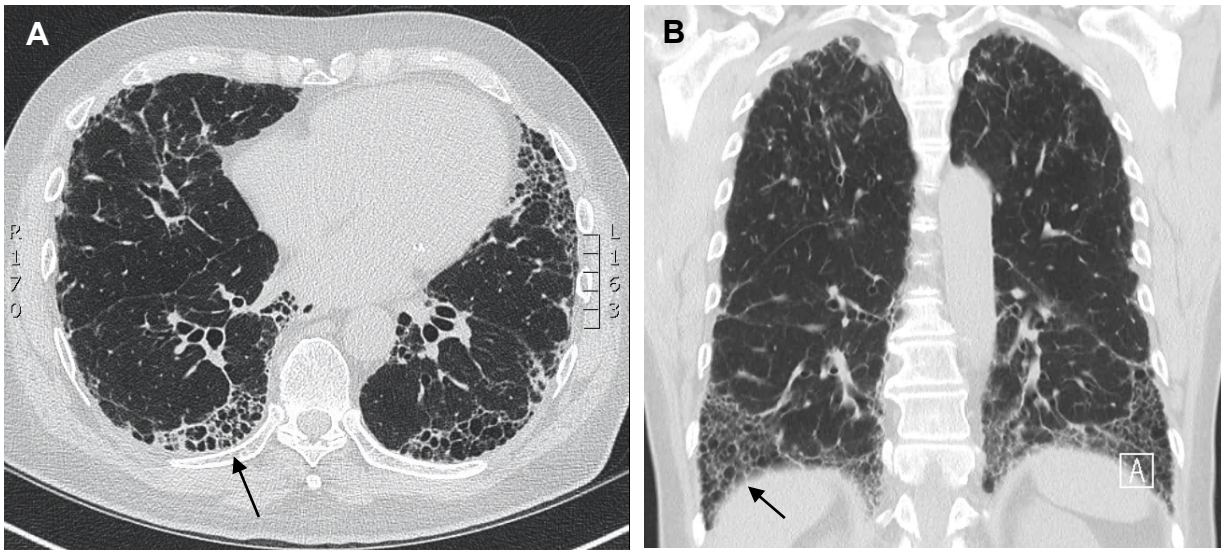


Figure 1.2 Radiological pattern of usual interstitial pneumonia (UIP). High-resolution CT (HRCT) images from a 62- year old patient with IPF demonstrating a usual interstitial pneumonia (UIP) pattern of fibrosis on **(A)** axial and (**B**) coronal sections. There is subpleural reticulation, traction bronchiectasis and lower lobe predominant, peripheral honeycombing (arrows).

BAL is performed by wedging a flexible bronchoscope within a selected bronchopulmonary segment, then instilling and withdrawing aliquots of normal saline to a total volume of up to 300mls[28]. For an accurate BAL differential cell count, at least

30% of the total volume instilled should be retrieved. The BAL immune cell profile can assist in the diagnosis of alternative ILDs which can mimic IPF on HRCT such as CHP, which is associated with a BAL lymphocytosis[28,29]. Acquisition of BAL has proved invaluable for research studies investigating macrophages[30] and the microbiome[31] in IPF.

In selected patients, a surgical lung biopsy may be obtained. A defining histopathological feature of UIP is 'temporal heterogeneity' where areas of dense interstitial fibrosis and architecturally distorted lung are juxtaposed with normal lung parenchyma[32]. UIP has a predilection for the subpleural and paraseptal regions[1]. At the interface between dense fibrosis and normal lung are active 'fibroblastic foci' which are aggregates of proliferating fibroblasts and myofibroblasts (Fig. 1.3 A). Three-dimensional imaging has shown these are complex and independent structures which are thought to occur at discrete sites of injury[33]. Collagens I, III, IV, V and VI, fibronectin, hyaluronan and versican have been identified at the myofibroblast core of fibroblastic foci[34]. Hyperplastic AT2 cells and reactive large, flattened epithelial cells are often seen overlying fibroblastic foci (Fig. 1.3 A)[35]. Marked remodelling of the parenchyma occurs with the development of mucin-filled "honeycomb cysts" which are curiously lined by ciliated bronchiolar epithelium[32,36] (Fig. 1.2 B).

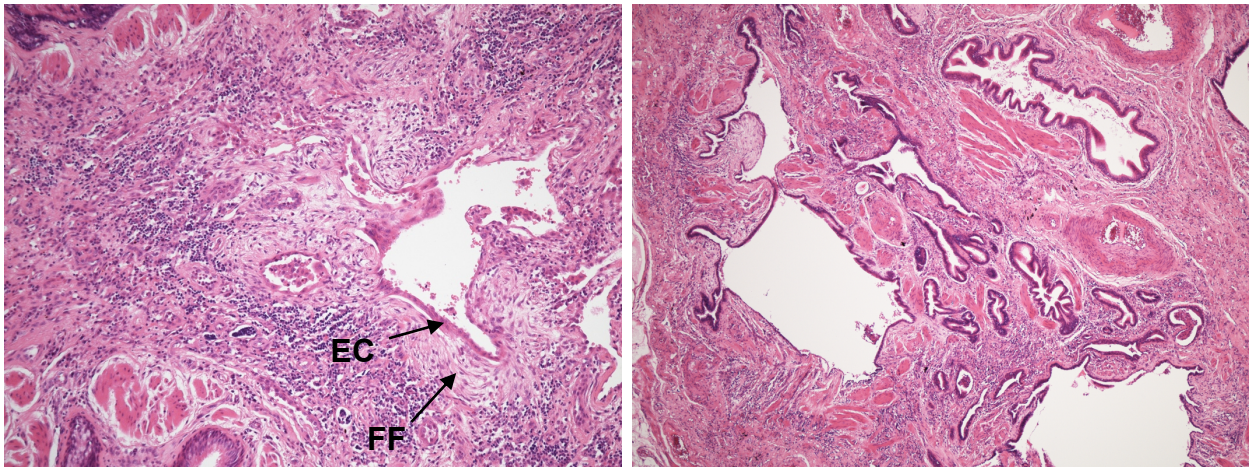


Figure 1.3 Histopathological features of UIP. Haematoxylin and eosin (H&E) stained lung tissue sections taken from IPF explant lung showing (**A**) fibroblastic foci (FF) covered with hyperplastic epithelial cells (magnification x100), and (**D**) microscopic honeycomb change (magnification x40). Images provided by Dr Alexandra Rice (Consultant Histopathologist, Royal Brompton Hospital, UK).

Clinical, radiological and histopathology data are integrated, and the accuracy and confidence of an IPF diagnosis is improved by a dynamic multidisciplinary discussion (MDD) between clinicians, chest radiologists and pathologists[37,38].

1.1.4 Treatment

PANTHER-IPF, a pivotal clinical trial published in 2012, demonstrated conventional treatment of IPF with a combination of prednisolone (a corticosteroid), azathioprine (an immunosuppressive) and *N*- acetylcysteine (NAC) increased the risk of hospitalisation and death compared with placebo[39]. This trial led to a paradigm shift in our therapeutic approach to IPF; a move away from immunosuppressive treatments towards novel antifibrotic agents, which were conditionally recommended in the 2015 international guidelines[40]. Furthermore the findings of this study support the fundamental hypothesis that IPF is driven by an aberrant wound healing response rather than inflammation[41].

Pirfenidone and nintedanib are novel antifibrotic drugs approved for use in UK and countries worldwide. The mechanism of action of pirfenidone has not yet been fully elucidated, but it is an orally active, synthetic, small-molecule consisting of a modified phenyl pyridone[42]. It is recognised to have anti-fibrotic, anti-inflammatory and anti-oxidant properties[42]. *In vitro* pirfenidone reduced primary human lung fibroblast proliferation and attenuated profibrotic transforming growth factor (TGF)- β signalling pathways[43]. The CAPACITY studies 004 and 006[44] were multinational, randomised controlled studies which compared pirfenidone to placebo in IPF patients with mild to moderate physiological impairment. Pirfenidone resulted in a significant reduction in FVC decline compared to placebo in study 004 but not in study 006. In view of these discordant results, a further multinational randomized controlled study - ASCEND[45] - was undertaken. This showed that after 52 weeks of treatment with pirfenidone the rate of decline in FVC was reduced by almost 50% compared to placebo. This effect of pirfenidone on disease progression translated into a 48% reduction in all- cause mortality at 1 year compared to placebo, in a pooled analysis of the ASCEND and CAPACITY studies[45]. Nintedanib is a small molecule inhibitor of the receptor tyrosine kinases fibroblast growth factor receptors (FGFRs), platelet derived growth factor

receptors (PDGFRs) and vascular endothelial growth factor receptors (VEGFRs)[46]. By blocking autophosphorylation of these receptors and downstream signalling pathways, nintedanib reduces fibroblast proliferation, migration and transformation to myofibroblasts[46]. Two replicate, multinational randomized controlled trials - IMPULSIS-1 and IMPULSIS-2 - showed nintedanib reduced the annual rate of FVC decline by approximately 50%, compared to placebo over 52 weeks[47]. Adverse events such as gastrointestinal upset and a photosensitive rash with pirfenidone, and diarrhoea with nintedanib limit tolerability and led to drug discontinuation in 20.9% and 26.3% respectively[48].

Whilst antifibrotic drugs slow disease progression they do not completely arrest or reverse pulmonary fibrosis, so there is an urgent need to better understand the early pathobiological events driving the development of the disease[49]. Lung transplantation is an option for some IPF patients, but is associated with a number of challenges given the frequently occurring comorbidities in this group such as pulmonary hypertension and coronary artery disease[50]. The 5-year survival post-transplantation is 47% – 53%, however a large proportion of patients die on the waiting list[51]. Acquisition of diseased parenchymal lung tissue from IPF patients undergoing transplant through biobanks, has been fundamental to research studies and has provided novel insight into disease pathogenesis[52,53].

1.2 Pathogenesis of IPF

1.2.1 Predisposing factors

The current pathogenic model of IPF involves a complex interaction between genetic susceptibility, environmental exposures and ageing[54].

1.2.1.1 Environmental exposures

The epithelial surfaces of the lung are constantly challenged by various inhaled environmental and occupational exposures that have been implicated in the development of IPF[55]. Cigarette smoking has been associated with an increased risk of both sporadic[56] and familial IPF[57]. Although not proof of causal links, meta-analyses of case- controlled studies discovered significant associations between metal dust, wood dust, farming or agricultural exposure, and IPF[56,58]. A study by Baumgartner et al. reported a significant association between metal dust and IPF that was only evident when the duration of exposure was greater than 5 years[59]. Air pollution may also play a role, with associations found between ozone (O_3) and nitrogen dioxide (NO_2) exposure and risk of acute exacerbations[60], and particulate matter of aerodynamic diameter 10mm (PM_{10}) and 2.5mm ($PM_{2.5}$) and mortality in IPF[61]. The high prevalence of gastro-oesophageal reflux (GER) in IPF[62], led to the hypothesis that, through microaspiration into the airways, GER may contribute to disease pathogenesis[63]. There have been conflicting results on the effect of antacid treatment on disease progression in IPF[64,65], and an underpowered, randomized controlled trial failed to demonstrate a reduction in FVC decline with anti-reflux surgery[66]. Interestingly, the presence of a hiatus hernia appears to be associated with disease progression in IPF patients taking an anti-fibrotic[67].

There is considerable interest in the role of an infective trigger in IPF[68]. A meta-analysis of 20 retrospective case- control studies suggested that viral infection with Epstein-Barr (EBV), cytomegalovirus (CMV), HHV-7 or HHV-8 was associated with an increased risk of IPF[69]. Methodological differences between the studies included in this analysis limit the ability to draw conclusions on the relationship between viruses and IPF. Interestingly, in at-risk first-degree relatives of familial interstitial pneumonia (FIP) patients, increased EBV and CMV DNA was detected in BAL[70]. Additionally, herpesvirus antigens and markers of endoplasmic reticulum (ER) stress were

expressed in AT2 cells, suggesting human herpes viruses may be cofactors in early disease pathogenesis. In contrast, Yin et al. used RNA-seq to explore virus gene expression in 28 IPF and 20 control lung tissue specimens and did not find any association between viruses and IPF[71].

High-throughput, molecular sequencing of the bacterial 16S ribosomal RNA gene in BAL samples has revealed the human airways are colonised with a community of microbes termed the 'lung microbiome', that differs compositionally in health and disease[72]. A study comparing the microbiome in IPF, COPD and healthy BAL samples, revealed the presence of distinct microbial communities, and a significantly higher bacterial burden in IPF which was associated with disease progression[31]. A response to the presence of these bacterial communities is evident in the host peripheral blood transcriptome[73]. It is unclear whether alterations to the microbiome play a causative role in IPF pathogenesis or occur by consequence of the immune and structural perturbations that take place during the disease[74]. A strategy attempting to modulate the microbiome using the broad-spectrum antibiotic, co-trimoxazole, failed to show an improvement in IPF outcomes compared to placebo in the EME-TIPAC randomised controlled[75].

1.2.1.2 Genetic risk

Studies investigating the genetic basis of familial IPF (also known as familial interstitial pneumonia) have discovered rare genetic variants (minor allele frequency <0.1%) in telomere-related pathways[76,77], and surfactant protein processing[78,79]. Telomeres are repetitive sequences which protect the ends of chromosomes from shortening during cell replication. Shortened telomeres were associated with poorer survival in IPF[80]. Whole- exome sequencing demonstrated variants in telomere- related genes (*TERT, RTEL1* and *PARN)* that have been implicated in familial disease, were also associated with an increased risk of sporadic IPF[81].

Findings from large genome wide association studies (GWAS) in IPF have paved the way to advancements in our understanding of the biological processes that drive IPF[82]. Collectively these studies have identified common variants (minor allele frequency >5%) in 17 genetic loci associated with sporadic IPF, which are linked to host- defence

(e.g. *MUC5B, TOLLIP*), epithelial cell- integrity (e.g. *DDP9, DSP*), and telomere maintenance (e.g. *TERC, TERT, RTEL1, PARN*)[83-85]. Together, these variants account for approximately one-third of the risk of developing familial and sporadic IPF[82]. The largest GWAS study in IPF published to date (2,668 IPF cases and 8591 controls) confirmed associations at 11 previously reported loci[86]. It also discovered further signals of association implicating mTOR signalling (*DEPTOR*) and mitotic spindle-assembly genes (*MAD1L1, KIF15).*

The *MUC5B* gain-of-function promoter variant rs35705950 is the strongest genetic risk factor for IPF[87]. *MUC5B* encodes a mucin glycoprotein which is the major constituent of airway mucous secreted by airway epithelial cells[88]. MUC5B and MUC5AC are the most abundant gel-forming mucins in the airways[89]. Eliminating inhaled microbes and particulate matter through mucociliary clearance, MUC5B plays a vital function in mucosal immunity. Mice deficient in Muc5b exhibited chronic lower airway bacterial infection, alveolar macrophage apoptosis and impaired phagocytosis, which was not found in mice deficient in Muc5ac[88]. Carriage of the *MUC5B* rs35705950 minor T allele increases the risk of both familial and sporadic IPF[90]. Intriguingly, subjects with the rs35705950 variant actually have an improved survival[91]. Immunohistochemical staining showed MUC5B was localized to airway secretory cells and honeycomb cysts in IPF[90]. The rs35705950 variant (G > T) has been associated with increased *MUC5B* promoter activity and MUC5B protein expression in bronchiolar epithelial cells in the IPF lung[92]. Tissue overexpression of MUC5B protein in carriers of the T allele is specific to IPF, and not observed in non-specific interstitial pneumonitis (NSIP) or control lung tissue[93]. In mice, overexpression of *Muc5b* in the airways and alveoli was associated with mucociliary dysfunction and more pronounced fibrotic response to bleomycin[94]. Interestingly *MUC5B* genotype has been associated with subclinical interstitial lung abnormalities (ILAs) on CT scan, with an increased risk in carriers of the minor T allele[95]. ILAs are associated with a higher rate of death secondary to pulmonary fibrosis[96], suggesting these may be a precursor to IPF in some cases. Importantly, Hunninghake et al. also showed that the *MUC5B* rs35705950 minor T allele is carried by 19% of individuals with no ILAs or definite fibrosis, suggesting other complex gene- gene and gene- environment interactions are at play to determine the risk of IPF[87].

Lending further support to the role of the epithelium in IPF, was the novel association found between SNP rs62025270 near *AKAP13* and IPF susceptibility[85]. A-kinase anchoring protein 13 (AKAP13) regulates RhoA, a molecule involved in profibrotic signaling pathways. Carriage of the minor A allele of rs62025270 was associated with increased expression of *AKAP13* mRNA in IPF lung tissue and immunohistochemistry revealed *AKAP13* to be localized to the alveolar epithelium and lymphoid follicles in IPF lung tissue.

Taken together, there are genetic risk factors that contribute to IPF predisposition and a complex interaction with environmental agents which ultimately determine inter-individual variability in disease behaviour[97]. The gene variants implicated suggest impaired mucosal host defense and increased cell senescence drive a susceptibility to environmental exposures[98]. This leads to repeated microinjuries at multiple sites in the distal airspace, which drive a fibrotic response.

1.2.2 Alveolar epithelial injury and dysfunction

The lung has a vast surface area for gas exchange of approximately $70m^2$ [99]. Thin alveolar type 1 (AT1) cells cover 95% of this surface, closely apposed to pulmonary capillary endothelial cells (PCECs) to facilitate efficient gas exchange[100]. Cuboidal alveolar type 2 cells (AT2) secrete pulmonary surfactant to reduce alveolar surface tension and prevent collapse during respiration. Genetic lineage-tracing experiments in mice have demonstrated AT2 cells are the principal stem cell population responsible for alveolar repair and regeneration of AT1 cells following injury[101]. A molecularly distinct subpopulation of AT2 cells termed alveolar epithelial progenitor (AEP) cells which express *Axin2* (a target of Wnt signalling), play a major role in the regeneration of the alveolar epithelium during lung injury[102]. Other epithelial progenitor populations contributing to distal lung regeneration have been described using lineage-tracing in mice, including bronchioalveolar stem cells (BASCs)[103] and cytokeratin 5 (Krt5)-expressing stem cells[104,105]. Signals from adjacent fibroblasts termed ‘mesenchymal alveolar niche cells’, expressing *Axin2* and *Pdgfr*α, support the self-renewal and differentiation of AT2 cells to AT1 cells[106].

It is widely accepted that injury to the alveolar epithelium is a key antecedent to fibrogenesis in IPF[35,41]. This was first suggested by electron microscopy studies of IPF lung tissue which showed areas of alveolar epithelial cell denudation[107]. Alveolar epithelial cells displaying features of apoptosis were found adjacent to collagen-rich fibroblastic foci in the fibrotic lung[108]. A key feature of IPF is the striking disturbance of alveolar re-epithelialisation[35]. For example, areas of hyperplastic epithelial change with squamous metaplasia were discovered in honeycomb cysts[109]. Immunohistochemical analysis demonstrated co-localisation of endoplasmic reticulum (ER) stress markers (ATF-6, ATF-4 and CHOP), with cleaved caspase-3 indicative of apoptosis in AT2 cells from IPF patients but not controls[110]. Targeted injury to AT2 cells with diphtheria toxin in a transgenic mouse model resulted in a significant increase in lung collagen content and fibrosis[111]. In mice, disruption of alveolar epithelial integrity by genetic deletion of tetraspanin (CD151) - which is essential for cell adhesion to the basement membrane - resulted in spontaneous age-related pulmonary fibrosis[112]. Fewer CD151-positive AT2 cells were found on histological evaluation of IPF lung tissue compared to controls, indicating loss of this molecule is relevant to the pathogenesis of IPF.

In view of the striking association of IPF with age, mechanisms including telomere attrition, cellular senescence and mitochondrial dysfunction have been linked to the development of the disease[113]. Senescence is a state of irreversible cell cycle arrest induced by multiple stressors including telomere attrition and DNA damage[114]. Senescent cells secrete factors known as the senescence-associated secretory phenotype (SASP) which can act in a paracrine manner to promote tissue dysfunction. Single-cell RNA-seq (scRNA-seq) has revealed significantly elevated expression of core senescence markers and related pathways in AT2 cells from IPF lung tissue compared to AT2 from controls[115]. Yao et al. generated a novel murine model of AT2 cell senescence by conditional knock-out of *Sin3a* in AT2 cells, which promoted progressive fibrosis with features comparable to human IPF[115]. Other work has suggested the miR-34 family of miRNA may contribute to increased senescence in AT2 cells in IPF[116].

Markers of telomere dysfunction and senescence were localised to AT2 cells in UIP pattern lung fibrosis[117]. Telomere dysfunction in AT2 cells results in cellular

senescence and limited regenerative capacity[118]. Tamoxifen-induced, conditional deletion of telomere repeat binding factor 1 (TRF1) in murine AT2 cells resulted in telomere shortening, spontaneous lung fibrosis and architectural distortion after 8 months of weekly injections[119]. It was associated with a higher proportion of senescent AT2 cells determined by staining for senescence associated β galactosidase (SA-β-gal).

Cell-type specific mitochondrial dysregulation has been identified in the fibrotic lung[120]. Electron microscopy has revealed enlarged, dysmorphic mitochondria with disrupted cristae in IPF AT2 cells compared to controls[121]. In highly fibrotic areas, mitochondrial changes were associated with BiP, a molecular chaperone and marker of ER stress. PTEN-induced putative kinase 1 (PINK1), regulates mitochondrial integrity and mitophagy, a process of selectively degrading damaged mitochondria[121]. Lower levels of *PINK1* transcript have been found in IPF AT2 cells, and mice deficient in PINK1 exhibit dysfunctional mitochondria and susceptibility to fibrosis.

In addition to alveolar epithelial cell injury, there is evidence of AT2 stem-cell dysfunction in IPF; IPF AT2 cells displayed reduced organoid colony forming efficiency *in vitro,* suggesting lower renewal capacity[122]. The extracellular matrix glycosaminoglycan, hyaluronan (HA), is expressed at lower levels on the surface of IPF AT2 cells and has been implicated in mechanisms governing their renewal. In mice, targeted genetic deletion of the enzyme responsible for hyaluronan (HA) synthesis, HA synthase 2 (*Has2*), or toll-like receptor 4 (*Tlr4)* in AT2 cells, led to reduced self-renewal and more severe fibrotic response to bleomycin[122].

The peripheral distribution of scarring in the IPF lungs led to the hypothesis that mechanical factors may catalyse pathogenesis. Cdc42, a member of the RhoGTPase family, is essential for regulating cytoskeleton reorganisation and cell shape change[123]. When *Cdc42* was genetically deleted in mice, AT2 cells could not differentiate into AT1 cells leading to alveoli overdistension, exposure to increased mechanical tension and progressive fibrosis[123]. Mechanical tension activated TGF-β signalling in AT2 cells in *Cdc42*-null mice.

IPF AT2 cells produce a plethora of growth factors, matrix metalloproteinases (MMPs) and chemokines which maintain the fibrotic milieu of the distal lung[124,125]. For example, MMP7 is highly expressed in fibrotic lungs and localised by immunohistochemistry to the epithelium[126]. MMPs degrade all components of the ECM but also influence the activity of matrix associated proteins including growth factors and other bioactive mediators which in turn modify cellular functions[127]. Immunohistochemistry studies of IPF lung tissue have demonstrated hyperplastic AT2 cells and bronchiolar epithelial cells synthesise transforming growth factor (TGF) – β1, a potent profibrotic mediator[128]. TGF-β promotes fibroblast proliferation and induces the secretion of ECM proteins by human lung fibroblasts[129]. Some of the histopathological features of human IPF were recapitulated in rat explant lung slices where TGF-β1 was overexpressed in AT2 cells using a retroviral gene transfection approach[130]. The integrin αvβ6, which converts latent TGF-β into its active form has also been detected in epithelial cells overlying fibrotic areas of the IPF lung[131].

In IPF lung tissue, AT2 cells closely associated with regions of matrix remodelling stained positive for N-cadherin and had lost apical-basal polarity, suggesting acquisition of a mesenchymal-like phenotype[103]. Co-localisation of AT2 cell markers with mesenchymal marker alpha-smooth muscle actin (α-SMA) was also demonstrated in the cystic airspaces of IPF lung tissue[132].

Dysregulation of developmental programmes may explain the aberrant behaviour of AT2 cells in IPF. There is evidence of increased activation of the canonical Wingless / integrase-1 (WNT) / β-catenin signalling pathway in IPF epithelial cells in areas of fibrotic remodelling[133]. WNT signalling is essential for normal lung development, but alterations in this pathway may contribute to the pathogenesis of IPF and other chronic lung diseases such as chronic obstructive pulmonary disease (COPD)[134,135]. Sonic hedgehog (SHH) is a morphogen responsible for lung patterning and branching morphogenesis[136]. Components of the Sonic signalling pathway are upregulated in IPF lungs and were localised in an apical distribution in bronchiolar and alveolar epithelial cells in IPF but not in normal lungs[136]. Additionally, fibroblast proliferation and collagen synthesis were increased with recombinant SHH.

Collectively these studies suggest disturbance of alveolar epithelial cell homeostasis is a critical factor driving IPF (Fig. 1.4). Cell-intrinsic and environmental factors contribute to AT2 dysfunction which is a key initiating event in the fibrotic cascade[137].

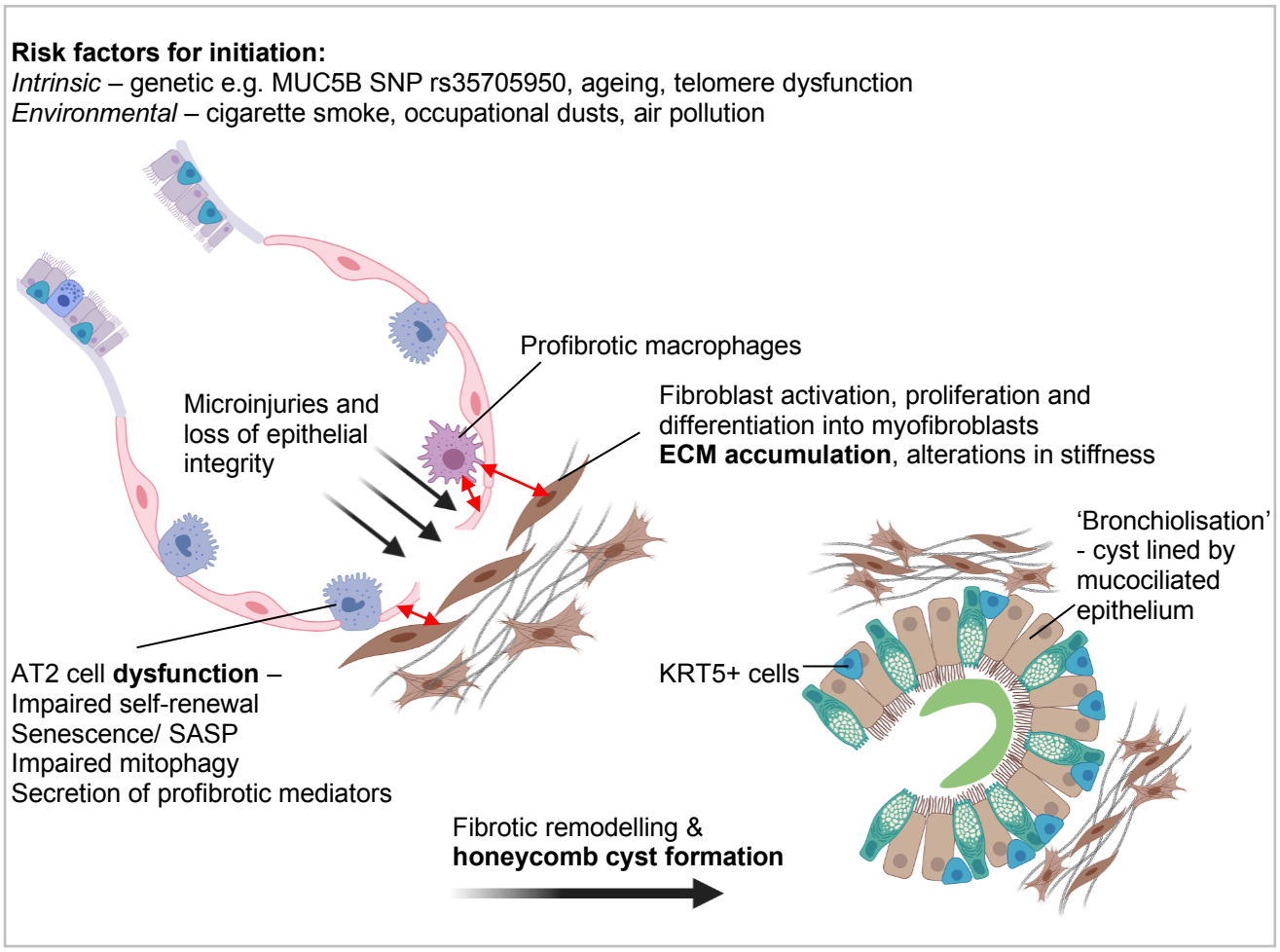


Figure 1.4 Current model of IPF pathogenesis. Alveolar epithelial dysfunction initiates a profibrotic cascade resulting in the relentless deposition of ECM by myofibroblasts. Pathological destruction of the lung parenchyma and remodelling occurs. Cytokeratin 5 (KRT5)-expressing progenitor cells populate areas of fibrotic remodelling and line honeycomb cysts, generating a mucociliated epithelium. Crosstalk (red arrows) between cell types drives pathogenesis. Figure created with BioRender.com.

1.2.3 Aberrant wound repair by activated fibroblasts

Fibroblasts are stromal cells which are instrumental in tissue repair and regeneration[138]. During injury they migrate into the wound area, proliferate and differentiate into myofibroblasts which have contractile properties and secrete collagen and other ECM components[138]. Differentiated myofibroblasts are characterised by the presence of stress fibres and α-smooth muscle actin (α-SMA). Their overriding goal in health is to facilitate wound closure and provide a matrix scaffold for organ structure. In conditions such as IPF, myofibroblasts persist and catalyse pathological remodelling of the ECM through dysregulated matrix turnover[139]. This results in the irreparable destruction of tissue architecture and loss of lung function resulting in respiratory failure.

Controversy exists over the origin of the expanded, pathogenic population of fibroblasts in IPF[139]. A study by Rock et al. suggested they arise from heterogenous, resident stromal cell populations rather than through the process of epithelial-mesenchymal transition (EMT)[140]. During EMT, epithelial cells acquire mesenchymal characteristics and behaviours, often in response to environmental signals[141]. Rather than contributing directly to the myofibroblast pool, ZEB1- induced EMT was shown to promote fibrosis through paracrine signalling to fibroblasts[142]. Fibrocytes are bone-marrow derived progenitors, identifiable through expression of markers including CD45, CD34 and CXCR4 and may also contribute to the pool of fibroblasts[143].

Functionally heterogenous populations of fibroblasts have been reported in the context of rheumatoid arthritis, an immune- mediated inflammatory disease affecting joints[144]. ScRNA-seq has revealed expansion of four transcriptionally distinct populations of fibroblasts residing in different anatomical regions of the lung in IPF[145]. Myofibroblasts expressing *ACTA2* were located around airways in the subepithelial region, *HAS1*^{hi} fibroblasts were found in subpleural regions, *PLIN2*+ lipofibroblast-like cells and *LUM*+ fibroblasts within the interstitial regions[145]. In a single-cell transcriptomics study by Tsukui et al., clustering of collagen- producing (*COL1A1+*), lineage negative (*CD31, CD45, CD235A, EPCAM-)* mesenchymal cells*,* identified a unique population of fibroblasts expressing *CTHRC1* (collagen triple helix repeat containing 1) in the fibrotic

lung[146]. This cluster of cells expressed high levels of ECM-related genes and were spatially resolved to the fibroblastic foci.

In normal tissue repair, myofibroblasts undergo apoptosis following wound closure[147]. Fibroblast survival plays a role in perpetuating fibrogenesis. IPF lung fibroblasts were shown to be resistant to FasL- induced apoptosis compared to control fibroblasts[148]. *In vitro* studies suggest lower levels of prostaglandin (PG)E_2, a major lung prostanoid, may lead to the disparity in cellular survival within the same microenvironment; with an increase in epithelial apoptosis and decrease in fibroblast apoptosis[148]. A relative deficiency of (PG)E_2 has been reported in the epithelial lining fluid of the fibrotic lung[149]. Signalling through p53 MAPK and PI3K/AKT, endothelin-1 and TGF-β have been shown to induce apoptosis resistance in normal primary human fetal lung fibroblasts (IMR-90)[150].

Polymerised collagen suppresses the proliferation of normal lung fibroblasts[151]. In contrast, IPF fibroblasts cultured on polymerised collagen undergo robust proliferation mediated by β1-integrin signalling, low levels of the tumour suppressor, phosphate and tensin homolog (PTEN) and increased phosphoionositide 3-kinase (PI3K)- Akt-S6K1 activity[151]. An aberrant PTEN/Akt axis and low levels of autophagy were shown to desensitise IPF fibroblasts to polymerised type 1 collagen-induced cell death[152].

1.2.4 Changes in the extracellular matrix (ECM)

The extracellular matrix (ECM) deposited by fibroblasts is far from inert. The 'core matrisome' (all the proteins in a given matrix) consists of approximately 300 proteins including collagens, proteoglycans and glycoproteins[153]. Additionally, there are 'matrisome-associated' proteins including ECM-bound growth and secreted factors, affiliated proteins and regulators[153,154]. Alterations to ECM composition and biomechanics in pulmonary fibrosis can directly modulate the phenotype and behaviour of resident cells[155]. IPF-derived ECMs enhanced fibroblast adhesion and proliferation compared to control ECM[156]. This effect was mediated by enhanced matrix cross-linking because inhibition of cross-linking enzymes transglutaminase 2 (TG2) and the lysyl oxidase (LOX) family during ECM generation, altered ECM structure and attenuated fibroblast adhesion and proliferation[156].

Fibroblasts cultured in decellularised lung matrices from IPF patients adopt a myofibroblast phenotype[157]. Similarly, the expression of myofibroblast differentiation marker α-SMA was increased in normal human lung fibroblasts cultured on stiff polyacrylamide gels designed to recapitulate the rigidity of the fibrotic lung[158]. Matrix stiffening changed fibroblast actin dynamics resulting in actin polymerisation and nuclear translocation of megakaryoblastic leukaemia factor-1 (MLK1), which ultimately led to an increase in α-SMA promoter activity[158]. This study provided mechanistic evidence linking mechanotransduction to the process of myofibroblast differentiation.

The type of ECM (fibrotic vs. control) has been shown to exert a greater influence on fibroblast gene expression than the effect of fibroblast origin (IPF vs. control)[159]. Using decellularized lung ECM from IPF patients and controls, Parker et al. discovered that ECM- driven gene expression changes were predominantly regulated at the level of translation[159]. ECM genes were activated by the fibrotic ECM suggesting a pathological positive feedback loop exists and this was catalysed by a reduction in expression of miR-29c, a microRNA which normally represses translation of these genes.

1.2.5 Profibrotic macrophages

Broadly, there are two distinct populations of macrophages in the human lung; airway macrophages (AMs) residing in the airway lumen and interstitial macrophages (IMs) in the parenchyma[160]. In reality, the situation is more complex and diverse subpopulations have been identified in studies of the murine[161,162], and human lung[163,164]. Moreover, the anatomical niche plays a critical role in shaping macrophage phenotype and function[165].

Macrophages are the most abundant immune cells in the airspace, comprising over 80% of the BAL fluid cell count[28,166], and detectable in sputum samples[167] and airway brushings[168]. Airway macrophages (AM) are strategically positioned in the airspace to carry out vital immune surveillance, respond to invading pathogens, process surfactant, and clear apoptotic cells and debris[169]. They are highly adaptive to their

microenvironment[170], and are regulated through interactions with the local epithelium[171].

The ontogeny of resident AMs and their relative contribution to lung disease compared with recruited monocyte-derived AMs has been a source of debate. Murine fate-mapping and parabiosis models indicated tissue-resident AMs in the lung derived from fetal monocytes then were self-maintained locally in the lung during homeostatic conditions[172,173]. Byrne et al. challenged this concept demonstrating that following human lung transplant, the majority of tissue-resident AMs in the donor lung were derived from the circulating monocyte pool of the recipient[174].

A shift from homeostatic, resident AMs to recruited monocyte-derived AMs has been observed in chronic respiratory diseases such as cystic fibrosis[167]. Genetic lineage tracing in murine models of pulmonary fibrosis revealed monocyte-derived, rather than tissue-resident AMs were crucial to the development of fibrosis and persisted in the lung for one year[162]. AT2 cell injury in a transgenic mouse model resulted in CCR2 mediated accumulation of non-resident $CD11b^{+}$ macrophages and $Ly6C^{hi}$ monocytes resulting in lung fibrosis[175]. In murine models of pulmonary fibrosis, liposomal chlodronate depletion of $Ly6C^{hi}$ circulating monocytes ameliorated fibrosis and reduced lung macrophage populations and markers of alternatively activated macrophages[176]. Traditionally, macrophages have been characterised as M1 (classically activated), producing pro-inflammatory cytokines directed against intracellular pathogens, or M2 (alternatively activated) associated with type 2 immune responses, parasite elimination and tissue repair[177]. This dichotomisation is not useful because macrophage activation state is dynamic, and a phenotypic spectrum is likely to exist[160].

Macrophage phenotype heterogeneity may give clues to disease pathology. For example, in IPF BAL there was a significant increase in the proportion of AMs lacking expression of transferrin receptor 1, CD71, which mediates the uptake of diferric transferrin[30]. $CD71^{-}$ AMs had impaired phagocytosis, expressed higher levels of monocyte marker CD14 and upregulated profibrotic genes compared to $CD71^{+}$ AMs. IPF patients with higher levels of $CD71^{-}$ AMs have a poorer prognosis[30].

Single-cell transcriptomic studies have exposed macrophage heterogeneity in distal lung tissue explants from controls and advanced IPF cases, describing three or four AM subpopulations[53,163,164]. Expansion of a profibrotic AM population expressing *SPP1,* encoding osteopontin, has been described in these studies, and is localised to the more fibrotic regions in the lower lobes[164]. Osteopontin is a secreted glycoprotein shown to drive proliferation and migration of human monocyte-derived macrophages[178].

The role of AM populations in the early, evolutionary phases of pulmonary fibrosis are yet to be investigated. Bronchoscopy studies focusing on the airway epithelial-immune environment, suggest there are populations of monocyte-derived and resident AMs in the healthy airway[179,180]. Situated with epithelial cells at an interface with the external environmental, AMs critically ensure appropriate immune tolerance to innocuous inhaled stimuli, and robust immune response to harmful pathogens[169]. Crosstalk mediated by connexin 43-containing gap junctions between resident AMs and alveolar epithelial cells modulated the immune response to inhaled LPS in a mouse model[181]. Computational approaches applied to scRNA-seq data have unveiled the complex epithelial-immune communication occurring in the human airspace[168,182]. There is a need to dissect epithelial-macrophage interactions in the airway that may, when dysregulated, promote a profibrotic milieu in IPF.

Distinct macrophage subsets orchestrate normal tissue repair[183]. In a murine model of skin wound repair, CD301b-expressing macrophages were shown to interact with and promote the proliferation of a subpopulation of myofibroblasts[184]. In three-dimensional (3D) fibrillar collagen gels, bone-marrow derived macrophages were attracted towards contractile myofibroblasts by the physical deformations they generated in the ECM[185]. This elegant work also demonstrated that detection of these mechanical signals by macrophages was achieved through stetch-sensitive membrane ion channels. This mechanism may also be relevant in IPF.

Adherens junction protein cadherin-11 (CDH11) has been identified between activated CD68-positive macrophages and myofibroblasts within fibroblastic foci in the lungs of IPF patients[186].The intimate contact achieved by CDH11- mediated adhesion, allowed latent TGF-β produced by primary mouse macrophages to be activated by lung

myofibroblasts, thus maintaining a profibrotic niche[186]. Interstitial macrophages are therefore intimately co-located and acting in concert with the key effector cells of fibrosis; myofibroblasts.

1.2.6 A role for airway epithelial cells in the pathogenesis of IPF

1.2.6.1 Contemporary views of the airway epithelial cell landscape

An expanding body of evidence suggests airway epithelial cell populations are intimately involved in the pathobiology of IPF. Immunohistochemistry studies of tissue sections have demonstrated bronchiolar abnormalities and the remarkable appearance of airway-like cells amidst severe parenchymal remodelling and cystic regions in IPF[36,187,188]. This has been termed 'bronchiolisation'. Numerous single-cell transcriptomic studies have corroborated this finding at a molecular level, emphasising a marked phenotypic shift from alveolar to airway epithelial cell populations in the distal lung[53,145,189].

Illuminating data from single-cell transcriptomic studies have transformed our views of the airway epithelial cell landscape in health and disease[52]. Early electron microscopy studies provided detailed morphological characterisation of the major epithelial populations in the human airway[190,191]. Throughout the respiratory tract there are specialised populations of epithelial cells with diverse functional roles[192] (Fig. 1.5). A principal function of the conducting airways is to filter the air we breathe, removing particulate matter and airborne antigens by mucociliary clearance[89]. Ciliated and secretory cell types line the luminal surface of the airways to perform this vital function. Secretory cells (including goblet and club cells) and submucosal glands, produce antimicrobial mediators and airway mucins which form the first layer of mucosal host defence[193]. Basal cells, lying beneath the surface epithelium attached to the basement membrane, account for one- third of all airway epithelial cells. They are the progenitor cells of the airway, capable of differentiating into other epithelial cell types during homeostatic turnover and are pivotal to repair and regeneration after injury[194]. The process of airway basal cell differentiation is governed by Notch signalling[195].

Immunohistochemistry and ultrastructural studies identified the existence of two basal cell subpopulations in the normal human airway, defined by the presence or absence of cytokeratin 14 (KRT14), hemidesmosomes and anchoring fibrils[196]. *In vivo* lineage-tracing studies in mice showed at homeostasis, there were two discrete KRT5+ basal cell subpopulations in the tracheal epithelium; basal stem cells (BSCs) and luminal precursors (BLPs)[197]. BLPs were derived from BSCs and estimated to comprise ~50% of the basal cell population. They were located in a basal position and persisted for 11 days prior to differentiating into luminal cell types[197]. These subpopulations were not distinguished by differential KRT14 expression. Single-cell transcriptomic studies have demonstrated basal cell diversity throughout the human airways. Depending on tissue sampling location within the tracheobronchial tree, three[198] or four[199] distinct basal cell subpopulations have been described on the basis of unique transcriptomic signatures. A population of 'suprabasal cells' characterised by higher expression of *KRT19* and *NOTCH3* was also reported in the healthy airway[179].

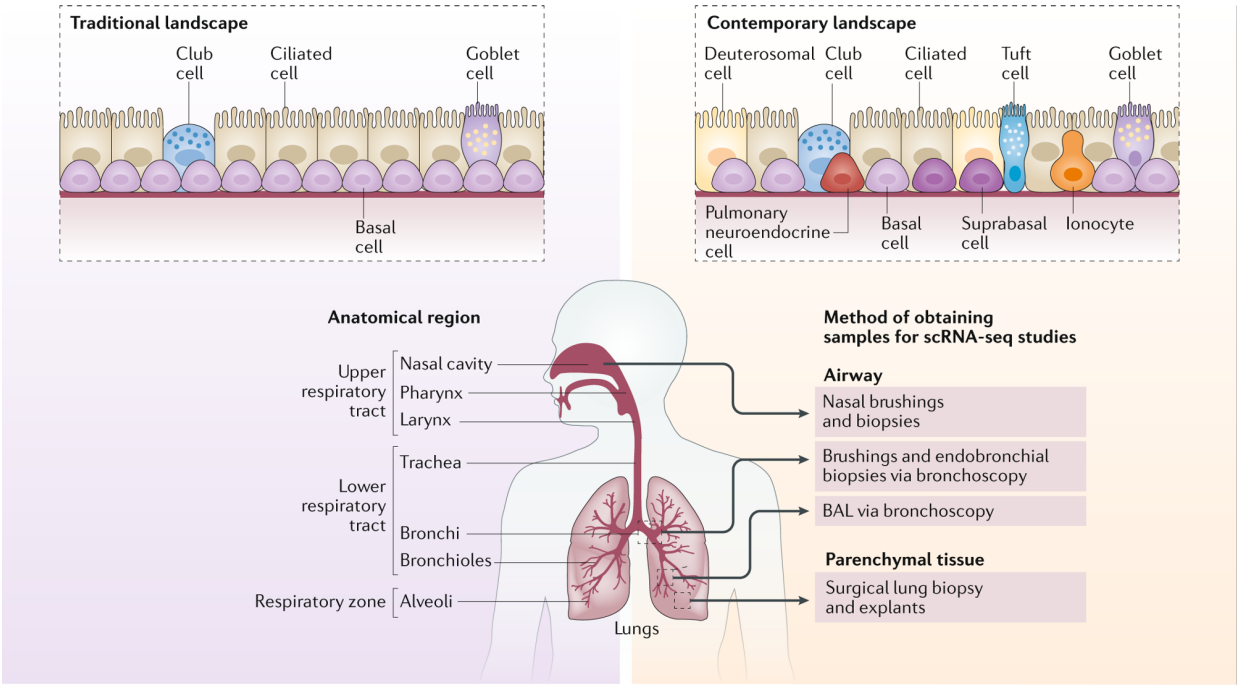


Figure 1.5 Contemporary map of the airway epithelial landscape captured by scRNA-Seq. Single-cell transcriptomic technology has enriched our view of the epithelial landscape, exposing rare cell types such as the pulmonary ionocyte, and novel cell states. The cellular composition defined by scRNA-seq is determined by the sampling site and approach. Image reproduced with permission of Springer Nature. From 'Hewitt, R.J., Lloyd, C.M. Regulation of immune responses by the airway epithelial cell landscape. *Nat Rev Immunol* (2021)'[52]

Profiling the transcriptome of individual cells, scRNA-seq technology has unveiled novel and rare airway epithelial cell types[200,201], and identified unique molecular cell states in health and disease[168]. The sentinel discovery of the cystic fibrosis transmembrane conductance regulator (CFTR)- expressing 'pulmonary ionocyte', and its relationship to established airway epithelial lineages was achieved through scRNA-seq[200,201]. Using an unsupervised computational approach, cells with similar gene expression profiles are grouped into clusters and assigned a biologically meaningful identity (annotated) based on consensus marker genes (Table 1.1)[202]. The starting material – biopsies, airway brushings or BAL – and anatomical location, are critical determinants of the cell populations identified using this technology. IPF scRNA-seq studies have explored the cellular composition of distal parenchymal lung tissue from explants, identifying epithelial, immune and stromal cell populations[145]. IPF BAL and airway brushings obtained via bronchoscopy have not yet been studied by scRNA-seq. Single-cell studies of the human airway using bronchoscopy samples (bronchial brushings and biopsies) demonstrated variations in the relative proportions and phenotype of epithelial and immune cells along the respiratory tract[168,179]. Deprez et al. showed the sampling method had an impact on the populations identified; brushings collected more luminal cell types (ciliated and secretory cells) and endobronchial biopsies acquired cells throughout the airway wall including from the stromal compartment and submucosal glands[179].

To study airway epithelial differentiation in human samples where lineage-tracing is not possible, computational tools have been developed to perform trajectory analysis using scRNA-seq data[203]. This bioinformatic approach leverages the transcriptional changes that occur during cellular differentiation and orders cells along their respective trajectories quantified by 'pseudotime'[203]. Pseudotime analysis has been used to unravel the dynamic relationship between cell types in the airway epithelium in health and disease[168]. When applied to primary nasal airway epithelial cells differentiating *in vitro*[204], and freshly harvested airway samples[179], it showed a transitory *FOXJ1+MUC5AC+* cell type lying in a trajectory between goblet and ciliated cells. These so-called 'mucous ciliated' cells were markedly increased in asthma but have not been studied in other respiratory diseases[168].

Table 1.1 Airway epithelial cell marker genes used for scRNA-seq studies. Reproduced with permission of Springer Nature. From 'Hewitt, R.J., Lloyd, C.M. Regulation of immune responses by the airway epithelial cell landscape. *Nat Rev Immunol* (2021)'[52]

Cell type	Selected human marker genes	Proposed function
Basal cells	*KRT5, TP63, KRT14, PDPN, NGFR, LGALS1, ITGA6, ITGB4, LAMA3, LAMB3, KRT15, S100A2, NPPC, BCAM, DST*	Main airway stem cells — regeneration, repair of epithelial lining
Suprabasal cells	*KRT5, TP63 low, KRT19, NOTCH3*	Intermediate between basal and club cells
Secretory — club and goblet cells	*SCGB1A1* (club cells), *MUC5AC* (goblet cells), *MUC5B, TFF3, SCGB3A1, BPIFB1, MSMB, SLPI, WFDC2*	Secretion of mucus, antimicrobial and anti-inflammatory peptides
Deuterosomal cells	*DEUP1/CCDC67, FOXJ1, FOXN4, CDC20B*	Precursor of ciliated cell characterized by genes critical to centriole amplification
Ciliated cells	*FOXJ1, PIFO, TPPP3, SNTN, FAM183A, LRRIQ1, DNAH12, C20orf85, SNTN, CAPS, TUBB4B, DNAH5, TSPAN1*	Clearance of mucus and debris
Mucous-ciliated cells	*FOXJ1, MUC5AC*	Differentiation intermediate between goblet and ciliated cells in homeostatic and diseased airways
Rare cell types		
Ionocyte	*FOXI1, CFTR, ASCL3, RARRES2*	Ion transport, fluid and pH regulation
Pulmonary neuroendocrine cells	*ASCL1, CHGA, CGRP, INSM1, HOXB5*	Environmental sensor — signals to CNS
Tuft/brush cells	*DCLK1, RGS13, POU2F3, ASCL2, LRMP,*	Immune and taste sensor, leukotriene synthesis, links to type 2 immunity
'Hillock' cells	*KRT13, KRT4*	Transitional cell type between basal and secretory; located in stratified, non-ciliated structures with high cell turnover; postulated to play a role in squamous barrier function and immunomodulation

1.2.6.2 Airway epithelial cell abnormalities in IPF

Studies have shown evidence of basal cell hyperplasia and atypia in the IPF distal lung[188]. Basal cells expressing transcription factor tumour protein p63 (TP63) - which are normally absent from the alveolar regions of the lung - extend into the alveolar space and line honeycomb cysts in IPF[188]. Multi-colour immunofluorescence analysis demonstrated subpopulations of $KRT5^{+}KRT14^{+}$ and $KRT5^{+}KRT14^{-}$ basal cells widely distributed in the IPF distal lung in varying configurations[205]. Supporting the concept that specific basal cell subpopulations may be involved in the disease process, single-cell analysis identified expansion of a 'secretory primed basal' (SPB) cell subset in fibrotic lung tissue[199]. This basal cell subset, identifiable by expression of CD66, was enriched in honeycomb cysts and expressed genes involved in mucin O-glycan biosynthesis.

The origin of the expanded population of KRT5+ basal cells in IPF is debated. Parallels have been drawn with murine models of H1N1 influenza- induced alveolar epithelial injury, in which clusters of $p63^{+}Krt5^{+}$ cells appear in damaged lung parenchyma[206]. There is general agreement from murine lineage-tracing studies that most of these $Krt5^{+}$ cells arise from a Krt5-negative progenitor source[104,207]. Given the species-related differences in cellular composition in the distal airways, it is unclear how murine lineage-tracing studies to ascertain cell-of-origin will translate to human disease[208]. Using time-lapse microscopy of murine lung slices, Vaughan et al. elegantly demonstrated migration of $Krt5^{+}$ cells derived from lineage-negative epithelial progenitors, into the alveoli following injury[104]. Interestingly, the traced Krt5 cells formed abnormal cystic structures following influenza injury, resembling honeycomb cysts found in human IPF. Strong Hes1/ HES1 expression indicating hyperactive Notch signalling was evident in these cysts in both mouse and human lungs[104]. Interestingly, histological analysis of human IPF lung revealed $KRT5^{+}p63^{+}$ basal cells overlying fibroblastic foci expressed markers associated with cell migration such as laminin-5 gamma-2, fascin and heat-shock protein 27[209]. It is tempting to speculate that severe alveolar damage and AT2 cell dysfunction in IPF, instigates a regenerative program of basal cell mobilisation from the distal airways[210]. Another possibility is that AT2 cells transdifferentiate toward $KRT5^{+}$ basal cells following parenchymal tissue destruction. Single-cell transcriptional profiling of human AT2 cells from normal and

fibrotic lungs suggested hypoxia and Notch signalling were drivers of a basal-like phenotype in AT2 cells[211]. Ingenuity pathway analysis revealed HIF1α was a key upstream regulator in fibrotic AT2 cells. After one passage in culture human AT2 cells upregulated *KRT5* mRNA[211].

Regardless of the cellular origin of these KRT5+ cells in human disease, their maladaptive response results in bronchiolised epithelial structures and irreversible loss of lung function. Interrogation of the IPF BAL cell transcriptome using microarrays found a significant association between the expression of an airway basal cell gene signature and increased mortality[212]. Airway basal cells identified by epithelial cell adhesion molecule (EPCAM) and KRT5 / 6 expression, comprised 0.25% BAL cells in IPF patients compared to 0.05% in age-matched healthy controls[212].

Single-cell transcriptomic work has provided evidence that more than one airway basal cell-like population is expanded in the remodelled fibrotic lung. A unique population of *KRT5*$^{-}$*KRT17*$^{+}$ epithelial cells have been discovered at the edge of fibroblastic foci in the IPF lung, which expressed some, but not all markers of basal cells[53,145]. These 'aberrant basaloid cells', co-expressed various mesenchymal and senescence-related genes, and transcription factor SOX9 which is involved in the development and repair of the distal airways[53]. The origin and function of these cells remains to be elucidated. Computational trajectory analysis using the 'Slingshot' R package suggested *KRT5*$^{-}$ *KRT17*$^{+}$ cells arise from transitional AT2 cells derived from either AT2 cells or secretory *SCGB3A2*$^{+}$ cells[145].

Immunostaining has shown honeycomb cysts contain regions of pseudostratified mucociliated epithelium, and 92% of the mucus cells expressed MUC5B[36]. MUC5B-expressing cells were found in 84% of distal conducting airways in normal lungs and were also abundant in submucosal glands[36]. Mucus cells within fibrotic regions expressed SRY (sex determining region Y) – box 2 (SOX2), a transcription factor typically associated with proliferation and differentiation of epithelial cells in the conducting airway[187]. Plantier et al. speculated that neuregulin-1 splice variant alpha (NRG1α) may drive an abnormal program of 'ectopic' airway epithelial cell

differentiation in IPF, as it does in vitro[187]. NRG1α was present in a subset of cells in the bronchiolised regions in IPF and elevated in IPF BAL compared to controls.

Taken together, these data emphasise the prominent changes that occur to the epithelial cell repertoire of the distal lung in IPF. Gaining traction is the notion that there is bronchiolisation of the distal lung, supported by evidence of an expansion in airway epithelial cell populations and even *COL15A1*- expressing vascular endothelial cells normally found in the peribronchial regions[53]. The studies described in this section typically utilised lung tissue from advanced cases of IPF, therefore the contribution of airway epithelial cells in the early stages of disease remains unknown. Furthermore, the data captures changes occurring in peripheral lung tissue sections, therefore it is unclear to what extent the proximal airway epithelium and associated cell types are involved. Lastly, some of the theories are derived from mouse data, which may not have relevance for a human disease associated with ageing.

1.2.7 Limitations of animal models of pulmonary fibrosis

The murine bleomycin model of pulmonary fibrosis is commonly used to provide mechanistic insight into the pathobiology of IPF[213]. This experimental model has facilitated critical advancement in our understanding of pathways of fibrogenesis, such as the central role played by TGF-β signalling[214]. However, there are important limitations to the translatability of this model to our understanding of human IPF. The bleomycin model fails to recapitulate all histopathological features of IPF and the progressive, irreversible nature of the disease[213,215]. In comparison to human disease, bleomycin-induced fibrosis develops rapidly (2 weeks after injury), is preceded by an acute inflammatory phase and largely resolves within 8 weeks[216]. The model is often performed in young mice, of a restricted genetic background, which are kept in specific pathogen free (SPF) conditions[216]. These factors can have a profound influence on the immune cell repertoire of mice and therefore translatability of findings to human disease[217].

Whilst the alveolar region of the lung is relatively well-conserved between species, there are key differences in structural anatomy and proximal-distal epithelial organisation between the murine and human airways[218]. For example, the

pseudostratified epithelium containing basal cells is restricted to trachea and proximal bronchi in mice yet extends throughout the conducting airways in humans[194].

For these reasons, the experimental work described henceforth in this thesis focuses on primary human cells and tissue samples acquired from patients with IPF and healthy controls.

1.3 **Summary**

A critical knowledge gap exists in our understanding of the contribution of the airway epithelial cell niche to the development of IPF. The diverse community of epithelial cells lining our airways regulate immune responses, defending the lower airspace from environmental threats. Alveolar epithelial cell injury is central to IPF pathogenesis, but as yet no study has determined if dysfunctional epithelial-immune cell interactions in the airway may be driving this. Given the remarkable epithelial aberrations and remodelling that take place in the distal lung in IPF, there is also a need to further interrogate airway epithelial phenotype and function in the conducting airway. Specifically, it will be important to establish if IPF airway epithelial cells are intrinsically 'primed' to promote pathological remodelling. Various properties of the fibrotic matrix have been shown to alter cellular phenotype but the extent to which airway epithelial cells are influenced by changes in the microenvironment are unknown. Functional studies using human samples are necessary to expose the mechanisms governing airway epithelial cell mobilisation within the fibrotic lung.

1.4 Hypothesis

Local tissue environmental signals influence airway epithelial cell phenotype and function and play an integral role in the pathogenesis of pulmonary fibrosis.

1.5 Aims

1. Characterise airway epithelial cell phenotype and function from patients with newly diagnosed IPF and healthy controls, under homeostatic conditions and following injury.
2. Determine the transcriptomic changes occurring in airway epithelial cells and airway macrophage populations in pulmonary fibrosis.
3. Explore cellular heterogeneity in the IPF airway epithelial niche and interrogate epithelial-immune cell interactions in the airway.
4. Determine how the local extracellular matrix (ECM) influences basal cell phenotype in the distal lung in health and during IPF.

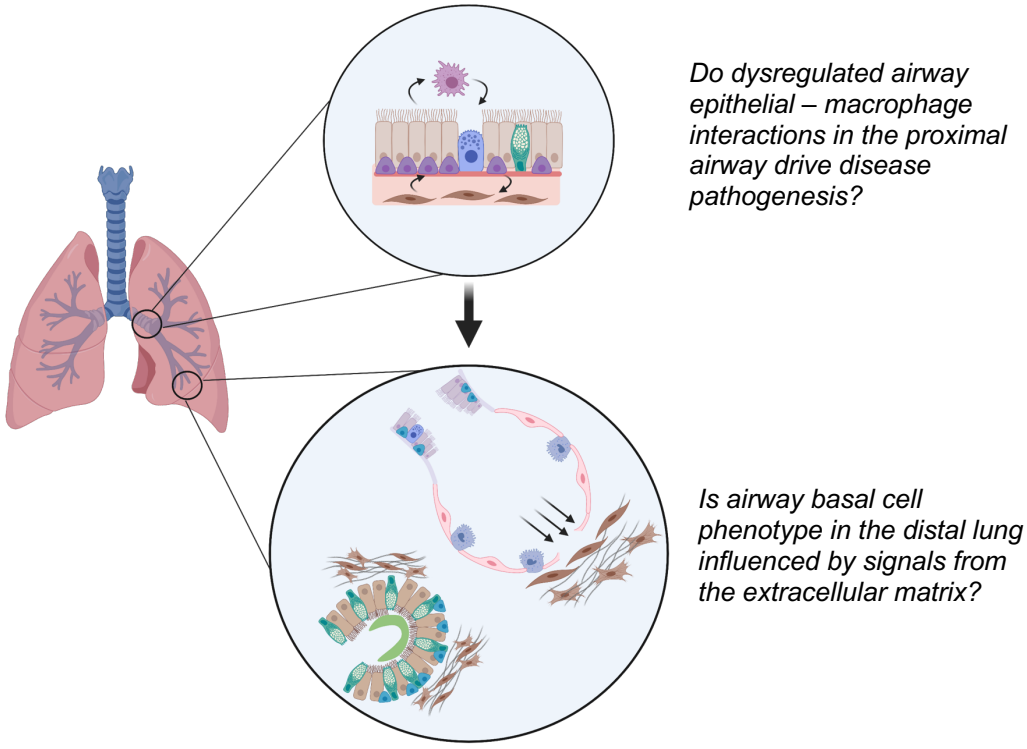


Figure 1.6 Summary of PhD hypothesis. Airway epithelial cell phenotype and function is influenced by signals from the local tissue microenvironment, which contribute to ongoing pathology. Figure created with BioRender.com.

CHAPTER 2

Materials and methods

2 Material and methods

2.1 **Subject recruitment**

2.1.1 IPF patients

Patients undergoing diagnostic work-up for interstitial lung disease (ILD) at the Royal Brompton Hospital (RBH) were prospectively recruited between March 2017 and March 2020. Subjects were included if clinical history and radiological pattern on high-resolution computed tomography (HRCT) scan were compatible with IPF (Chapter 1, Fig. 1.2). Only those subjects receiving a consensus diagnosis of IPF following a multi-disciplinary discussion in line with international guidelines[2], were included in the analysis. Pulmonary function testing was carried out prior to bronchoscopy. Composite physiological index (CPI)[19] was calculated for each IPF patient recruited using the formula:

CPI = 91 – (0.65 x %predicted DL_{CO}) – (0.53 x %predicted FVC) + (0.34 x %predicted FEV_1)

DL_{CO} – diffusing capacity of the lung for carbon monoxide (haemoglobin-adjusted)
FVC – forced vital capacity
FEV1 – forced expiratory volume in 1 second

CT data was scored as previously described[219] with fibrosis (FS) and honeycombing (HC) to the nearest 5% per lobe, and traction bronchiectasis (TB) on an 18-point scale. All human material was collected following written informed consent and in accordance with the Research Ethics Committee (NRES reference:15/SC/0101).

2.1.2 Healthy controls

Healthy control subjects comprised staff members at RBH and Imperial College London, and ex-smokers who were ineligible (on account of normal lung function) for inclusion in a locally run early COPD study. Subjects were recruited through local poster advertisements and emails sent to all staff members. Healthy controls included non-smokers or those with a previous history of smoking and normal lung function (FEV1 >80% predicted for age and height, FEV1/FVC ratio >70%), no diagnosis of ILD or symptoms suggestive of chronic lung disease. After a screening telephone call to determine eligibility, they were invited for a more detailed health screening questionnaire, spirometry and bronchoscopy. All human material was collected

following written informed consent and in accordance with the Research Ethics Committee (NRES reference:15/LO/1399).

2.2 **Human samples**

2.2.1 Flexible bronchoscopy with bronchoalveolar lavage (BAL) and bronchial brushings

A total of 98 IPF patients and 25 healthy controls consented and underwent a flexible bronchoscopy and BAL in accordance with international guidelines[1,28,220] (Fig. 2.1). Bronchoscopies were performed with lidocaine local anaesthesia and subjects lightly sedated with midazolam. Consistent with published data[221], the procedure was generally well-tolerated and there were no complications.

BAL was collected from a total of 95 IPF patients and 25 healthy controls and contributed to the transcriptomic work described in Chapter 4 and other research projects in our lab[30,174,219,222]. For this, the bronchoscope was wedged in the right middle lobe bronchus and four 60-ml aliquots of warmed, sterile saline were instilled and aspirated using a syringe. Lavage aliquots were pooled for each patient and the volume and BAL appearance were recorded for all samples. Of the BAL fluid returned, 50mls was used for hospital diagnostics and the surplus (~50 – 70mls) was used for research. BAL samples were transported on ice to the laboratory at Imperial College London.

Bronchial brushings were acquired from 35 IPF patients and 22 healthy controls for use in the *in vitro* studies described in Chapter's 3 and 6, and the transcriptomic studies in Chapter's 4 and 5. Under direct visualisation, two to three bronchial brushings (Olympus; #BC-202D-3010) were taken in the right main bronchus and placed directly into DMEM (Dulbecco's Modified Eagle Medium, Gibco) transport media when used for cell culture (Section 2.3), Buffer RLT (Qiagen) with 1% β-mercaptoethanol (β-ME) when being processed for bulk RNA-sequencing studies (Section 2.14), or HBSS (Hank's Balance Salt Solution, Gibco) supplemented with penicillin (1000 U/ml) and 1 mg/mL streptomycin (1mg/ml) for single-cell RNA-sequencing (Section 2.16). Bronchial brushings were transported at room temperature when used for cell culture, or on ice for RNA sequencing studies.

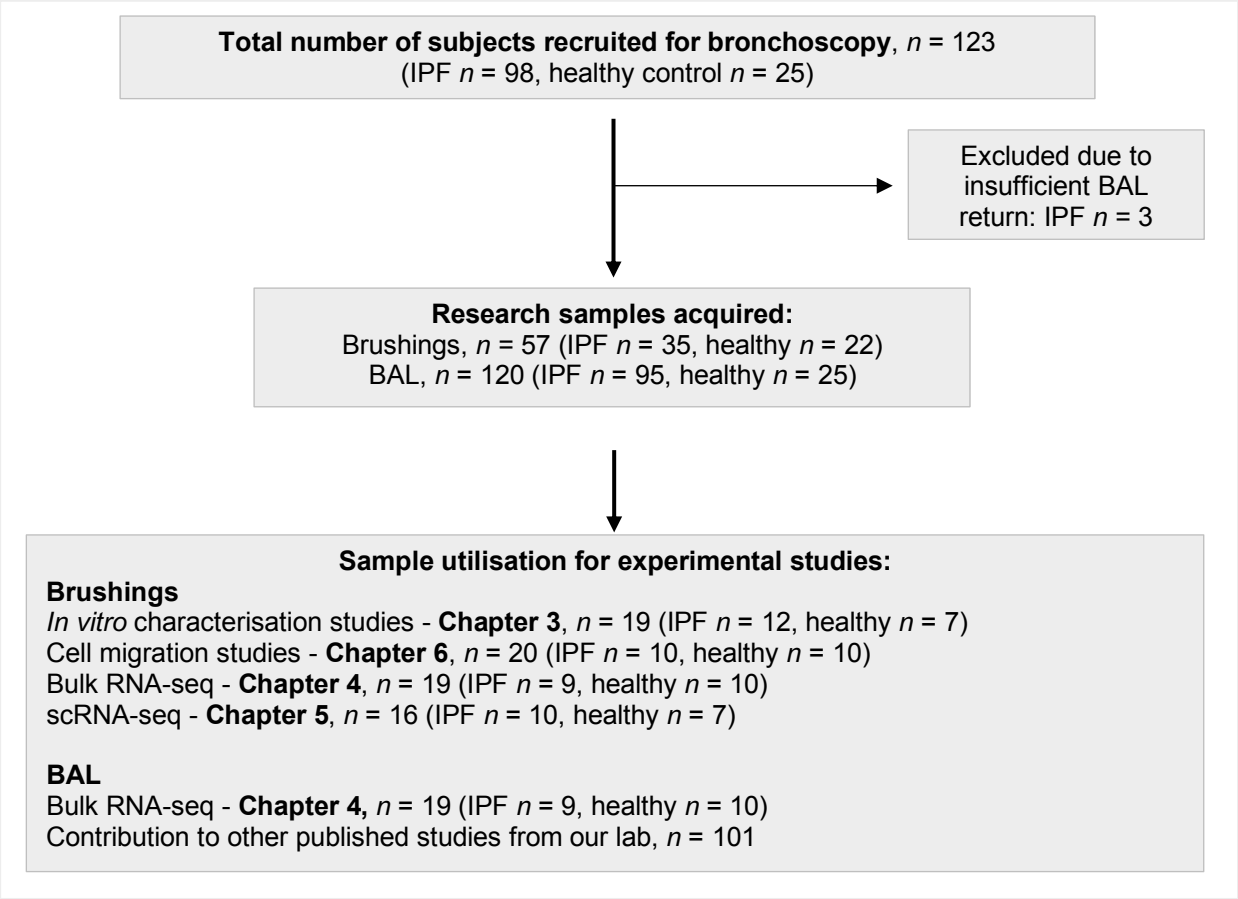


Figure 2.1 Subject recruitment and sample utilisation for experimental studies. A flow diagram indicating how human samples were used in this thesis. All subjects underwent BAL as part of a diagnostic work-up for ILD. From a subset of individuals, brushings were also taken from the right main bronchus for research purposes only. The sample utilisation box provides an indication of how participants samples were distributed between experimental studies. Some samples were used for more than one experiment e.g. 2 brushings were taken for culture studies and 2 brushings for transcriptomics. The sample numbers noted here for the sequencing studies were prior to any exclusion at the quality control stage.

2.2.2 Blood sampling

Prior to bronchoscopy, whole blood samples were collected from each subject in labelled EDTA vacutainer tubes (BD) via venepuncture. The samples were inverted 8-10 times before storing at -80^{o}C for future DNA extraction and genotyping.

2.2.3 Fresh human lung tissue specimens

Non-fibrotic control lung tissue specimens were obtained from patients undergoing lung cancer resections at the Royal Brompton Hospital (REC 15/SC/0101). The specimens consisted of normal lung tissue taken as distant from the tumour as

possible to minimise any impact on the cells. The tissue was transported in DMEM (Gibco) with Primocin (InvivoGen), on ice. This tissue was used for the isolation of primary human lung fibroblasts and for precision- cut lung slices (PCLS).

2.2.4 Archived formalin fixed paraffin embedded (FFPE) human lung tissue

Formalin fixed paraffin embedded (FFPE) lung parenchyma and proximal airway sections cut 4 - 5μm thick were procured from the Royal Brompton Hospital pathology archive with ethical approval (REC - 15/SC/0569). All specimens were reviewed by consultant histopathologists Professor Andrew Nicholson and Dr Alexandra Rice. IPF sections were obtained from explant lung tissue or surgical lung biopsies and exhibited a characteristic usual interstitial pneumonia (UIP) histological pattern. Normal lung tissue was obtained during non-small cell lung cancer resections from sites remote to the tumour.

2.3 **Cell culture**

2.3.1 Characteristics of primary bronchial epithelial cell (PBEC) donors

PBECs were isolated from brushes derived from participants (Table 2.1) and expanded in culture for experimentation. There was a male predominance in both IPF and healthy control groups. IPF subjects were older with a mean age 70 years, compared to controls 57 years. The majority of participants were ex- or never-smokers, however it is important to note for downstream functional studies, that IPF subjects ID_5224 and ID_5233 were current smokers. Pulmonary function tests revealed a spectrum of disease severity in the IPF group. Composite physiological index (CPI) - a measure of the extent of fibrosis in IPF patients - ranged between 24 (indicating less severe disease) to 64 (indicating more severe disease). Consistent with the known association between *MUC5B* (rs35705950) genotype status and IPF disease susceptibility, 67% IPF patients were heterozygous (GT) for the minor *T* allele, compared to 29% of healthy controls[223].

Table 2.1 Characteristics of PBEC donors for functional *in vitro* studies in Chapter 3.

Subject ID	Diagnosis	Age	Sex	Smoking status	FEV1 (% predicted)	FVC (% predicted)	FEV1/FVC	DLCO	CPI	*MUC5B* (rs35705950) genotype	Proliferation	Wound healing	Flu challenge	Co-culture
1017	Healthy	36	M	NA	NA	NA	NA	NA	NA	G/G	✓	✓		
1018	Healthy	65	M	NA	NA	NA	NA	NA	NA	G/T	✓	✓	✓	
1028	Healthy	63	M	NA	NA	NA	NA	NA	NA	G/G	✓	✓	✓	
1036	Healthy	63	M	Ex	1.98 (62.93)	2.91 (71.67)	0.68	NA	NA	G/T	✓	✓	✓	
1037	Healthy	55	F	Never	1.52(59.05)	1.77 (54.84)	0.86	NA	NA	G/G	✓	✓	✓	
1042	Healthy	63	M	Never	NA	NA	NA	NA	NA	G/G			✓	
1052	Healthy	52	F	Never	2.83 (107.6)	3.64 (111)	0.78	NA	NA	G/G			✓	
1290	IPF	70	M	Never	1.61 (57.00)	2.11 (57.30)	0.76	24.13	64.33	G/T		✓	✓	
1292	IPF	79	M	Ex	2.49 (93.77)	2.97 (83.20)	0.84	57.77	41.24	G/T	✓	✓		
1300	IPF	76	M	Ex	1.98 (76.81)	2.84 (83.05)	0.70	55.57	36.98	G/T	✓			
4127	IPF	77	F	Never	1.81 (103.74)	2.20 (102.41)	0.83	34.76	49.40	Undetermined		✓		
4130	IPF	72	M	Ex	2.76 (104.18)	3.66 (105.57)	0.75	53.01	36.01	NA	✓	✓		✓
5223	IPF	76	M	Ex	2.05 (77.41)	2.26 (64.21)	0.91	36.90	59.31	G/G	✓	✓	✓	
5224	IPF	52	M	Current	2.42 (74.41)	2.62 (64.72)	0.92	34.29	59.71	G/T	✓	✓	✓	
5233	IPF	66	M	Current	NA	NA	NA	NA	NA	G/T	✓	✓	✓	
5260	IPF	63	M	Ex	4.16 (127.87)	4.92 (118.21)	0.85	61.67	31.74	T/T		✓		
5354	IPF	74	F	Ex	1.76 (91.95)	2.36 (101.45)	0.75	42.41	40.93	G/T		✓		
5359	IPF	63	M	Ex	3.06 (96.06)	3.99 (97.98)	0.77	58.29	33.84	G/T		✓	✓	
5371	IPF	77	M	Ex	3.14 (125.6)	3.70 (111.04)	0.85	76.93	24.85	G/T			✓	

FEV1 – forced expiratory volume in 1 second, FVC – forced vital capacity, DL_{CO} – diffusing capacity of the lung for carbon monoxide (haemoglobin-adjusted). Tick indicates sample was used for *in vitro* experiment.

2.3.2 Isolation and culture of PBECs

Freshly collected bronchial brushings were agitated in media to detach cells and centrifuged (300x*g*, 5 minutes). The resultant cell pellet was resuspended in serum-free, Airway Epithelial Growth Medium (PromoCell) containing manufacturer's supplements and Primocin 100µg/ml (InvivoGen), added to protect against microbial contamination. The cell suspension was seeded in a 25cm^2 flask (Corning) pre-coated with purified type 1 bovine collagen (PureCol®, Advanced BioMatrix) diluted 1 in 100 in distilled water and incubated for 30 mins at 37^{o}C. Cells were cultivated at 37^{o}C in a humidified atmosphere of 5% CO_2. Media was changed every 48-hours and cells were passaged when 80% confluent using Tryple (Invitrogen). Cell counting and viability assessment was performed with 0.4% trypan blue (Sigma-Aldrich) and a haemocytometer using a similar approach to that described in Section 2.11.2. Live, unstained cells were counted. Cells were cryopreserved (1 x 10^6 / cryovial) at passage #2 in a freezing solution containing 80% PromoCell, 10% DMSO (Sigma-Aldrich) and 10% FBS (Gibco). Cryovials were stored in a freezing container (Mr Frosty, Thermo Scientific) overnight at -80^{o}C prior to transfer to liquid nitrogen. When required, PBECs were thawed quickly in a 37^{o}C water bath and cultured in a pre-coated 75cm^2 flask (Corning) until 80% confluent, then passaged (Passage #3) for use in experiments.

2.3.3 Air-liquid interface (ALI) culture

Passage #3 PBECs were seeded (4.0 x 10^4/ insert) on purified collagen 1 (Advanced BioMatrix) coated Transwell polyester membrane inserts (6.5mm diameter; 0.4µm pore size; Corning). PneumoCult-Ex media (STEMCELL) containing manufacturer's supplements, was added to apical and basal chambers and replaced 48-hourly. When confluent, the apical media was removed to promote mucociliary differentiation, and PneumoCult-ALI medium with manufactuer's supplements (STEMCELL) and Primocin (InvivoGen) was added to the basal chamber. From week 2 following air-exposure, mucus was gently washed from the apical surface with warmed PBS and stored for further analysis. Cells were maintained in ALI culture for 21 days, then inserts were either lysed in 350µl Buffer RLT (QIAGEN) for future gene expression analysis or fixed in 4% PFA (Electron Microscopy Sciences) for immunofluorescence microscopy.

2.3.4 Isolation and culture of primary human lung fibroblasts (HLFs)

HLFs were obtained from IPF surgical lung biopsy specimens at the Royal Brompton Hospital. Control HLFs were derived from lung tissue specimens collected during lung cancer resections (REC 15/SC/0101). For these, normal lung tissue was taken as distant from the tumour as possible to minimise any impact on the cells. Further details of the demographic and clinical characteristics of each participant are provided in the appendix.

HLF were isolated from tissue using an established technique[224]. Biopsied tissue was cut into 1mm^3 pieces and distributed evenly in wells of a 6-well plate. Linear, parallel scratches were made with a needle to encourage fibroblast outgrowth which could be seen after ~1 week. Explant cultures were maintained in DMEM (Gibco) with 10% FBS (Gibco) and Primocin (InvivoGen) at 37^{o}C in a humidified atmosphere of 5% CO_2. HLFs were cryopreserved (1 x 10^6 / cryovial) at passage #2 in a freezing solution containing DMEM (Gibco) with 10% DMSO (Sigma-Aldrich) and 10% FBS (Gibco). Cryovials were stored in a freezing container (Mr Frosty, Thermo Scientific) overnight at -80^{o}C prior to transfer to liquid nitrogen. When required, HLFs were thawed quickly in a 37^{o}C water bath and cultured in a pre-coated 75cm^2 flask (Corning) until 80% confluent, then passaged (passage #3) for use in experiments.

2.3.5 Epithelial-fibroblast co-culture

Primary HLFs (passage #3) were seeded (0.5 x 10^5/ insert) on the basal surface of a Transwell polyester membrane insert (6.5mm diameter, 0.4µm pore size; Corning) that had been inverted. HLFs adhered over 2 hours before the Transwell was placed upright and IPF PBECs (passage #3) were seeded on the apical surface (1.0 x 10^5 / insert). Media containing 50% DMEM (Gibco) and 50% Airway Epithelial Cell Growth Medium (PromoCell) supplemented with manufacturer's supplements and Primocin (InvivoGen), was added to apical and basal compartments to maintain the co-culture. After 1 week in submerged culture, the apical surface was air-exposed and basal media was supplemented with retinoic acid (50nM) to promote mucociliary differentiation[225]. After 30 days of ALI culture, the inserts were harvested for TEM analysis at RBH and immunofluorescence.

2.4 **Immunofluorescence staining and imaging**

2.4.1 Cytospins of bronchial brushes

To determine cellular constituents of a bronchial brush a cytospin was prepared. Cells were detached from the brush by agitation and centrifuged as previously described. The cells recovered were counted using a haemocytometer (~2.0–4.0 x 10^4), resuspended in PBS and centrifuged (18 *x g*, 4 minutes). The cells were fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Sciences) for 20 minutes, washed in PBS, then permeabilised with permeabilisation buffer (eBioscience) for 20 minutes (room temperature). A PAP pen (Sigma) was used to demarcate the area on the slide to be stained. The cells were blocked with 2% BSA in PBS for 30 minutes (room temperature) then stained with primary antibodies to acetylated alpha-tubulin (ciliated cells) and cytokeratin 5 (basal cells), detailed in Table 2.2 for 2 hours at room temperature. F-actin, an abundant cytoskeletal protein allowing distinction of cell morphology, was stained with Phalloidin (Alexa Fluor 488; Invitrogen). A goat anti-mouse secondary antibody conjugated to Alexa Fluor 568 (1:200; Life technologies) was then added for 30 minutes (room temperature) and nuclei counterstained with DAPI (1:2000; Sigma-Aldrich). Mowiol 4-88 (Sigma-Aldrich) was used to mount the 13mm coverslip prior to imaging on Leica SP5 inverted confocal microscope equipped with four lasers (405nm, 488nm, 543nm, 633nm). Images were acquired using the LAS AF platform (Leica) and analysed using Fiji software **(**http://fiji.sc)[226].

The same immunostaining approach was also used to identify cytokeratin 13- positive cells on cytospins of bronchial brushings (Section 4.4.6.5). Details of the primary antibody are provided in Table 2.2. To bind the primary antibody, a donkey anti-goat secondary antibody conjugated to Alexa Fluor 488 (Invitrogen) was used.

Table 2.2 List of antibodies used for confocal microscopy

Antigen	Host/ isotype	Clone	Fluorochrome	Manufacturer / Cat. No.	Dilution factor*
Cytospin IF – confirmation of cell types					
Alpha tubulin (acetyl K40)	Mouse/ IgG2B	6-11B-1	Unconjugated	Abcam #ab24610	1:200
Cytokeratin 5	Rabbit/ IgG	EP1601Y	AF 647	Abcam #ab193895	1:200
Cytospin IF – confirmation of cytokeratin 13 positive cells					
Cytokeratin 13	Goat/ IgG	Polyclonal	Unconjugated	Abcam #ab79279	1:100
Basal cells in submerged culture					
ZO-1	Mouse/ IgG1	ZO1-1A12	AF 488	Invitrogen #339188	1:200
EpCAM / CD326	Mouse/ IgG2b	9C4	AF 488	BioLegend #324209	1:100
Cytokeratin 5	Rabbit/ IgG	EP1601Y	AF 647	Abcam #ab193895	1:200
PBEC- ALI culture – confirmation of cell types					
ZO-1	Mouse/ IgG1	ZO1-1A12	AF 488	Invitrogen #339188	1:100
α- tubulin (acetyl K40)	Mouse/ IgG2B	6-11B-1	Unconjugated	Abcam #ab24610	1:100
MUC5AC	Mouse/ IgG1	Polyclonal	Unconjugated	Abcam #ab78660	1:100
PBEC- ALI flu challenge model					
ZO-1	Mouse/ IgG1	ZO1-1A12	AF 488	Invitrogen #339188	1:100
Influenza A, H1N1	Goat, IgG	Polyclonal	Unconjugated	Abcam #ab20841	1:100
PCLS distal lung					
CD326 (EpCAM)	Mouse/ IgG2b	9C4	AF 488	BioLegend #324209	1:100
CD206	Mouse/ IgG1	15-2	APC/ Cy7	BioLegend #321120	1:100
Cell-derived matrix – validation of decellurisation					
Collagen I	Rabbit/ IgG	Polyclonal	Unconjugated	Novus biological #NB600-408	1:100
Fibronectin	Rabbit/ IgG	Polyclonal	Unconjugated	Abcam #ab2413	1:100

Cell-derived matrix – BC migration experiments					
Collagen I	Mouse/ IgG1	COL-1	Unconjugated	Sigma-Aldrich #C2456	1:100
Fibronectin	Mouse/ IgG1	FN-15	Unconjugated	Sigma-Aldrich #F7387	1:100
FFPE distal lung tissue					
Cytokeratin 5	Rabbit/ IgG	EP1601Y	Unconjugated	Abcam #ab52635	1:200
Keratin 5	Rabbit	Poly19055	Unconjugated	BioLegend #905501	1:200
α-smooth muscle actin	Mouse/ IgG2	1A4	Cy3	Sigma-Aldrich #C6198	1:500
FFPE proximal lung tissue					
Cytokeratin 5	Rabbit/ IgG	EP1601Y	Unconjugated	Abcam #ab52635	1:200
Cytokeratin 13	Mouse/ IgG	AE8	Unconjugated	Abcam #ab16112	1:200

*all primary antibodies were diluted in blocking buffer

2.4.2 Basal cells in submerged culture

PBECs in media (PromoCell, supplemented as previously) were seeded (3.0 x 10^4/ well) on a collagen I (PureCol®, Advanced BioMatrix) coated 96-well μ-plate (Ibidi). When confluent, the cells were fixed, permeabilised and blocked as described in Section 2.4.1. The cells were incubated with primary conjugated antibodies to ZO-1 (tight junction marker), EpCAM (epithelial cell adhesion molecule) and cytokeratin-5 (Table 2.2) overnight at 4°C. Cell nuclei were counterstained with DAPI (1:2000 dilution). Imaging was carried out with a Leica SP5 inverted confocal microscope. Images were acquired using LAS AF (Leica) and analysed using Fiji software (http://fiji.sc**)[226].**

2.4.3 Basal cells on ECM proteins

UV-sterilised glass coverslips were pre-coated in ECM proteins as described in Section 2.21.1. PBECs were seeded at a density of 1.0 x 10^4 cells per well. After 24 hours of incubation these cells were fixed in 4% PFA (Electron Microscopy Sciences). As per previous immunostaining methodology, Phalloidin (Alexa Fluor 488; Invitrogen) was used to stain actin cytoskeleton and DAPI as a nuclear counterstain. The cells

were coverslip mounted with ProLong Gold Antifade Mountant (Invitrogen). The purpose of this parallel experiment was to gain insight into morphological changes that occur on different ECM components. High-resolution images were acquired by Dr Leo Carlin (Beatson Institute, Glasgow, UK) using a 40x oil objective on Zeiss 880 microscope using Airyscan detectors. Maximum intensity projection of Airyscan z-stacks were generated in Imaris (Bitplane).

2.4.4 Differentiated ALI cultures

After completing the ALI culture experiments, the Transwell inserts (Corning) were fixed, permeabilised and blocked as described in Section 2.4.1. Cells were stained overnight at 4^{o}C with primary antibodies to ZO-1 (tight junction protein), alpha tubulin (ciliated cells) and mucin 5AC (secretory cells) (Table 2.2). After washing in PBS, the secondary antibodies added for 30 minutes (room temperature) were goat anti-mouse IgG secondary antibody conjugated to Alexa Fluor 568 (1:200 dilution; Life technologies) and donkey anti-rabbit IgG conjugated to Brilliant Violet 421 (1:200 dilution; BioLegend). Nuclei were counterstained with DAPI (1:2000). To validate antibody specificity, negative control slides without primary antibody were prepared. The membrane was cut from each Transwell insert using a scalpel and mounted on a glass slide with ProLong Gold Antifade Mountant and a 13mm coverslip. Imaging was carried out as described in Section 2.4.1.

Differentiated ALI cultures infected with influenza A virus (Section 2.9.1) were processed using the same procedure. Cells were stained overnight (4^{o}C) with primary antibodies to influenza A, ZO-1 and cytokeratin-5 (Table 2.2). A donkey anti-goat secondary antibody conjugated to Alexa Fluor 647 (1:200; Invitrogen) was then added for 30 minutes (room temperature) and nuclei counterstained with DAPI (1:2000; Sigma-Aldrich). The inserts were mounted and imaged as described above.

2.4.5 Precision Cut Lung Slices (PCLS)

Fresh human lung control tissue was obtained from surgical resections performed at the Royal Brompton Hospital as described in Section 2.2.3. A method developed in our lab for performing PCLS on mouse lung tissue[227], was adapted for human lung tissue specimens. The lung tissue was slowly inflated with pre-warmed (37^{o}C) 4%

(w/v) low melting point agarose (Thermo Scientific) using a 21G needle and syringe, and placed on ice for 30 minutes to allow the agarose to set. An automated vibrating microtome (Compresstome VF-300 Tissue Slicer; Precisionary Instruments LLC) was used to cut 450μm slices which were incubated in a 24-well μ-plate (Ibidi) in DMEM (Gibco) supplemented with 10% FBS (Gibco) and Primocin (InvivoGen) at 37°C, 5% CO_2. The lung slices were fluorescently labelled with antibodies as indicated in Table 2.2. After 1.5 hours of incubation at 37°C, the antibodies were removed and washed with media. The lung slices immersed in a small volume of media (50μl) to ensure close apposition of the slices with the bottom of the well. Imaging was carried out on a heated stage using a Leica SP5 inverted confocal microscope with 10x objective. A timelapse z-stack image series was captured and analysed in using LAS AF (Leica) and analysed using Fiji software.

2.4.6 Cell-derived matrices (CDMs) from primary HLFs

Upon completion of the experiment described in Section 2.20, CDMs in a 24-well μ-plates (Ibidi) were gently washed in PBS and fixed in 10% formalin (Sigma-Aldrich) for 15 minutes at room temperature. The fixative was removed, CDMs were washed in PBS then blocked with 30% horse serum (Gibco) for 1 hour at room temperature. CDMs were stained overnight at 4°C with primary antibodies to ECM constituent's collagen I and fibronectin. After careful washing in PBS, secondary antibodies diluted in 30% horse serum were added for 30 minutes at room temperature. For the first experiment, designed to validate the decellularization procedure, the secondary antibody used was goat anti-rabbit IgG conjugated to Alexa Fluor 546 (1:200; Invitrogen). For the basal cell migration experiments, the secondary antibody used was goat anti-mouse Alexa fluor 568 (1:200; Invitrogen). In both experiments Phalloidin-iFluor 488 (Abcam) was used as per manufacturer's instructions and nuclei were counterstained with DAPI (1:1000; Sigma-Aldrich) for 20 minutes. The staining solution was removed from each well and Vectashield® Antifade Mounting Medium was added. Imaging was carried out using a 63x oil objective on a Leica SP5 inverted confocal microscope. Images were acquired using the LAS AF platform (Leica) and analysed using Fiji software **(**http://fiji.sc**)**

2.4.7 Paraffin embedded lung tissue sections

FFPE distal lung tissue sections of 4μm thickness were obtained from the Histology department at the Royal Brompton Hospital as described in Section 2.2.4. The tissue sections were deparaffinized by heating at 60°C for 1 hour then immersing three times for 5 minutes in Gentaclear (Genta Medical, UK). A graded alcohol series (100%, 90%, 70%) followed by PBS wash was used to rehydrate the tissue. Antigen retrieval was performed by heating the sections in 0.01M sodium citrate solution (pH 6.0) for 15 minutes in a microwave at 800 watts. A PAP pen (Sigma) was used to demarcate the area on the slide to be stained. Prior to staining the sections were blocked to prevent non-specific primary antibody binding with 10% goat serum (Sigma-Aldrich) for 1 hour (room temperature). The sections were stained overnight at 4°C with primary antibodies to cytokeratin 5 (basal cells) and alpha-smooth muscle actin (myofibroblasts). After washing in PBS, secondary antibodies diluted in 10% goat serum were added for 30 minutes at room temperature. The secondary antibodies used were goat anti-rabbit IgG conjugated to DyLight® 488 (1:200; Invitrogen). Nuclei were counterstained with DAPI (1:1000; Sigma-Aldrich) or anti-Histone H3 (1:200; C-terminus) AF 594 (#819405; BioLegend) for 20 minutes. ProLong Gold antifade mountant (Life Technologies) was applied directly to the tissue sections. As a negative control, tissue stained only with the secondary antibody was used. The sections were viewed on a Leica SP5 inverted confocal microscope using Leica LAS software. Images were acquired on a Leica SP5 inverted confocal microscope equipped with four lasers (405nm, 488nm, 543nm, 633nm), using a 10x, 20x and 63x oil objective lens, and Leica Application Suite X (LAS X) software. Images were processed in Fiji[226]. To obtain concurrent second harmonic generation images, stained tissue sections were also imaged with collaborators Fred Fercoq and Leo Carlin (Beatson Institute, Glasgow, UK) using Upright Zeiss 880 with spectral detection (32 channels from 410nm to 695nm) using a 20x/1.0 water immersion objective lens. More information on this approach is provided in Section 2.19.1.

An identical immunostaining approach was adopted to identify cytokeratin 13- positive cells and keratin 5 (basal) cells lining conducting airways in FFPE proximal lung tissue sections. For this experiment the secondary antibodies used were a goat anti-rabbit

IgG conjugated to DyLight® 488 (1:200; Invitrogen) and a donkey anti-mouse Alexa Fluor 568 (1:200; Invitrogen).

2.5 Histology processing of ALI Transwell inserts

After completing the ALI culture experiments, the Transwell inserts (Corning) were fixed in 4% PFA (Electron Microscopy Sciences) for 20 minutes at room temperature then washed in PBS. The membrane was cut from the Transwell inserts and carefully submerged in 4% low melting point agarose (Thermo Scientific) which had been warmed to 45^{o}C in a 24-well plate. To avoid damaging the epithelial cell layer during the process, it was essential that the agarose was not too viscous when performing this step. Ensuring the Transwell membrane was centred in the well, the 24-well plate was placed on ice to encourage the agarose to solidify at room temperature. The solidified agarose block containing the embedded membrane was carefully removed and placed in a histology cassette. Graded ethanol dehydration was performed by placing the cassette in a series of concentrations; 30%, 50% and 70% ethanol, each for 15 minutes. Paraffin embedding, sectioning (4μm), mounting, hematoxylin and eosin (H&E) and Periodic acid- Schiff (PAS) staining was performed by Lorraine Lawrence (NHLI, Imperial College London). Sections were imaged using a Leica DM2500 microscope and Leica Application Suite v4.7 software.

2.6 Transmission Electron Microscopy (TEM)

Co-cultures and airway brushings were submitted to Andrew Rogers (EM department, Royal Brompton Hospital) for transmission electron microscopy. ALI-cultured epithelial cells co-cultured with fibroblasts on porous membranes were excised from 6.5mm inserts and fixed with 2.5% glutaraldehyde (TAAB) in 0.05M sodium cacodylate buffer (pH 7.2) at room temperature for 1 h. Airway brushings in airway epithelial growth media (PromoCell) were fixed in an equivalent volume of 5.0% glutaraldehyde for 1 h, pelleted by centrifugation at 2500rpm (~1000g) for 5 min. The pellet was rinsed without resuspension in cacodylate buffer, the supernatant discarded and 60μl molten 1.0% agar added to the resuspended sample to form a loose pellet when set.

Sections of inserts and cell pellets in 2.0ml Eppendorf tubes were further rinsed in cacodylate buffer twice for 5 mins and post-fixed in buffered 1.0% osmium tetroxide

(TAAB) for 1 h at room temperature. The samples were then washed in purified water then dehydrated through an ascending ethanol series of 70%, 90% and 100%. A transition to 100% Araldite (Agar Scientific, UK) was achieved through an ethanol:Araldite series; 50:50, 25:75 mixtures then 100% Araldite overnight. Samples were embedded in moulds with 100% Araldite and polymerised over 3 days at 60^{o}C. ROIs were selected by light microscopy in microtome-cut $1\ \mu m$ thick toluidine blue stained, Araldite-embedded survey sections. Ultrathin 70nm sample sections were mounted on 200 mesh thin- bar copper grids and contrast stained with 1.5% uranyl acetate in methanol and a Reynolds lead citrate solution for examination at 120KV by TEM (JEOL 1400+; JEOL Ltd.) at the Royal Brompton Hospital, London. Micrographs of the co-cultures and airway brushings were acquired at magnifications ×10000 – x00000 (AMT XR16 camera; Deben UK Ltd.).

2.7 **Functional assessment of PBECs in submerged culture**

2.7.1 Proliferation assay

PBECs (passage #3) from IPF patients and healthy controls were seeded (2.0 x 10^4/ well) in a collagen coated 24-well plate (Corning) in airway epithelial cell growth media (PromoCell). JuLi Stage real-time cell-history recorder (NanoEnTek,Seoul, Korea) was used to capture time-lapse brightfield images of the culture (3 positions/ well) every hour for 48-hours. Proliferation data was generated by analysis of these images by JuLi Stage software.

2.7.2 Wound closure assay

PBECs (passage #3) from IPF patients and healthy controls were grown to confluence in a collagen I (PureCol®, Advanced BioMatrix) coated 24-well plate. A plastic pipette tip (P1000) was used to deliver a single, uniform, linear wound in each well. JuLi Stage real-time cell-history recorder (NanoEnTek,Seoul,Korea) was used to capture time-lapse brightfield images of the wound every 30 minutes for 48-hours. Wound closure data was generated by the JuLi Stage software from these images.

2.8 **Functional assessment of PBECs in ALI culture**

2.8.1 Transepithelial electrical resistance (TEER)

TEER is a validated, non-invasive technique for quantifying the integrity of tight junctions in cell culture models[228]. TEER of ALI cultures was measured on days 7, 14 and 21 post- air exposure using an EVOM2 Epithelial Voltohmmeter (World Precision Instruments) as per manufacturer's instructions. 100μl warmed PBS was added to the apical chamber of each Transwell (Corning) 5 minutes before measuring TEER. The electrode was place vertically in each well in a consistent position to measure TEER. Between wells the electrode was rinsed with 70% ethanol then distilled water.

2.8.2 Ciliary beat frequency (CBF)

CBF was measured using a well-established method at the Royal Brompton Hospital[229]. Cells from ALI cultures were harvested from the insert using a pipette tip, resuspended in PneumoCult- ALI media (STEMCELL) and examined in a chamber coverslide on the heated stage (37^{o}C) of a Leica DMLB light microscope (100x oil immersion lens). A digital high-speed video camera (Kodak Ektapro Motion Analyser,Model 1012) linked to CamLink software was used to record beating cilia at 400 frames/second, shutter speed 1 in 2000. CBF was calculated using the formula: CBF = 400 / (number of frames for 10 beats) x 10. Ten CBF measurements were taken per sample.

2.9 **Influenza virus infection of differentiated epithelial cells**

2.9.1 Influenza A virus infection

PBECs from IPF (*n*=8) and healthy controls (*n*=7) were differentiated in an ALI culture for 21 days as previously described. The apical surface was washed with PBS for 30 minutes (37^{o}C) to remove any mucus that can inhibit influenza virus infection. Influenza A virus (A/England/195/2009) was provided by Dr James Harker (Imperial, NHLI) and diluted in serum-free DMEM (Gibco). As a negative control, the same strain of virus was UV-inactivated using a Stratalinker UV 1800 Crosslinker (Stratagene). ALI culture infection was carried out in line with a previously published approach[230]. The ALI cultures were infected apically with active or UV-inactivated influenza virus to

achieve a multiplicity of infection (MOI) of 0.01[231] (there were approximately $5x10^5$ cells per Transwell insert). After 1 hour of incubation at 37°C, viral inocula were removed and the PBECs were washed with PBS. Infected ALI PBECs were maintained with PneumoCult- ALI media (STEMCELL) at 37°C, 5% CO_2. At 0-, 24-, 48- and 72-hours post-infection, TEER was measured and serum-free DMEM (Gibco) was added to the apical surface for 30 minutes (37°C) to collect viral particles and soluble mediators. Conditioned media from the apical surface and basal compartments were harvested at these timepoints and stored at -80°C for further analysis. Upon completion of the experiment, PBECs were fixed with 4% PFA for immunofluorescence analysis using the approach described in Section 2.4.4 or lysed with 350µl Buffer RLT (QIAGEN) with 1% β-mercaptoethanol and stored at -80°C for RNA extraction.

2.9.2 Lactate dehydrogenase (LDH) assay

Epithelial cytotoxicity/ cytolysis was assessed by measuring LDH released from damaged cells using a Cytotoxicity Detection Kit (LDH) (Roche). As per the manufacturer's instructions, 50µl of cell-free conditioned media was harvested at a defined experimental timepoint and combined with an equal volume of the assay reaction mixture (prepared by mixing 250µl of the catalyst solution with 11.25ml of the dye solution) in a 96-well flat-bottomed microplate. The microplate was incubated for 10 minutes at room temperature, protected from light. Fresh PneumoCult-ALI medium was used as a background control, and to serve as a positive control (100% cell lysis), cell-free conditioned media was harvested from PBEC ALIs treated with 1% Triton-X. Absorbance was measured at 490nm using a SpectraMax i3x plate reader with SoftMax Pro 7 software (Molecular Devices).

2.9.3 LEGENDplex™ cytometric bead array

A LEGENDplex™ (BioLegend) bead-based multiplex assay (Human Inflammation Panel 1, 13-plex; #740809) was used to simultaneously quantify 13 human inflammatory mediators including IL-1β, IFN-α2, IFN-γ, TNF-α, MCP-1 (CCL2), IL-6, IL-8 (CXCL8), IL-10, IL-22p70, IL-17A, IL-18, IL-23, and IL-33. Conditioned media harvested from the apical and basal chambers of PBEC ALI cultures 72hrs post-

influenza A infection was thawed and diluted 2-fold with assay buffer. The assay was carried out as per the manufacturer's instructions in a 96-well V-bottom microplate. The standards were run in duplicate and samples run in triplicate. Briefly, the samples were incubated with bead sets of different sizes and internal fluorescence intensities that are each conjugated with a capture antibody specific for one of the 13 analytes. Like a sandwich immunoassay, biotinylated detection antibodies detect specific analytes bound to the capture beads. Streptavidin-phycoerythrin (SA-PE) was then added which produces a fluorescent signal proportional to the amount of analyte bound which is quantifiable on a flow cytometer. The samples were run on a BD LSRFortessa II flow cytometer. The Human Inflammation Panel uses two sets of beads which could be identified on forward scatter (FSC) and side scatter (SSC) properties. The internal fluorescent dye was detected using the APC channel. The concentration of the analyte was determined from the average MFI of triplicate samples, using a standard curve in the LEGENDplex™ Data Analysis Software v8.0 (BioLegend).

2.10 **Flow cytometry of epithelial cells**

2.10.1 Airway brushings

A validated flow cytometry protocol was adopted[232] to characterise PBECs isolated from airway brushings. The staining protocol was optimised using an expanded 16-HBE cell line (Lonza) to ensure sufficient cell numbers for antibody titrations, isotype controls and fluorescent-minus one (FMO) controls. 16-HBEs ($1x10^5$ /well) cultured in MEM (Gibco) and PBECs ($5x10^4$/well) from two airway brushings were transferred to a 96-well round bottomed plate (Corning). Cells were washed in PBS (Gibco), centrifuged (400*xg,* 4°C, 3 minutes) and the pellet resuspended in 40µl Zombie Aqua™ Fixable Viability dye (1:1000; BioLegend). After incubating for 20 minutes at 4°C in the dark, cells were washed and resuspended in Human TruStain FcX Fc receptor blocking solution (1:100; BioLegend) for 10 minutes in the dark to block non-specific Fc-mediated interactions. Cell surface staining was performed with primary antibodies (Table 2.3) diluted in FACS buffer (1Mm EDTA, 25mM HEPES, 1% FBS in PBS) for 30 minutes at 4°C in the dark. Cells were washed then fixed in IC fixation buffer (eBioscience) for 10 minutes at room temperature. Following this, cells were washed and permeabilised with permeabilization buffer (eBioscience) for 20 minutes.

Intracellular staining was achieved by incubating the cells with antibody (Table 2.3) diluted in permeabilization buffer for 30 minutes at 4^{o}C in the dark. The cells were washed and re-suspended in FACS buffer prior to flow cytometry. A live-dead control sample was prepared by heating cells for 2 minutes at 70^{o}C. Fluorescence minus one (FMO) controls for each antibody were studied by omission of one antibody from the antibody panel. Compensation was performed using UltraComp eBeads (Invitrogen) for each antibody as per manufacturer's instructions. Flow cytometry was performed using a BD LSR Fortessa III cell analyser. FlowJo7 was used for data analysis **(**https://www.flowjo.com/solutions/flowjo/downloads**).** The gating strategy employed is shown in Fig. 2.2. Debris, doublets and dead cells were excluded and BECs identified as those which did not express leucocyte antigen CD45 but did express both pan-cytokeratin and EpCAM.

Table 2.3 Antibodies used for PBEC flow cytometry

Antigen	Host/ isotype	Clone	Fluorochrome	Manufacturer / Cat. No.	Dilution factor*	Method
Airway brushings						
CD45	Mouse/ IgG1	H130	BV 605	BioLegend #304042	1:100	EC
EpCAM / CD326	Mouse/ IgG2b	9C4	AF 488	BioLegend #324209	1:100	EC
Pan-cytokeratin (4, 5, 6, 8, 10, 13, 18)	Mouse/IgG1	C-11	PE	Abcam #ab52460	1:100	IC
Submerged culture						
EpCAM / CD326	Mouse/ IgG2b	9C4	AF 488	BioLegend #324209	1:100	EC
CD45	Mouse/ IgG1	H130	BV 605	BioLegend #304042	1:100	EC
Cytokeratin 5	Rabbit/ IgG	EP1601Y	AF 647	Abcam #ab193895	1:200	IC
Cytokeratin 14	Rabbit/ IgG	EP1612Y	PE	Abcam #ab210414	1:5000	IC
P63*	Rabbit/ IgG	EPR5701	Unconjugated	Abcam #ab124762	1:1000	IC

*conjugated with Goat Anti-Rabbit IgG conjugated to BV 421 (1:200; BD)
EC – extracellular stain, IC – intracellular stain

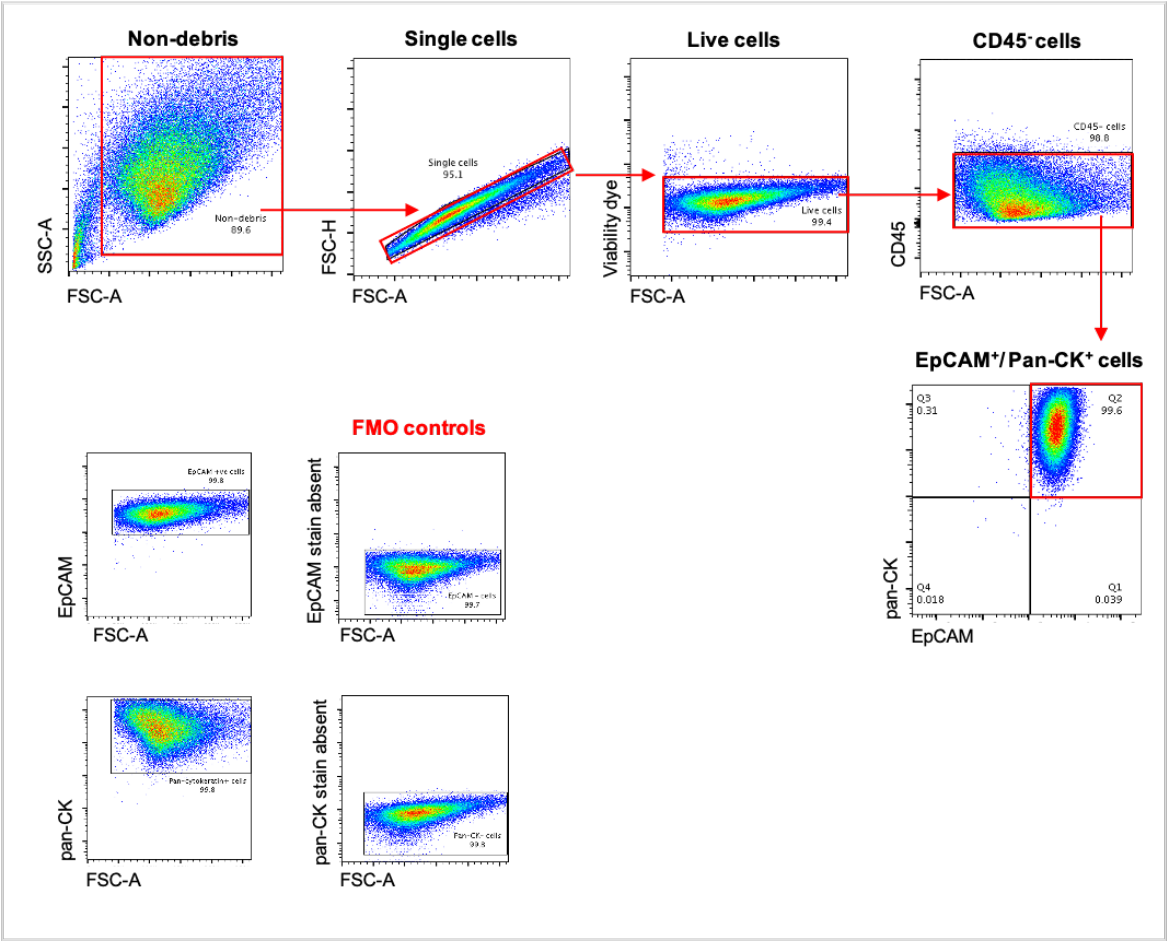


Figure 2.2 Gating strategy for epithelial cells using 16-HBE cell line. FSC-A vs. SSC-A were used to gate cells by size and granularity, excluding cell debris. Single cells were subgated by FSC-A/ FSC-H properties. A fixable viability dye was used to select live cells. CD45 -negative cells selected. The expression of EpCAM (cell surface) and pan-cytokeratin (intracellular) were used to determine epithelial cells within the $CD45^-$ population. Cell % from the parent gate are noted on each plot. Fluorescence minus one (FMO) controls for epithelial antibodies are illustrated below. FSC-A: forward scatter area; SSC-A: side scatter area; pan-CK – pan-cytokeratin.

2.10.2 Epithelial cells in submerged culture

Cultured primary airway epithelial cells (passage 3) were plated in a 96-well round bottomed plate (2 x 10^5/ well). Cells were washed in PBS (Gibco), centrifuged (500*xg,* 4^{o}C, 5 minutes) and the pellet resuspended in 50µl LIVE/DEAD fixable near-IR dead cell stain (1:1000 dilution in PBS; #L10119; Invitrogen). After a 20-minute incubation in the dark at room temperature, excess viability dye was removed with a PBS wash. Intracellular fixation and permeabilization buffer (eBioscience) was added to the cells overnight at 4^{o}C in the dark. Cells were washed in PBS and stained with 150 µl unconjugated primary antibody anti-p63 (Table 2.3), and Human Fc block (1:50; BD

Pharmingen) diluted in permeabilization buffer (eBioscience) for 30 minutes at 4°C in the dark. Cells were washed in permeabilization buffer and centrifuged. Secondary antibody staining was performed with Goat Anti-Rabbit IgG conjugated to BV421 (1:200; BD) diluted with permeabilization buffer in 2% BSA. Cells were washed and stained with remaining conjugated primary antibodies diluted in permeabilization buffer for 30 minutes at 4°C in the dark. After a final washing step, the cells were resupended in FACS buffer. Flow cytometry was performed using a BD LSR Fortessa III cell analyser. FlowJo7 was used for data analysis **(**https://www.flowjo.com/solutions/flowjo/downloads**).**

2.11 **Airway macrophage enrichment from BAL for bulk RNA-sequencing**

2.11.1 BAL processing

Human BAL acquired during bronchoscopy was preserved on ice in a 50ml centrifuge tube (Corning) and processed within 4 hours using a protocol established by our lab[30,174]. Whole BAL was filtered through a 70μM sterile strainer and centrifuged at 400 *x g* for 5 minutes at 4°C (Hettich Rotina 420R Centrifuge) to form a cell pellet. If necessary red cell lysis buffer (155mM NH_4CL, 10mM $KHCO_3$, 0.1mM EDTA, pH 7.3) was added for 10 minutes. The cell were washed and resuspended in 1- 3mls FACS buffer (1mM EDTA, 25mM HEPES and 1% FBS in PBS) depending on pellet size.

2.11.2 Cell counting

Total BAL cell counts were obtained using a staining solution consisting of 0.02% (w/v) crystal violet solution (Sigma) with 700mM acetic acid in PBS. Cells in suspension were added to the staining solution in a 1:10 dilution, and 10 µl was added to the haemocytometer. The number of cells in 4x4 grid in three areas of the haemocytometer grid were counted and averaged. The number of cells per ml was calculated using the following formula:

$$\text{Average cell count x dilution factor (10) x } 10^4 = \text{No. of cells/ ml}$$

2.11.3 Fluorescence- activated cell sorting (FACS)

Flow cytometry studies have shown mannose receptor, CD206 to be consistently expressed on AMs and therefore a useful marker for isolating this population[166]. BAL cells (5 x 10^6) were centrifuged (400 *x g*, 4°C, 3 minutes) in a 15ml centrifuge tube (Corning), washed in PBS and resuspended in 100µl of LIVE/DEAD fixable near-IR dead cell stain (1:1000 dilution in PBS; #L10119; Invitrogen). After a 10-minute incubation in the dark at room temperature, excess viability dye was removed with two PBS washes. Cell surface staining was performed with conjugated antibodies (Table 2.4) diluted in FACS buffer (1mM EDTA, 25mM HEPES and 1% FBS in PBS) in the presence of human Fc block (1:100; BD Pharmingen), for 20 minutes at 4°C in the dark. After staining, the cells were washed and resuspended in 1mls FACS buffer prior to cell sorting. Compensation was performed using UltraComp eBeads (Invitrogen) for each antibody as per manufacturer's instructions. Cell sorting was carried out by a member of Imperial College's Flow Cytometry Facility using a BD FACSAria III Cell Sorter. AMs were sorted as live, $CD45^+CD3^-TCR^-CD206^+$ cells, summarised by the gating strategy depicted in Chapter 4. A total of 150,000 - 250,000 cells were sorted into FACS buffer, pelleted and lysed in 350µl Buffer RLT (QIAGEN) containing 1% β-mercaptoethanol and stored at -80°C.

Table 2.4 Antibodies used for airway macrophage sorting.

Antigen	Clone	Fluorochrome	Manufacturer	Cat. No.	Dilution factor
FACS					
CD3	OKT3	FITC	BioLegend	317306	1:100
CD206	15-2	PerCP/Cyanine5.5	BioLegend	321122	1:100
CD45	H130	PE-Texas Red	Invitrogen	MHCD4517	1:100
TCR αβ	IP26	BV421	BioLegend	306722	1:25
MACS					
CD206	15-2	APC/Cyanine7	BioLegend	321120	1:50
Flow cytometry for purity assessment					
CD45*	H130	PerCP/Cyanine5.5	BioLegend	304028	1:100
CD206*	15-2	APC	BioLegend	321110	1:100
CD56	*HCD56*	*PE/Dazzle 594*	*BioLegend*	*318348*	*1:100*
CD117 (c-kit)	*104D2*	*BV605*	*BioLegend*	*313218*	*1:100*
Siglec-8	*7C9*	*PE*	*BioLegend*	*347104*	*1:100*
CD19	*H1B19*	*BV421*	*BioLegend*	*302234*	*1:100*
FcεRIα	*AER-37*	*AF700*	*BioLegend*	*334630*	*1:100*
CD3	*OKT3*	*PE/Cyanine7*	*BioLegend*	*317334*	*1:100*
CD14	*M5E2*	*BV711*	*BioLegend*	*301838*	*1:100*
CD177	*MEM-166*	*FITC*	*BioLegend*	*315804*	*1:100*

*Antibodies used for purity assessment

Antibodies listed in italic were included in the staining panel as were required by members of the lab for immunophenotyping BAL. They were not analysed in this project.

2.11.4 Magnetic-Activated Cell Sorting (MACS)

When FACS was unavailable due to the time of sample collection, MACS® Cell Separation (Miltenyi Biotec) based on CD206 expression was performed as per the manufacturer's protocol. After BAL processing, 5–10 x10^6 cells were washed, centrifuged (400*xg*, 4°C, 3 minutes) and incubated with APC/Cy7 conjugated anti-CD206 antibody (1:50; Biolegend) (Table 2.4) and Human Fc Receptor Blocking Solution, TruStain FcX™ (1:100; BioLegend) diluted in 500µl MACS buffer (2mM EDTA, 0.5% FBS in PBS). After a 15-minute incubation at 4°C in the dark, the cells were washed three times and resuspended in 80µl MACS buffer. Magnetic labelling

was performed by incubating the cell suspension with 20µl anti-Cy7 MicroBeads (Miltenyi Biotec) for 15 minutes in the dark at 4°C. The cells were washed three times, resuspended in 500µl MACS buffer and loaded onto a pre-rinsed MACS LS column (Miltenyi Biotec) in a MidiMACS™ magnet. The LS column was washed three times in MACS buffer and unlabelled cells which passed through (negative fraction) collected. Cell separation was achieved by removing the LS column from the magnet and using 5mls MACS buffer with the plunger supplied to firmly elute magnetically labelled cells (positive fraction). The cells were counted and a total of 250,000 sorted cells were pelleted and lysed in 350µl Buffer RLT (Qiagen) with 1% β-mercaptoethanol (β-ME) and stored at -80°C.

Sort purity was confirmed by flow cytometry using cells from unsorted BAL and the CD206 enriched fraction and represented as the % CD206+ cells of total cells (non-debris) in the sample.

2.11.5 Flow cytometry to assess sample purity

To confirm purity of MACS enriched samples prior to RNA-seq, flow cytometry was performed on cells from unsorted BAL (2.5 - 5.0 x 10^5 cells) and the CD206 enriched fraction (1.0 x 10^5). Although not relevant for this project, a wider panel of antibodies were used in the flow staining protocol (Table 2.4) to perform general BAL immunophenotyping for other members of the lab. Poonam Ghai (Research Assistant, Imperial College London) performed staining and flow cytometry. In a 96- well plate, cells were centrifuged (400*xg*, 4°C, 3 minutes), washed in PBS then resuspended in 10µl (1:1000 dilution in PBS) of LIVE/DEAD fixable near-IR dead cell stain (#L10119; Invitrogen). The live/dead stain was not used for MACS enriched samples because it has almost identical excitation and emission properties to the anti-CD206 APC/Cy7 antibody used. After a 10-minute incubation in the dark at room temperature, excess viability dye was removed with a PBS and FACS buffer washes. Cell surface staining was performed with conjugated antibodies (Table 2.4) diluted in 25µl FACS buffer in the presence of human Fc block (1:100; BD Pharmingen), for 20 minutes at 4°C in the dark. After staining, the cells were washed twice in FACS buffer prior to adding 50µl eBioscience Intracellular (IC) Fixation Buffer (Invitrogen) for 20 minutes at 4°C in the dark. The cells were washed twice and resuspended in 150µl FACS buffer. To improve

gating accuracy, fluorescence minus one (FMO) controls were also prepared which were cells stained with all fluorophores except one. A BD LSRFortessa II Flow Cytometer was used to acquire data. I analysed the raw FCS files using FlowJo v10 (FlowJo LLC) using a gating strategy previously published by our lab[222] (Fig. 2.3). Any enriched CD206+ AMs determined to be less than 80% pure were excluded from RNA-seq analysis.

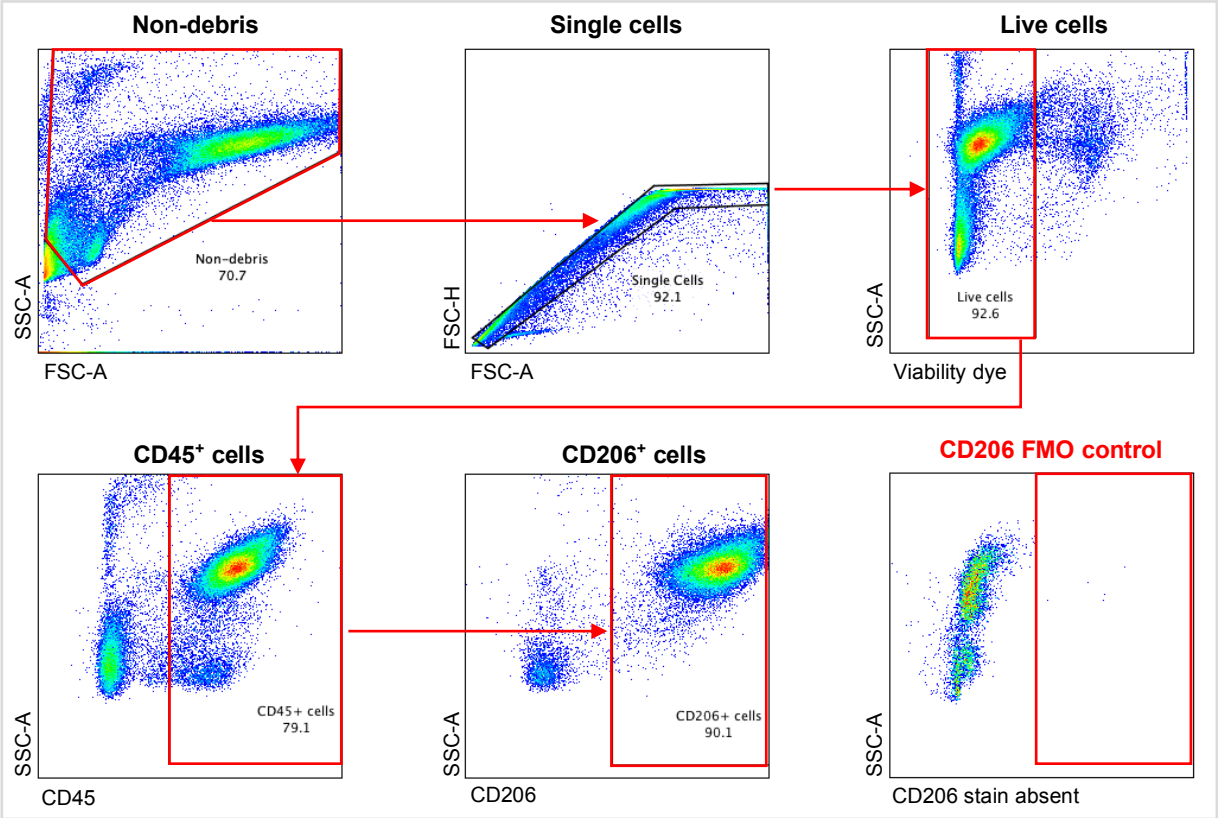


Figure 2.3 Gating strategy for airway macrophages. FSC-A vs. SSC-A were used to gate cells by size and granularity, excluding cell debris. Single cells were subgated by FSC-A/ FSC-H properties. A fixable viability dye was used to select live cells. From live, CD45+ cells, CD206+ cells were identified. Cell % from the parent gate are noted on each plot. Fluorescence minus one (FMO) control for CD206 is illustrated. FSC-A: forward scatter area; SSC-A: side scatter area.

2.12 **RNA extraction and quality control**

Cells were lysed by the addition of 350µl Buffer RLT (QIAGEN) containing 1% β-mercaptoethanol (β-ME). RNA was extracted using the RNeasy MiniKit (QIAGEN) as per manufacturer's instructions obtained at https://www.qiagen.com/gb/resources**.** This system purifies high-quality RNA from cells using silica-membrane RNeasy spin columns. RNA concentration and RNA integrity number (RIN) was confirmed using 1µl of the RNA sample with the Agilent RNA ScreenTape assay and Agilent 2200 TapeStation instrument and software, as per manufacturer's protocol. Purified RNA was stored at -80^{o}C for future experimentation.

2.13 **Wound healing PCR array**

The expression of a panel of genes involved in wound healing pathways, was evaluated by real-time RT-PCR using an RT^2 Profiler PCR Array (#PAHS-121Z; QIAGEN) in a 96-well plate format. Each plate contained primer assays for 84 wound healing genes (Appendix, Table 7.1) and 5 housekeeping genes for normalisation of the data. They also incorporated controls to test for genomic DNA contamination, and the efficiency of the reverse-transcription and PCR reactions. RNA samples harvested from epithelial cells from the influenza challenge model and basal cell migration experiments, were used for this PCR array as per the manufacturer's instructions. RNA purified and stored as per Section 2.12 was thawed on ice. A total of 300ng RNA from each sample was used to synthesise cDNA with the RT^2 First Strand Kit (QIAGEN), incubating at 42^{o}C for 15 minutes, then at 95^{o}C for 5 minutes in a Veriti 96-well Thermal Cycler (Applied Biosciences). The cDNA was mixed with the SYBR Green ROX qPCR Mastermix (QIAGEN) and aliquoted into the wells of the RT^2 Profiler PCR Array. PCR was performed on a ViiA 7 instrument (Applied Biosystems) using Fast 96-well block. The following cycling conditions were programmed; 1 cycle at 95^{o}C for 10 minutes; 40 cycles at 95^{o}C for 15 seconds and 60^{o}C for 1 minute, with a ramp rate adjusted to 1^{o}C/s. The real-time PCR threshold cycle (C_T) values were analysed using the GeneGlobe Data Analysis Center (Qiagen) **(**https://geneglobe.qiagen.com/gb/analyze/**).** Within this software, sample quality control was performed, and the housekeeping gene(s) displaying smallest variance across samples was selected. The expression level of a gene is represented as $\mathbf{2^{-C_T}}$, because there is an inverse proportional relationship between the threshold cycle (C_T)

and the level of gene expression, and the amount of product doubles with each cycle. A C_T cut-off of 35 was set for the analysis. Normalised gene expression values are calculated as the difference between the C_T value of the gene of interest (GOI) and the average C_T value of the selected housekeeping genes (HKG):

$$2^{-[C_T(GOI) - C_T(AVG\ HKG)]} = 2^{-\Delta C_T}$$

Fold- change is calculated using the delta delta C_T method[233] by dividing the normalised expression in IPF samples by the normalised expression in healthy control samples:

$$\frac{2^{-\Delta C_{T(IPF)}}}{2^{-\Delta C_{T(CON)}}} = 2^{-\Delta\Delta C_T}$$

A Student's t-test of the replicate normalised gene expression values for each gene in the IPF group and healthy control groups was performed. PCA and hierarchical clustering was performed in ClustVis[234] **(**https://biit.cs.ut.ee/clustvis/**).** Volcano plots were generated using VolcaNoseR[235] **(**https://huygens.science.uva.nl/VolcaNoseR/**).**

2.14 **Bulk RNA-seq of human airway brush and airway macrophages**

2.14.1 Preparation of cDNA libraries

From each sample, 200ng total RNA was used for cDNA library preparation using reagents in the Illumina TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) according to the manufacturer's Low Sample (LS) protocol. Briefly, the polydenylated (polyA) mRNA (capturing coding RNA and forms of non-coding RNA) was purified using poly-T oligo attached magnetic beads, fragmented under high temperatures and reverse transcribed with random primers into first strand cDNA. The next step involved second strand cDNA synthesis, incorporating dUTP in place of dTTP to quench the second strand during amplification. A single 'A' base was added to the 3' ends of the blunt double-stranded cDNA fragments, indexing adapters were ligated and 15 PCR cycles were used to enrich DNA fragments with adapters and amplify DNA in the library. The size of the final library product (~260 bp) for each sample was validated using a D1000 screentape on an Agilent Technologies 2100 Bioanalyser according to the manufacturer's protocol. The libraries were normalised and pooled for sequencing in the Advanced Sequencing Facility at the Francis Crick Institute.

2.14.2 Sequencing

The cDNA libraries were sequenced on an Illumina HiSeq 4000 sequencer and typically generated ~ 25 million paired-end reads (100 bp) per sample. The Advanced Sequencing Facility at the Francis Crick Institute (London, UK) performed the sequencing, demultiplexed the data, and provided raw FASTQ files for downstream analysis.

2.15 **Bulk RNA-seq data analysis**

2.15.1 Quality control

The RNA-Seq data analysis pipeline is depicted in Fig. 2.5. First, the raw paired-end FASTQ files for all samples underwent quality control using FastQC (Babraham Bioinformatics, Babraham, UK) and the data was aggregated into a single report using a Python package called MultiQC v1.6[236]. The mean quality value (Phred score) across each base position, per sequence quality scores and GC content were evaluated. All samples passed quality control.

2.15.2 Alignment

Prior to alignment, poor quality bases with an average quality score <10 were trimmed from 3' end to improve read alignment. The following illumina adapters were also trimmed from 3' end using using Cutadapt version 1.9.1 within Strand NGS:

Read 1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
Read 2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Raw reads were aligned to the human transcriptome (Genome Reference Consortium Human Build 38 GRCh38 (NCBI); Ensembl Genes and Transcripts (2016.12.01)) in StrandNGS v3.2. Strand NGS aligner uses COBWeb, a new proprietary algorithm which handles short and long reads, and allows user-defined gaps and mismatches. Default parameters were used for the alignment; minimum percent identity = 90%, maximum percent gaps = 5%, and maximum number of novel splices = 2%. Default output parameters were also maintained so that only 1 match per read with a match length ≥25 was reported. Read pairs with more than 5 valid matches

were ignored. The alignment report was reviewed and any samples with less than 70% uniquely matched reads were excluded from subsequent analysis.

2.15.3 Quantification and normalisation

The quantity of cDNA generated in the library prep and the number of reads that align to a parent gene, reflect the abundance of mRNA transcript present in the sample. Gene expression values are therefore derived from the quantification of raw read counts. The mapped and aligned reads (including partial reads) were quantified in StrandNGS v3.2. The DESeq approach[237] was used to compute 'normalised counts' for each gene in the sample from the raw read counts ($r_{i\,j}$). DESeq first calculates the geometric mean (m_i) of the raw counts for each gene across all the samples. For each sample it derives a normalisation factor (N_j) from the median of the values $r'_{i\,j}$, where $r'_{i\,j} = r_{i\,j} / m_i$. Normalised counts $n_{i\,j}$ for each gene are then calculated as $r_{i\,j} / N_j$. Normalised counts were log-transformed to generate 'normalised signal values' which allow better data visualisation.

2.15.4 Filtering for low counts

Low abundance genes were removed from downstream analysis by filtering based on a lower cut-off of 5 raw reads counts. Selection of this raw count threshold was determined by data presented in Section 4.4.5. Only genes with greater than 5 raw counts in at least 7 samples were included in the downstream analysis. This sample number was determined by the number of samples in the smallest group, minus 1.

2.15.5 Principal components analysis (PCA)

PCA is a linear dimensional reduction approach which summarises complex datasets by its top 'principal components'[238]. Typically, most of the variation in the data is described in the first few principal components (eigen vectors). All filtered genes were used in Strand NGS to perform PCA. The data was projected onto the top principal components (PC1 and PC2) to visualise major trends in the dataset.

2.15.6 Hierarchical clustering

Data was baseline transformed to the median of the normalised signal values of all samples. To identify groups of samples with unique transcriptomic profiles it is

necessary to select genes which display the most highly divergent levels of expression between samples[239]. Genes displaying low variability in expression and adding statistical noise were excluded by this approach. Therefore, two clustering approaches were used based on, (1) the top 5% most variant genes in the dataset, and (2) genes exhibiting greater than 4-fold deviation in expression up or down to the median expression calculated across all samples using a python script custom written by the developers of Strand NGS. Hierarchical clustering was performed using Pearson's centred distance metric and average linkage rule. The dendrogram was explored and any gene clusters ('modules') showing a distinct expression pattern in IPF compared to controls, were taken forward for further analysis.

2.15.7 Differential expression analysis

To identify differentially expressed genes between IPF and healthy controls in each dataset, the R package DESeq2 v1.6.1[240] was used within Strand NGS. The DESeq2 package is tailored to perform RNA-seq differential expression analysis where there is high variance (overdispersion) in gene count data and often experiments involve small sample numbers. A Wald test was used to determine the statistical significance of the log2 fold change differences, and the Benjamini – Hochberg procedure[241] was used for multiple-testing adjustment (pAdj <0.05).

2.15.8 Gene ontology (GO) Enrichment Analysis

GO enrichment analysis[242,243] was performed online at http://geneontology.org using gene sets from distinct clusters in the dataset. This website utilises GO annotations (GO Ontology database DOI: 10.5281/zenodo.4081749 Released 09-10-2020) with the PANTHER Classification System[244] (Released 28-07-2020) to find GO terms which are over-represented in a dataset. GO enrichment analysis was performed with a Fisher's exact test and only representative GO biological processes for FDR P <0.05 were reported.

2.15.9 Ingenuity Pathway Analysis (IPA)

Core analyses of upregulated and downregulated gene sets from the AM differential expression analysis were performed in IPA (QIAGEN Inc.). Based on gene expression changes in the dataset, this software package provides insight into signalling and

metabolic pathways activated, upstream regulators, interacting gene networks and biological function using curated information from the QIAGEN Knowledge Base[245]. There were several components to the analysis:

(1) **Canonical pathway analysis**, where the significance (-log(p-value) of the overlap of genes in the dataset with well-established, canonical pathways in the QIAGEN Knowledge Base was calculated by the right-tailed Fisher's Exact Test. Overall canonical pathway activation/ inhibition was predicted using the IPA z-score algorithm.

(2) **Upstream regulator analysis** was to predict upstream regulators (any molecule) that may be activated or inhibited to explain the gene expression changes observed in the dataset. The statistical significance was calculated by the right-tailed Fisher's Exact Test. Predicated activation z-score predicts the activation state of the upstream regulator based on the expression of genes downstream; only absolute z-scores $\geq$2 (activated) were considered significant and included in the analysis.

(3) **Network analysis** generated non-directional gene interaction networks based on the uploaded gene set[246]. Each network was described by the top 3 most significant biological functions. A network score - calculated as the negative exponent of the right-tailed Fisher's exact test – was used to rank networks according to the relevance to genes in the dataset. The higher the score, the greater number of 'network eligible molecules' in the dataset.

2.15.10 Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis (GSEA)[247] was performed using 947 genes associated with Pulmonary Fibrosis (MESH ID: D011658) from the curated Comparative Toxicogenomics Database[248] Gene-Disease Associations dataset, obtained via the Harmonizome portal[249]. GSEA determines whether members of an a priori defined gene set tend to occur toward the top or bottom of a list of all genes ranked according to their differential expression between two classes[247]. The AM expression dataset for IPF and healthy controls was prepared in a **.gct** format using ENSEMBL IDs and

normalised signal values for all filtered genes and uploaded to GSEA v4.1.0. ‘Collapse dataset to gene symbols’ was set to FALSE and the maximum size threshold to 1000 (to account for the size of the Pulmonary Fibrosis gene set). The number of permutations was set to 1000 for the analysis.

2.15.11 CIBERSORTx

CIBERSORTx is a computational method for ‘digital cytometry’, which can estimate cell type abundances from bulk RNA-seq data[250]. A matrix containing normalised signal values from all filtered genes was deconvoluted using the ‘Impute Cell Fractions’ function and a microarray derived leukocyte gene signature matrix of 22 immune cell types (LM22)[251]. B-mode batch correction was performed to remove technical differences between the LM22 microarray signature matrix and our bulk RNA-seq dataset. Quantile normalisation was disabled and 100 permutations were used for statistical analysis.

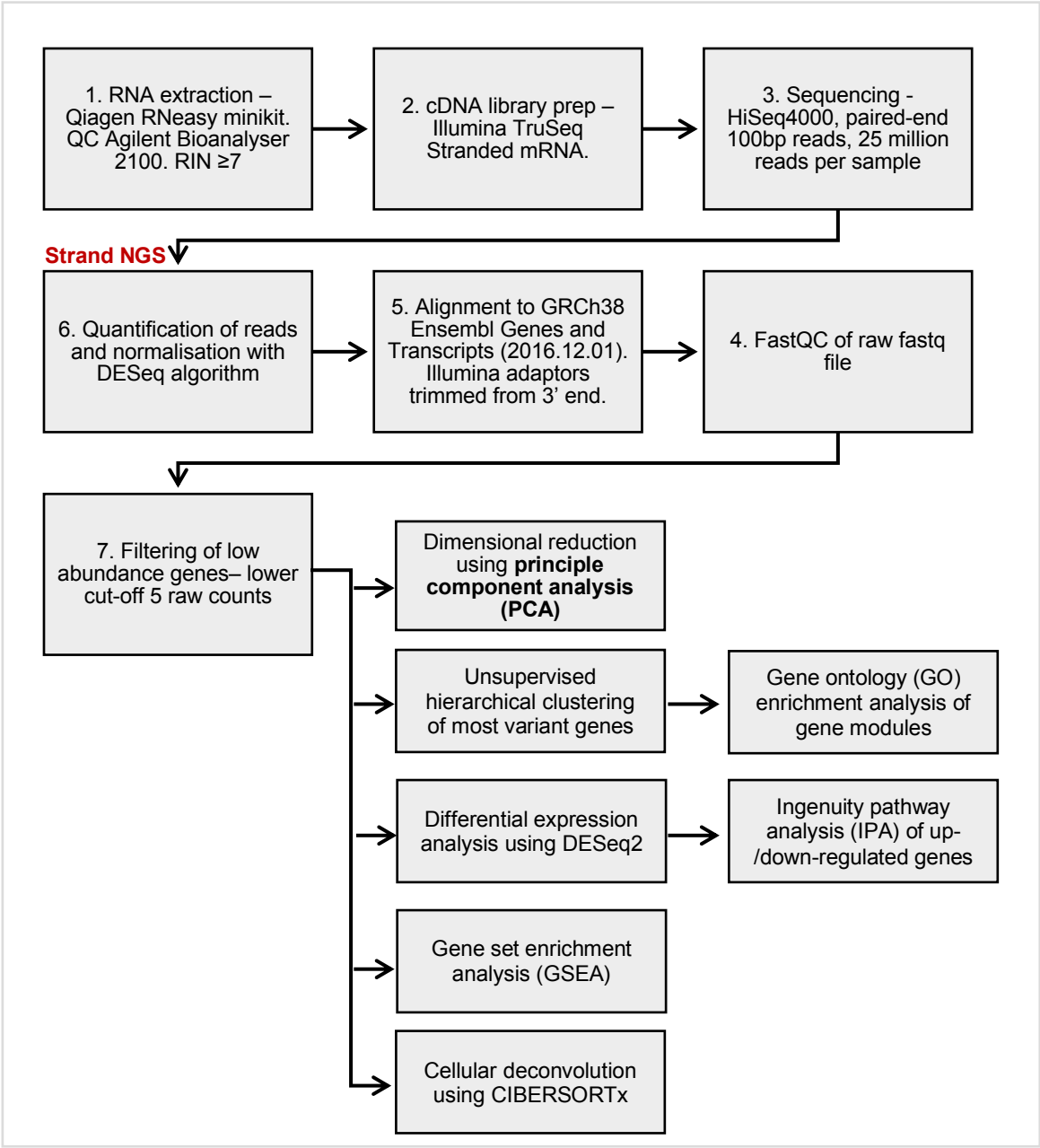


Figure 2.4 Bulk RNA-seq analysis pipeline

2.16 **Single-cell RNA-seq of human airway brushes**

2.16.1 Isolation of single cells from human airway brushes

A detailed protocol used in a previous publication[168] was kindly provided by Professor Martijn Nawijn (University of Groningen, the Netherlands) via personal communication. Two bronchial brushes were collected as described in 2.2.1 and transported on ice to the O'Garra Lab at Francis Crick Institute, London, UK. Here, the

brushes were centrifuged (1200rpm, 5mins, 4°C) and the resulting cell pellet resuspended in 1ml HBSS with 1mg/ml collagenase D from *Clostridium histolyticum* (Roche) and 0.1mg/ml DNase I from bovine pancreas (Roche) to and transferred to a FACS tube. The cell suspension was incubated for 1 hour at 37°C, with gentle agitation to avoid cell clumping. Dissociated cells were then passed through a 70μm filter, centrifuged (1200rpm, 5mins, 4°C) and the resulting cell pellet resuspended in HBSS with 0.04% UltraPure BSA (Invitrogen). The single-cell suspension was cleared of red blood cells using ACK red cell lysis buffer (Gibco). A final single-cell suspension of airway epithelial cells in 300μl HBSS with 0.04% UltraPure BSA was processed for scRNA-seq. The viability and concentration of each single-cell suspension was measured using trypan blue (Sigma) and the Eve automatic cell counter.

2.16.2 Chromium 10x Genomics library construction and sequencing

Single-cell library construction and sequencing was performed by Jimena Perez-Lloret (Advanced Sequencing Facility, The Francis Crick Institute). Approximately 1.0 x 10^4 – 1.3 x 10^4 cells were loaded for each sample into a separate channel of a Chromium Chip G for use in the 10X Chromium Controller (#PN-1000120). The cells were partitioned into nanoliter scale Gel Beads in emulsions (GEMs) and lysed using the 10x Genomics Single Cell 3′ Chip v3.1 GEM, Library and Gel Bead Kit (#PN-1000121). cDNA synthesis and library construction were performed as per the manufacturer's instructions (Chromium Single Cell 3' Reagent Kits v3 User Guide). The RNA was reversed transcribed and amplified using 12 cycles of PCR. Libraries were prepared from 10μl of the cDNA and 12 cycles of amplification. Each library was prepared using Single Index Kit T Set A (#PN-1000213) and sequenced on the HiSeq4000 system (Illumina) using 100 bp paired-end run to an approximate depth of 50,000 reads per cell. Libraries were generated in independent runs for the different samples.

2.17 **Single-cell RNA-seq data analysis**

2.17.1 Pre-processing

Single-cell RNA-seq data analysis was performed in collaboration with Probir Chakravarty (Bioinformatics and Biostatistics, The Francis Crick Institute), who executed all coding for the project. The scRNA-Seq pipeline is summarised in Fig. 2.5.

Raw reads were initially processed by the Cell Ranger (10X Genomics) v3.0.2 pipeline, which deconvolved reads to their cell of origin using the UMI tags, aligned these to the human transcriptome (version GRCh38-3.0.0) using STAR (v2.5.1b) and reported cell-specific gene expression count estimates. Raw feature count matrix from the 'EmptyDrops' algorithm output of Cell Ranger were used to correct for Ambient RNA using the FastCAR R package with default settings **(**https://github.com/LungCellAtlas/FastCAR**).** All subsequent analyses were performed in R v3.6.0 using the Seurat (v3) package[252] **(**https://satijalab.org/seurat/**).** Genes were considered to be expressed if the estimated (log10) count was at least 0.1. Filtering was then performed to remove from consideration; cells with less than 1000 mapped reads; cells expressing fewer than 500 genes and cells for which mitochondrial genes made up greater than 3 standard deviations from the median per patient. Each dataset was normalised using the 'logNormalize' function with a scale factor of 10,000. The top 2000 highly variable genes were found using the 'FindVariableGenes' function and the data centred and scaled using the 'ScaleData' function. PCA decomposition was performed and, after consideration of the eigenvalue 'elbow-plots', the first 40 components were used to construct the UMAP plot for each individual sample. Doublets were flagged using the DoubletFinder package[253], and rates calculated per sample, but not removed from downstream analyses.

2.17.2 Sample integration

To improve sample integration, further filtering was performed to exclude cells with less than 100 mapped reads; cells expressing fewer than 200 genes and cells for which mitochondrial genes made up greater than 3 standard deviations from the median per patient. Samples from IPF and healthy controls were then integrated using Seurat's 'IntegrateData' function using 2000 integration anchors and the first 50 components were used to construct the integrated UMAP plot. To evaluate the contribution of each subject to the integrated clusters obtained, a UMAP was generated for each sample using the integrated UMAP 'space'.

2.17.3 Cluster annotation

Cluster specific markers were identified using the 'FindMarkers' function using default settings (min.pct=0.25 and logfc.threshold = 0.25), which uses the Wilcoxon rank sum test to compare each cell belonging to one cluster versus all other cells. Using the statistically significant gene markers generated per cluster by the 'FindMarkers' algorithm, cell type annotation was achieved using; (1) marker genes manually identified from the literature[163,168,179], (2) the scCatch package[254], and (3) GSEA using marker gene lists from the literature[163,179].

2.17.4 Subclustering

To study basal cell populations in more detail without the influence of other cell clusters, cells annotated as basal cells (cluster 5) and suprabasal cells (cluster 3) of the integrated dataset were subclustered in Seurat. Using the approach previously described, annotations were assigned based on the gene markers generated per cluster by the 'FindMarkers' algorithm. To filter any contaminating immune cells, only cells expressing epithelial lineage marker, *EPCAM* (log2 normalised expression >0.25) but not immune cell lineage marker, *PTPRC* (log2 normalised expression <0.1) were taken forward for subclustering.

2.17.5 Differential expression analysis

Differential genes expressed between IPF and healthy controls per cluster were identified using the FindMarkers function using the following parameters (min.pct=0.1 and logfc.threshold = 0.25, and min.pct=0.1 and logfc.threshold = 0.1). Pathway and process enrichment analysis using differentially expressed genes was carried out using GSEA[247,255] and PANTHER GO[242-244].

2.17.6 Trajectory inference using Monocle 3

Monocle 3[256,257] was used to infer epithelial cell differentiation trajectory in IPF and healthy airway samples. Immune cell clusters were removed and only epithelial cells (*EPCAM* log2 normalised expression >0.4; *PTPRC* log2 normalised expression <0.1) were taken forward. Basal cells represent the earliest developmental stage in epithelial differentiation, so the root-node of the trajectory was selected by taking the average

of the co-ordinates of each cell in cluster 5 on the UMAP. Default parameters for Monocle 3 were used in the analysis, except for the cluster cells function in which we used k = 3, res = 10^{-5}.

2.17.7 Trajectory inference using Slingshot

Lineage trajectories were constructed by Slingshot[258] (version 1.2.0) running on R (version 3.6.1) with a PCA plot used as dimensionality reduction. Slingshot performs well at identifying multiple trajectories from noisy single-cell data[258]. Immune cells were removed and only cells with *EPCAM* expression greater than 0.4 and *PTPRC* expression less than 0.1 taken forward for trajectory analysis. The start of the lineage was set as basal cell cluster 5 and the terminal clusters were set as ciliated clusters 8, 9, 10, and 14, and ionocyte cluster 15, based on prior knowledge of lineage relationships in the human airway epithelium[259]. The trajectories were considered the developmental pseudotime of each lineage. Gene expression dynamics along pseudotime per lineage were identified by the R package gam (version 1.20) and significantly changing genes were selected with p values less than 1 x 10^{-5}.

2.17.8 Cell-cell interactions using CellPhoneDB

Interactions between cell clusters were determined using CellphoneDB[260,261] v2.0 **(**https://github.com/Teichlab/cellphonedb**)** using the integrated Seurat object after subsetting into cells from IPF or healthy subjects. Using scRNA-seq data, CellPhoneDB infers enriched receptor–ligand interactions between two cell types based on expression of receptor by one cell and corresponding ligand by another. Only ligands or receptors present in at least 30% of the cells in a cluster were considered in this analysis. Detailed information regarding the ligand-receptor database and statistical approach are available online at https://www.cellphonedb.org/documentation**.** Briefly, random permutation of the cluster labels of each cell was performed 1000 times. The combined mean of the average receptor expression level in a cluster and average ligand expression in the interacting cluster were computed for each permutation. A null distribution was established by pairwise comparison of each receptor-ligand pair between two interacting clusters. A p-value for the cell-type specificity of a receptor-ligand pair was calculated by the proportion of means which are the same or greater than the actual

mean. To select biologically relevant interacts in the dataset, highly enriched interactions (based on the number of significant pairs) are prioritised by CellPhoneDB. Interactions between clusters were visualised using Cytoscape v3.7.0 **(**https://cytoscape.org/**),** where the node (cell type) size is proportional to the total number of interactions of each cell type, and the edge thickness is proportional to the number of interactions between the connecting clusters. Any clusters with fewer than 20 interactions were not included in this network analysis.

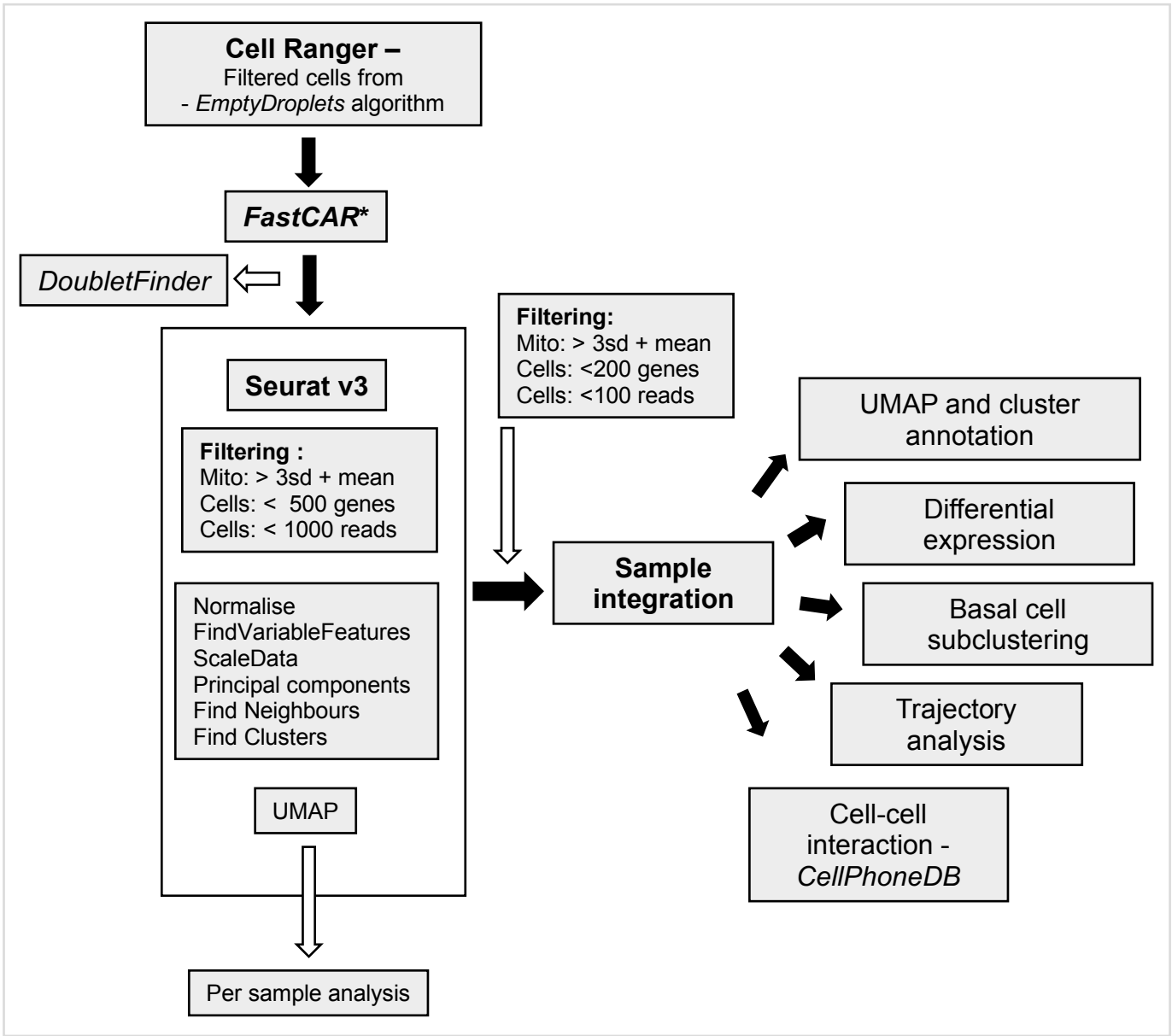


Figure 2.5 Single-cell RNA-seq pipeline

2.18 *MUC5B* genotyping assay

MUC5B SNP *rs35705950* genotyping was performed by Poonam Ghai (Research Assistant, NHLI, Imperial College London). The polymorphism of interest is a G/T transversion substitution. Genomic DNA was isolated from stored whole blood samples using the FlexiGene DNA Kit (#51206;QIAGEN) as per manufacturer's protocol. All subjects were genotyped for the *MUC5B* polymorphism using a TaqMan SNP Genotyping Assay (Assay ID: C_1582254_20; Cat. #4351379; Applied Biosystems) from 20ng DNA as per manufacturer's instructions. Reactions were carried out in a 96-well reaction plate using a ViiA7 (Applied Biosystems) Real-Time PCR system with the Fast 96 – Genotyping- Standard protocol.

2.19 Second harmonic generation imaging and analysis

2.19.1 Image acquisition

Lung tissue sections stained as described above were imaged with collaborators Fred Fercoq and Leo Carlin (Group Leader) at the Beatson Institute, Glasgow, UK. Images were acquired with a Zeiss 880 2-photon microscope (Beatson Advanced Imaging Resources, Beatson Institute) equipped with a 32 channel Gallium arsenide phosphide (GaAsP) spectral detector (Carl Zeiss) using 20x/1 NA water immersion objective lens. Samples were excited with a tuneable multiphoton laser (680-1300nm) set up at 900nm or 100nm and/or normal lasers (488nm, 561nm and 633nm) depending on immunostaining used, and signal was collected onto a linear array of the 32 GaAsp detectors in lambda mode with a resolution of 8.9nm over the visible spectrum. Spectral images were then unmixed with Zen software (Carl Zeiss) using references spectra acquired from unstained tissues (tissue autofluorescence and Second Harmonic Generation), DAPI-stained tissue or beads labelled with AF488-, Cy3-, AF594- or AF647-conjugated antibodies. Images were acquired in each of the following regions: alveolar (ALV-NORM, ALV-IPF), airway basement membrane (BM-NORM, IPF-NORM), perivascular space (PVS-NORM, PVS-IPF), honeycomb cyst (CYST) and fibrotic (FIBRO).

2.19.2 Collagen texture analysis

To analyse the structural characteristics of collagen fibres in each region, a variety of metrics were obtained using IMARIS and OrientationJ and GLCM plugins for Fiji/

ImageJ (Table 2.5). Importantly, IMARIS and OrientationJ utilise properties of the original image, and are so-called first-order statistics. In contrast, grey-level co-occurrence matrix (GLCM) texture analysis derives statistical measures by considering the spatial relationship of pixels in an image, so-called 'second-order statistics'. Regional enumeration of fluorescently labelled KRT5+ cells associated with collagen matrix was done manually in Fiji.

Table 2.5 Collagen analysis measurements

Measurement	Software	Description
Collagen intensity	Imaris Software, Oxford Instruments	Mean intensity of SHG signal in collagen fibres.
Collagen density	Imaris Software, Oxford Instruments	Collagen area represented as a percentage coverage of the total image area analysed.
Coherency	OrientationJ plugin, Fiji	Measures fibre orientation and coherency; a value of zero indicates perpendicular fibres, a value of 1 indicates parallel fibres.
Contrast	GLCM plugin, Fiji	Measures local variations in the grey-level co-occurrence matrix; a value of zero indicates a constant image with no variation.
Entropy	GLCM plugin, Fiji	Statistical measure of randomness, ranging from zero to infinity.
Correlation	GLCM plugin, Fiji	Linear dependency of grey levels on those of neighbouring pixels; higher values for similar grey-level regions.
Inverse Difference Moment (IDM)	GLCM plugin, Fiji	Local homogeneity or smoothness across an image; a value of 1 indicates a grey levels of the pixel pairs are similar
Angular Second Moment (ASM)	GLCM plugin, Fiji	Measures the number of repeated pixel pairs indicating uniformity of distribution of grey level in the image; a value of 1 indicates a constant image.

Image J plugins obtained from the following websites –
https://imageJ.nih.gov/ij/plugins/texture.html
http://bigwww.epfl.ch/demo/orientation/

An explanation of GLCM features was obtained from –
https://www.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html.

2.20 **Cell-derived matrices (CDM) from primary human lung fibroblasts**

The approach used to generate CDMs from primary HLFs was based on a protocol published by Kaukonen et al.[262] HLFs isolated from IPF (*n*=3) and control (*n*=3) subjects as described in Section 2.3.3, were seeded in wells of a 24-well μ-plate (Ibidi) at a density of 4.0 x 10^4. The cultures were maintained in DMEM (Gibco) media supplemented with 10% FBS (Sigma-Aldrich) and Primocin (InvivoGen) in an incubator (37ºC, 5% CO_2). The cells formed a confluent monolayer and media was replaced every 48 hours for 21 days to allow CDM production by HLFs. Medium was removed and the cells washed with PBS. HLFs were denuded by adding 500μl warmed sterile- extraction solution (1ml of NH_4OH, 250μl of Triton-X-100 and 48.75ml PBS passed through a 0.45μm filter) for approximately ~1 minute whilst closely observing the integrity of the matrix using a light microscope. The extraction buffer was gently removed and the CDM washed carefully with PBS to avoid disruption. Residual cellular DNA was removed by incubating CDMs (37ºC, 30 minutes) with 500μl DNase I (10μg ml^{-1}; Roche). The CDM was gently washed with PBS and matrix integrity was validated using a light-microscope. The CDM was stored in sterile PBS containing Primocin (InvivoGen) at 4ºC until required for basal cell migration experiments and immunofluorescence imaging. Notably we did not use gelatin-coated coverslips or ascorbic acid in the media to promote cross-linking of collagen as suggested in the protocol by Kaukonen et al.[262] as did not want to artificially alter the properties of the matrix in a way that may influence the cell migration results or MS-proteomic data generated.

2.21 **Basal cell migration experiments**

2.21.1 Basal cell migration on ECM substrates

Individual wells of 48-well tissue culture plates (Corning) were coated with purified human collagen type I (Sigma-Aldrich, #CC050), type III (Sigma-Aldrich, #CC054), type IV (Sigma-Aldrich, #C6745), fibronectin (Sigma-Aldrich, #F2006) and recombinant human versican (RayBiotech, #230-00833-50) at a concentration of 50μg/ml, for 1 hour in an incubator (37ºC, 5% CO_2) in line with previously published methods from our lab[263]. Primary human bronchial epithelial cells (PBECs) from healthy controls (*n*=5 donors) and IPF patients (*n*=5 donors), expanded to passage #3

as previously described, were seeded into each well at a density of 2.5 x 10^3 cells. After a period of 30 minutes to adhere to the culture plates, the cells were imaged with a 4x objective on the JuLI™ Stage Real-Time Cell History Recorder every 15 minutes for 48 hours. The JPEG images generated were imported to Nikon NIS Elements v.4.50 for analysis. The images were saved as a new sequence in the Nikon '.nd2' format, and pixel size and time step were updated to match the JuLI™ Stage system.

2.21.2 Basal cell migration through CDMs

Decellularized CDM was generated in 24-well μ-plates (Ibidi) as described in Section 2.20. The plate was briefly centrifuged (Hettich Centrifuge, Rotina 420R) at 1200rpm for 1 minute to ensure the CDM was adherent to the bottom of the well, and 200μl Airway Epithelial Cell Growth media (PromoCell) was added. PBECs were expanded as described in Section 2.3.2. The cells were passaged (P3) and the cell suspension incubated with CellTracker™ Deep Red Dye staining solution (Invitrogen #C34565, 1:1000 dilution in 1ml serum-free DMEM) for 40 minutes at 37°C in a Falcon tube. The cells were centrifuged (1200rpm, 5 minutes, 4°C), staining dye removed and resuspended in Airway Epithelial Cell Growth media (PromoCell) for counting. Basal cells were added at a concentration of 1.0 x 10^4 cells per well ensuring an even distribution using the light microscope. The cells were incubated for 2 hours (37°C, 5% CO_2) to allow the cells to adhere to the CDM. The fluorescently labelled basal cells were imaged on the CDMs in collaboration with David Gaboriau (Microscopy Specialist, Facility for Imaging by Light Microscopy, Imperial College, UK). A Zeiss Axio Observer Z1 Inverted Widefield Microscope with Lumencor SpectraX LED illumination and Zen Blue acquisition software was used to image the cells. The microscope chamber was heated to 37°C and maintained with 5% CO_2.

After calibrating the plate, Köhler illumination was adjusted, and positions in each well were added manually. Images were captured in the far-red channel (excitation 625–655 nm, emission 665–715 nm, for CellTracker DeepRed) and in transmitted light, using a Plan-Neofluar 10x/0.30 Ph1 objective. The system had a Hamamatsu Flash4.0 camera, and 2x2 binning was used, giving a 1024x1024 pixel image with a pixel size of 1.30 μm. Two positions with good cell coverage were manually selected per well, and their x, y and z coordinates saved. A Z-stack of five slices with a step of 7 μm was

acquired for each position, covering 28 μm in total. The stacks were centred on the z value saved when selecting the position, and these Z-stacks were acquired every 5 minutes for 24 hours. To minimise subtle displacement of the delicate matrix in the well associated with stage-related turbulence in the culture media, the stage speed was reduced to 40%.

At the end of the experiment the CDMs with adherent basal cells were fixed in 10% formalin for 15 minutes at room temperature, washed in PBS and stored at 4^{o}C for subsequent immunofluorescent imaging.

2.22 **Cell tracking analysis using Nikon NIS software**

2.22.1 Tracking basal cell migration on ECM substrates

Cell tracking analysis was carried out in collaboration with David Gaboriau (Microscopy Specialist, Facility for Imaging by Light Microscopy, Imperial College, UK). The JPEG Images generated by the JuLI real-time cell history recorder were imported to Nikon NIS Elements to create an image sequence. The sequence was calibrated (1.13μm/ pixel) and cropped to 48 frames (12hrs). This duration was selected because beyond 12hrs, automated tracking became less reliable as images were crowded with proliferating cells. To remove any artefact associated with microscope drift, the sequence was aligned to the previous frame. A median filter with a kernel size of 9 was applied to improve accuracy of spot detection, as it blurred a large proportion of the debris and maximised the detection of cells. Spot detection using the brightfield channel was set to detect dark spots with a typical diameter of 27μm and with a contrast setting of 8. Bright objects were removed. Once dark spots were detected, automated tracking was performed by the software algorithm 'Track Binaries' and then manually inspected as a means of quality control. False tracks such as those associated with debris or those which jumped to nearby cells were removed from analysis. If overall tracking was deemed less than 90% accurate on the basis of manually checking the tracks, the whole 12 hrs sequence was excluded. The raw cell migration data generated by Nikon NIS software was exported to MotilityLab **(**http://www.motilitylab.net/**)** an online resource allowing the quantitative and statistical

analyses of cell tracks, and GraphPad Prism 8 for further analysis and presentation. Only cells tracked for the entire 12-hour duration were included in the final analysis.

For each cell the following tracking measurements were reported;
(1) track length is the sum of all track segment distances.
(2) track speed is the track length divided by the time elapsed from the beginning to final position.
(3) straightness ratio (directionality ratio), calculated by the straight-line distance between the start and end point of the cell track *(d)*, divided by the track length *(D)*. A straight cell trajectory equates to a straightness ratio of 1, a curved trajectory equates to a ratio of 0.
(4) mean squared displacement (MSD), based on particle movement theory, can be used to characterise the type of motion displayed by cells[264]. MSD measures the average distance travelled by a cell over a specified time interval. NikonNIS-Elements calculates MSD as a sum of the squared distances from the beginning of the track. A more detailed explanation of the calculation of MSD is described by Wang et al.[264]

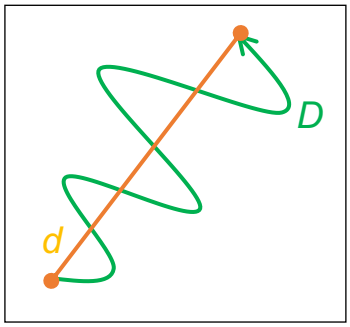


Figure 2.6 Calculation of straightness ratio. *d* – straight line distance between the start and end point of the cell track, *D* – track length.

2.22.2 Tracking basal cell migration through CDM (3D)

Images captured using the Cy5 and brightfield channels of the Zeiss Axio Observer were pre-processed in Icy[265], an open source bioimage analysis software. In a custom-made automated Icy protocol, the image sequences were cropped to 12 hours (144 frames, 5- minute intervals) and a maximum intensity projection of the Z-stack (7μm slices) performed. The resulting cropped and projected sequence was saved as a .TIFF file and imported into Nikon NIS Elements software for tracking analysis. The image sequence was calibrated (1 pixel = 1.30 μm, and T-step = 300secs) and images aligned to previous frame to correct for drift. Spot detection was performed on the Cy5

channel with the detector set to detect bright, circular objects of different sizes of typical diameter 27 μm with contrast set to 140. Dark objects were removed. Automated tracking was optimised using a random motion, constant speed model. A maximum gap size of 3 was allowed between tracks. Tracks were terminated when no object was closer than 2 times the standard deviation from a predicted position. A maximum object speed of 100 μm/s was set to exclude rapidly moving floating debris. Tracks with 12 or less frames were deleted. The automated tracks were manually checked for accuracy. The raw cell migration data generated by Nikon NIS software was exported to MotilityLab an online resource allowing the quantitative and statistical analyses of cell tracks, and GraphPad Prism 8 for further analysis and presentation. Only cells tracked for the entire 12-hour duration were included in the final analysis.

2.23 **MS-based proteomics of cell-derived matrices**

2.23.1 Isolation of enriched cell-derived ECM and MS sample preparation

This method was used to reduce the complexity of protein samples for MS analysis by removing cellular components and enriching for ECM proteins[266]. All steps were carried out at 4°C to minimise proteolysis. Primary human lung fibroblasts from 6 donors (*n* = 3 controls, *n* = 3 IPF) were cultured in a 10cm petri dish and maintained for 21 days in DMEM (Gibco) and 10% FBS to allow matrix deposition. Following a PBS wash, the cells were incubated for 30 minutes in 1ml extraction buffer (10 mM Tris, 150 mM NaCl, 1% [vol/vol] Triton X-100, 25mM EDTA, protease inhibitor in MilliQ Ultrapure water) to solubilise cellular proteins, and samples were then centrifuged at 14,000xg for 10 minutes. The remaining pellet was incubated for 30 minutes in 0.5ml alkaline detergent buffer (20 mM NH4OH and 0.5% [vol/vol] Triton X-100 in PBS) to further solubilise cellular proteins and disrupt cell–ECM interactions. Samples were then centrifuged at 14,000xg for 10 minutes. The pellet was then incubated for 30 minutes at room temperature, then for 20 minutes at 65ºC in 100µl DNase buffer (0.4µg Benzonase Nuclease (Millipore) in 2.4ml PBS) to degrade DNA. Samples were centrifuged at 14,000xg for 10 minutes. To yield the ECM fraction, the pellet was resuspended in 5x reducing sample buffer (100mM Tris pH 6.8, 25% [wt/vol] glycerol, 10% [wt/vol] SDS, 0.1% bromophenol blue, and 10% [wt/vol] β-mercaptoethanol in MilliQ Ultrapure water) and heat denatured at 70 ºC for 20 minutes. The samples were then centrifuged at 14,000xg for 10 minutes. The remaining pellet was resuspended

in S trap lysis buffer (5% SDS with 50 mM TEAB pH 7.5) and transferred to a 130 µl AFA tube (Covaris Inc, Woburn, USA,Covaris) to yield the ECM fraction.

2.23.2 Western blotting

Proteins were resolved based on molecular weight using SDS-PAGE. Briefly, 2µl of 5x SDS-PAGE loading buffer was added to 10µl of sample, incubated for 10 mins at 95^{o}C and loaded onto a precast NuPAGE 4-12% BisTris gel (ThermoFisher). The gel was run at 200V for 40-50 minutes. The proteins were transferred (30V for 70 minutes) to a nitrocellulose membrane (Whatman). Instant*Blue* (Sigma-Aldrich) Coomassie staining was used to visualise the proteins. Membranes were blocked in casein blocking buffer overnight at 4^{o}C, then probed with primary antibodies to Lamin B1 (#12586; Cell Signalling), fibronectin (#F3648; Sigma-Aldrich) and α-tubulin (#T6199; Sigma-Aldrich) diluted in PBS with 0.05% (vol/vol) Tween20 (PBS-Tween) for 1 hour. Membranes were washed in PBS-Tween before adding secondary antibodies; goat anti-rabbit IgG (H+L) Alexa Fluor 680 (#10585543; ThermoFisher), anti-mouse IgG (H+L) DyLight 800 4X PEG conjugate (#5257; Cell Signaling) diluted in Milli-Q Ultrapure water for 1 hour. Membranes were washed in PBS-Tween then scanned using an Odyssey infrared (IR) imaging system (LI-COR Biosciences).

2.23.3 MS data acquisition

The resulting ECM-enriched fractions were prepared using the S-Trap™ proteolytic digestion protocol carried out by Maryline Fresquet (Lennon Lab, Wellcome Trust Centre for Cell-Matrix Research, Manchester, UK). Briefly, the proteins were subjected to reduction and alkylation then further denatured by acidification and exposure to methanol. The samples were bound to the S trap columns via centrifugation, and the captured proteins digested using trypsin (1:10 enzyme:protein ratio). Digested peptides were eluted and submitted for mass spectrometry analysis. Data were acquired by LC-MS/MS using an UltiMate® 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA) coupled to a QE HF (Thermo Fisher Scientific, Waltham, MA) mass spectrometer.

2.23.4 MS data analysis

Raw MS files were imported into Proteome Discoverer v2.4.1.15 SP1 (PD) (Thermo Scientific), a program that integrates all the different steps in a quantitative proteomics experiment (including MS/MS spectrum extraction, peptide identification and quantification) into automated workflows. The data were analysed using the standard Sequest™ HT-Percolator workflow and the default consensus workflow for label-free quantification. The Sequest HT search was performed against the Homo sapiens database (SwissProt and TrEMBL version 2017-10-25, taxonomy 9606, 42252 and 124124 sequences respectively).

The hierarchical clustering of the proteomic data and the principal component analysis plots were computed and visualised using ClustVis (https://biit.cs.ut.ee/clustvis/). Proteome Discoverer uses the ANOVA method to calculate the statistical significance of differential protein expression between sample groups, with a lower p-value indicating the abundances are likely to be statistically different. The Benjamini-Hochberg procedure was used to adjust for multiple comparisons. Protein interaction network analysis was performed using STRING v11 **(**https://string-db.org/**).** STRING is a database of known and predicted protein-protein interactions. The list of matrix proteins identified, and their respective fold change (based on calculated abundance ratios) were submitted to the search under proteins with values/ranks. The proteins were mapped onto the *Homo sapiens* network, with a FDR stringency of 1%. The active interaction sources selected were experiments and co-expression databases and the network edges represents the confidence (strength of data support). A medium confidence score of 0.4 was applied.

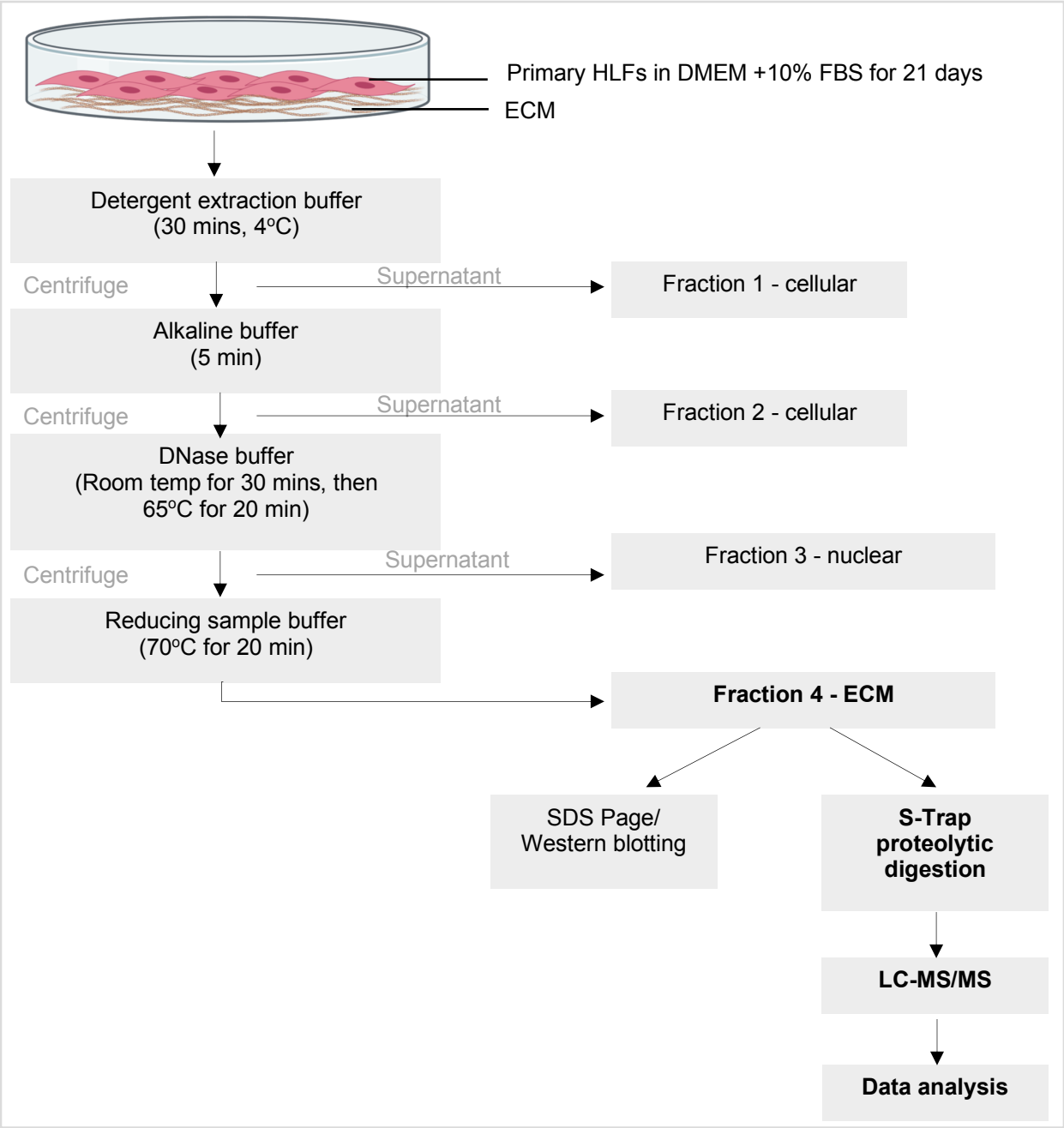


Figure 2.7 Proteomics workflow for human lung fibroblast -derived matrices.

Illustration created with https://biorender.com/.

2.24 **Statistical analysis (excluding transcriptomic and proteomic data)**

Statistical analysis of transcriptomic and proteomic data are described in the relevant sections. For all other data, statistical analysis was carried out using Prism v9 (GraphPad Software, LLC.). To determine if the data was from a Gaussian distribution the D'Agostino-Pearson omnibus normality test was performed. This test calculates the skewness and kurtosis of the data to quantify how far the distribution is from Guassian. If the p- value returned for this test was high (>0.05), it was assumed the data were sampled from a Gaussian distribution and a parametric test was used for statistical analysis; for unpaired data this was an unpaired t test, and for paired data this was a paired t-test. If the normality test p-value was ≤0.05, the null hypothesis was rejected and therefore data were not sampled from a Gaussian distribution so a nonparametric test was used; for unpaired data this was a Mann-Whitney test and for paired data this was a Wilcoxon matched-pairs signed rank test. For cell migration experiments where multiple cells from multiple donors were studied under different experimental conditions, a nested one-way ANOVA with Tukey's multiple comparison test was used. For quantitative analysis of SHG data acquired from different lung regions from multiple IPF and control donors, an ordinary one-way ANOVA with Tukey's multiple comparisons test was performed. For correlation analyses, Pearson correlation coefficients were computed and a two-tailed P value, ≤0.05 considered significant. The statistical test used for each analysis are described in the accompanying figure legend and significant differences noted by asterisk(s): *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

No *a priori* power calculations were performed for the group numbers used in the experiments described in this thesis. It was an exploratory study using primary human cells therefore the sample size was determined by the feasibility of recruitment and sampling in the available timeframe. Primary cells are a limited resource therefore it was not possible to perform pilot experiments to guide a power calculation. For the *in vitro* work, sample sizes were chosen based on the numbers that had shown a biologically meaningful difference in similar cell culture experiments in the lab. Power computation for transcriptomic studies is more complex and there is no standardised approach. Sample size was based on the number of subjects included in other RNA-

seq studies in which significant differences were detected. The single-cell data from this project will be integrated with Martijn Nawijn's "discovAIR Lung Cell Atlas" dataset which will satisfy sample size requirements.

CHAPTER 3

In vitro phenotyping and functional studies of the IPF airway epithelium

3 *In vitro* phenotyping and functional studies of the IPF airway epithelium

3.1 Background and rationale

Investigations into the role of epithelial dysfunction in IPF, have predominantly focused on alveolar type II epithelial (AT2) cells, situated at the epicentre of disease pathology[267]. Many of these studies have used immortalised cell lines such as adenocarcinoma A549 cells, given the practical challenges associated with isolating and culturing primary alveolar epithelial cells from fibrotic lung tissue specimens[268]. Few studies have used primary alveolar epithelial cells[269] or performed *in vitro* characterisation of other lung epithelial cell types such as those involved in 'bronchiolisation' and honeycomb cyst formation in the IPF distal lung[187]. A growing body of evidence suggests functionally distinct epithelial cell types throughout the respiratory tract play an integral role in mucosal immunity and repair[52]. We sought to better characterise airway epithelial cells from IPF patients which, although situated proximally in the respiratory tract, share the same genetic and environmental exposure risk factors for the disease as AT2 cells.

For diseases with airway-focused pathology such as asthma and COPD, air-liquid interface (ALI) cultures have been utilised as a physiologically relevant tool to study the phenotype and function of fully differentiated airway epithelial cells[270-273]. Culturing primary bronchial epithelial cells (PBEC) at an air-liquid interface on microporous membranes in Transwells promotes basal cell differentiation. Despite differences in the microenvironmental milieu *in vitro*, PBEC-ALI cultures generate a mucociliated, pseudostratified epithelium modelling the cellular composition found *in vivo*[274]. Studies using human nasal and bronchial epithelial cells have demonstrated a good correlation between the transcriptome of cells collected directly from the airway and those cultured and differentiated at ALI *in vitro*[275,276]. Disease-specific cell-intrinsic properties are retained in epithelial ALI cultures for example, asthmatic bronchial epithelial cells demonstrated a deficient interferon- beta response to rhinovirus-16 infection *in vitro*, compared to healthy controls[277]. This suggests *in vivo* epigenetic programming of epithelial cells is preserved *in vitro*[271]. PBEC-ALI culture models, have been used to study the airway epithelial cell response to diverse environmental exposures including influenza virus[230], SARS-CoV-2[278], cigarette smoke[279] and diesel

exhaust[280]. Our understanding of epithelial cell differentiation[204], tight junction mediated barrier integrity[272,281], wound repair[282] and cilia function[283] have advanced through use of PBEC-ALI culture models.

PBEC- ALI cultures do not capture the complex interactions that occur with other key cell types known to orchestrate immune responses and fibrosis in the airway microenvironment, such as fibroblasts and macrophages. Co-culture models provide an opportunity to dissect the interplay between different cell types in a particular niche[271]. In these co-culture models, cells can make direct physical contact, or interact indirectly through soluble mediators. A combination of primary cells and cell lines have been used for the accompanying cell types in COPD[284] and IPF[285] co-culture models. There have been no co-culture models designed specifically to examine the role of airway epithelial cells with interacting cell types in IPF. Correll et al. studied primary AT2 cells on floating collagen gels which contained lung fibroblasts, showing poorly differentiated AT2 cells suppressed fibroblast matrix and growth factor gene expression[286]. Evidence of disturbed epithelial barrier integrity with disorganised zonula occludens-1 (ZO-1) tight-junctions and increased paracellular permeability was discovered in an indirect co-culture model using normal primary AT2 cells and IPF lung fibroblasts[287].

In this Chapter, advanced microscopy techniques coupled with cell culture models are used to study primary airway epithelial cells derived from IPF patients and healthy controls under homeostatic conditions, and during injury and infection.

3.2 **Hypothesis**

There are intrinsic differences in cultured IPF airway epithelial cells compared to healthy controls, that lead to impaired barrier integrity and dysregulated immune responses.

3.3 **Aims**

1. Isolate primary airway epithelial cells from bronchial brushings taken from IPF patients and healthy controls and establish a functional ALI culture model.

2. Perform phenotypic characterisation of cells isolated directly from airway brushings and in cell culture using microscopy techniques.

3. Evaluate epithelial repair responses to mechanical and viral injury.

4. Explore cell-cell interactions in IPF using indirect and direct co-culture models and human precision cut lung slices (PCLS).

3.4 **Results**

3.4.1 Phenotyping airway epithelial cells from bronchial brushings using microscopy techniques

Before studying IPF airway epithelial cells *in vitro*, microscopy techniques were used to identify the epithelial cell types harvested by bronchial brushings. To examine cell morphology and ultrastructure, transmission electron microscopy (TEM) was carried out by Andrew Rogers (EM Department, Royal Brompton Hospital) using two freshly harvested bronchial brushings from a patient with pulmonary fibrosis. This revealed preserved epithelial strips displaying a typical pseudostratified organisation in line with previous reports for the normal human airway[288,289]. Ciliated columnar epithelial cells could be identified with multiple cilia projecting into the airway lumen (Fig. 3.1 A and B). Cilia could be seen arising from basal bodies situated in the cytoplasm beneath the apical surface membrane and displayed a normal (9+2) microtubular arrangement. As previously described[288], these cells had abundant mitochondria, reflecting the high-energy demands of motile cilia. This feature is noteworthy since it is relevant to the single-cell RNA-seq analysis described in Chapter 5. Ciliated cells were juxtaposed to secretory cells containing large, mucous granules with mitochondria and other organelles situated at the periphery of the cell. Basal cells were ovoid shaped, did not project to the luminal membrane and contained a large nucleus relative to the cell cytoplasm, consistent with previous reports[290]. Desmosome intercellular junctions

between epithelial cell types were present, and also cytokeratin filaments were observed in the cytoplasm of basal cells. The basement membrane, sub-epithelial structures and stromal cells were not identified because unlike bronchial biopsies, brushings collect the more superficial cell layers[179]. Overall, there did not appear to be any distinct morphological abnormalities in airway epithelial cells from the subject sampled. Although TEM provided unparalleled ultrastructural detail of the cell types present, it is a labour-intensive technique. Therefore, confocal microscopy was used to validate cell identities on cytospins prepared from IPF brushings (Fig. 3.2 A – D). This confirmed that brushings comprised predominantly of epithelial cells expressing the transmembrane glycoprotein; epithelial cellular adhesion molecule, EpCAM. There was a smaller subset of cells which did not stain positive for EpCAM and were likely immune cell contaminants, a finding which is explored further in Chapter 5. Cilia were identified by positive staining to alpha-tubulin, and basal cells stained positive for cytokeratin-5. Detailed cellular phenotyping by confocal microscopy and other techniques such as flow cytometry, is restricted by the antibodies available to epithelial cell markers. To overcome this issue and complement the microscopy data presented in this section with comprehensive molecular phenotyping of cell populations on IPF airway brushings, single-cell RNA-sequencing was carried out and is discussed in Chapter 5.

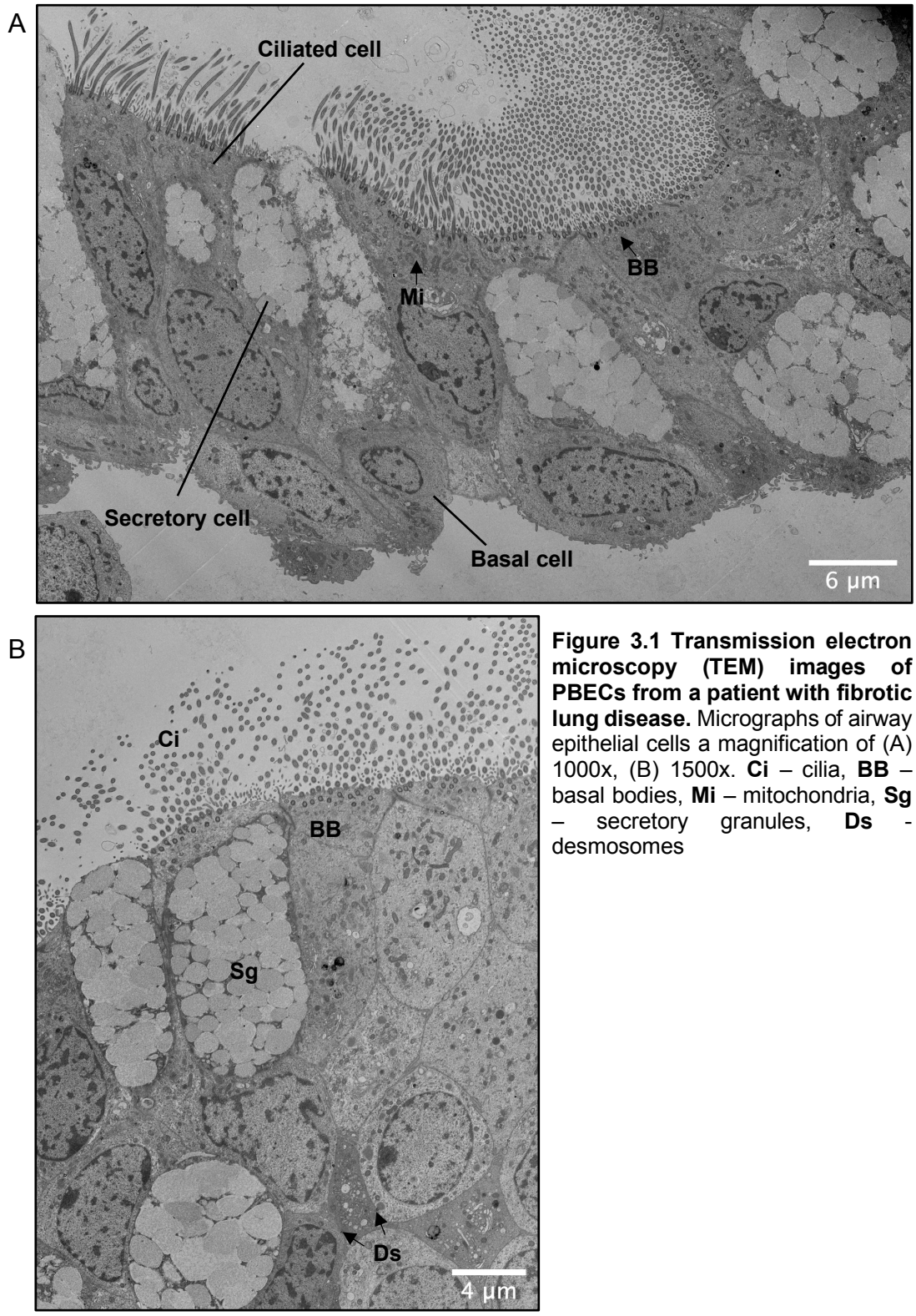


Figure 3.1 Transmission electron microscopy (TEM) images of PBECs from a patient with fibrotic lung disease. Micrographs of airway epithelial cells a magnification of (A) 1000x, (B) 1500x. **Ci** – cilia, **BB** – basal bodies, **Mi** – mitochondria, **Sg** – secretory granules, **Ds** - desmosomes

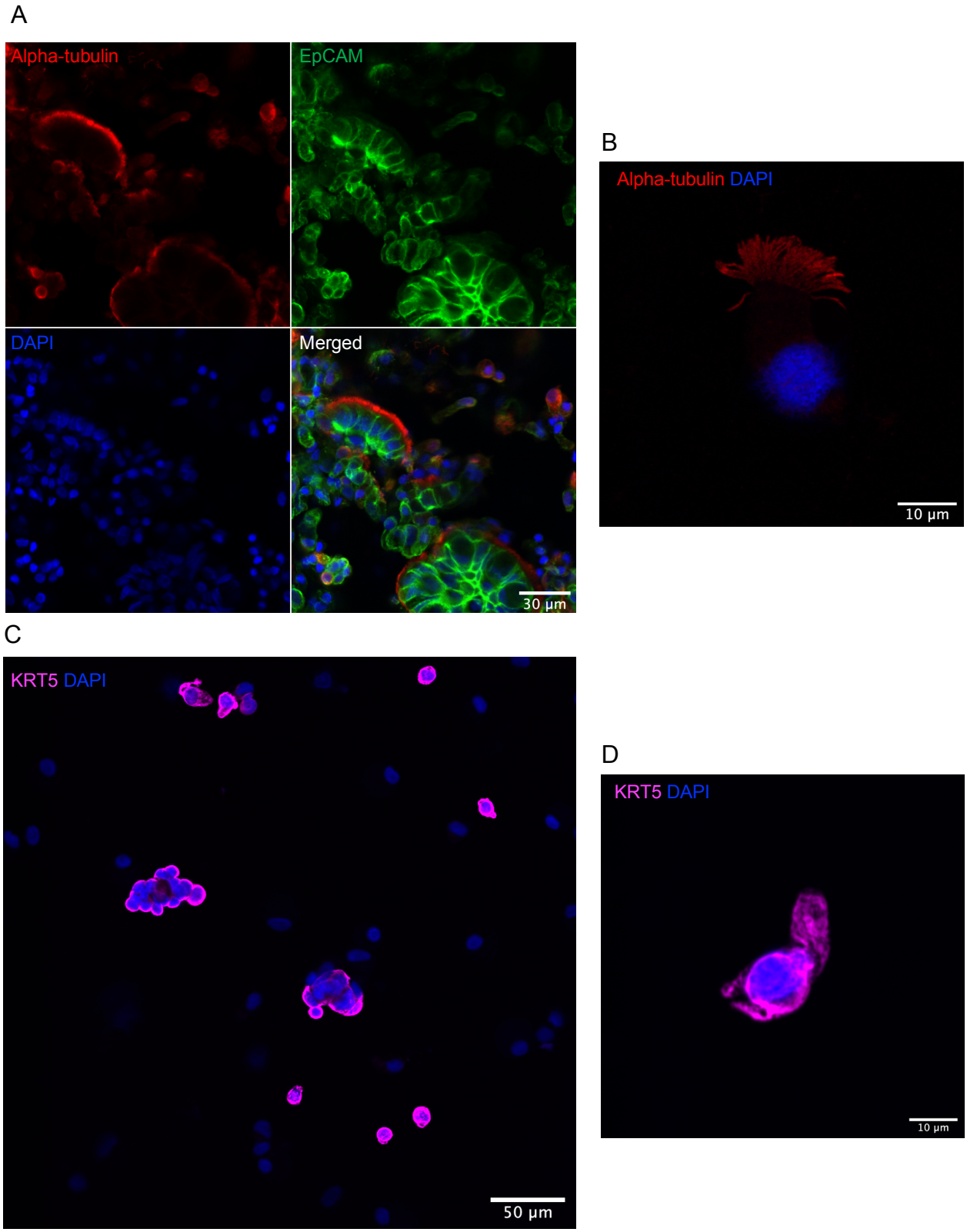


Figure 3.2 Immunofluorescence microscopy of IPF bronchial brush cytospin preparations. (A) Strips of ciliated epithelial cells staining positive for EpCAM (green) and cilia marker, alpha-tubulin (red) (B) High-resolution image of a single ciliated epithelial cell (C) Clusters of basal cells staining positive for marker, cytokeratin 5 (magenta) (D) High-resolution image of a single basal cell. Images acquired using Leica SP5 confocal microscope. Images representative of cytospins from n = 3 subjects.

3.4.2 Culture characteristics of IPF airway epithelial cells

PBECs isolated directly from airway brushings were expanded in a T25 flask to passage 1 (P1), then a T75 flask to passage 2 (P2) in airway epithelial growth media prior to cryopreservation. Cells were passaged when they were 80% confluent in the flask. IPF PBECs took a greater number of days to reach the confluency required for passaging at P1 and P2 (Fig. 3.3 A). There was no difference in total cell number between IPF and healthy donors at each passage (Fig. 3.3 B). The baseline characteristics of the donors are noted in Table 2.1 (Chapter 2). A possible explanation for the delay in reaching 80% confluency by IPF PBECs, was a reduction in cell proliferative capacity. Therefore, under identical culture conditions using PBECs derived from a separate cohort of IPF and healthy subjects (detailed in Table 5.1), a JuLI^{TM} Stage Real-Time Cell History Recorder was used to measure cell proliferation over 48 hours. Change in cell confluence over time was calculated by JuLI^{TM} Stage software using images captured every 15 minutes. Cell seeding density was consistent between all donors, and the relative change in confluence was calculated from the starting percentage confluence. Using this approach to precisely quantify PBEC proliferation over a 48-hour period, there was no demonstrable difference in proliferation between IPF and healthy PBECs under homeostatic conditions (Fig. 3.4). Of note, PBECs were used at passage 3 for this assay, and had been previously cryopreserved following the conventional cell expansion approach described above.

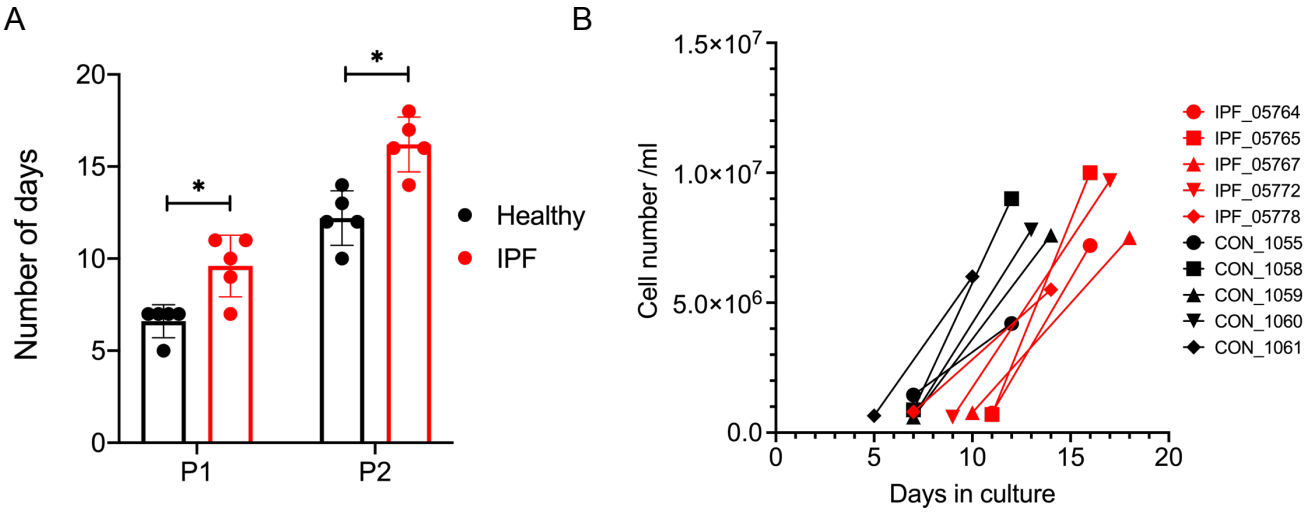


Figure 3.3 PBEC culture characteristics. PBECs from IPF patients (*n*=5) and healthy controls (*n*=5) were isolated from two airway brushings and expanded in culture. Baseline characteristics shown in Table 5.1, Chapter 5. (**A**) Number of days to reach 80% confluence for passage 1 and passage 2. (**B**) Cell number counted per donor at passage 1 and 2, represented according to the number of days in culture. Statistics: Two-tailed, Mann-Whitney test; p <0.05. Mean ± SD.

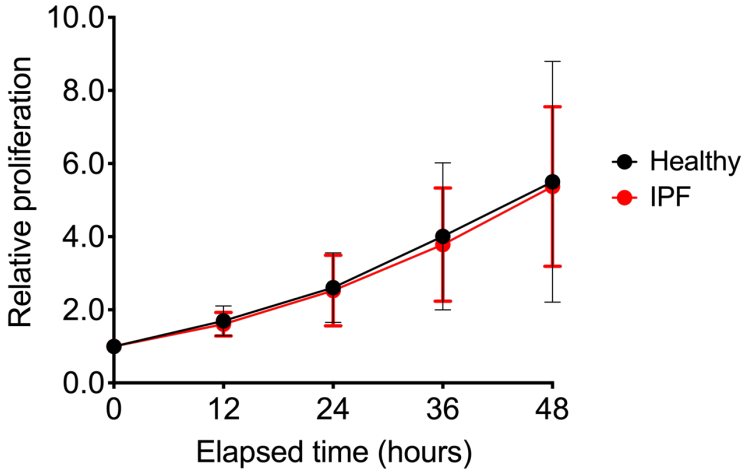


Figure 3.4 PBEC proliferation assay. PBECs (passage 3) from IPF patients (*n*=6) and healthy controls (*n*=5) were cultured under normal conditions and cell confluence captured every 15 minutes using the JuLI™ Stage system. Only 12-hour datapoints are displayed for figure clarity. Mean ± SD.

3.4.3 Epithelial wound healing assay

An intact repair response following injury is critical for airway epithelial homeostasis[291]. Loss of alveolar epithelial integrity is considered a salient event propagating the development of IPF[267]. Therefore, a wound healing assay[292] was performed, in which a linear wound was created in a confluent monolayer of PBECs, and wound closure was studied over 48hrs using the JuLI Stage Real-Time Cell History Recorder. A wound of consistent width (~550µm) was created by scratching the epithelial monolayer with a P1000 pipette tip. IPF (*n*=10) and healthy (*n*=5) PBECs were visualised every 15-minutes migrating into the cell-free area. Greater than 50% of the wound had been closed within 12hrs by IPF and healthy PBECs (Fig. 3.5). Overall, there was no statistically significant difference in wound healing capacity (assessed by two-tailed, unpaired t-test of area under the curve) observed between IPF and healthy PBECs. At 48 hours, all healthy PBECs had achieved over 90% wound closure. In contrast, there was a greater range of wound closure achieved by IPF PBECs, with subject ID_1290 reaching only 65% closure.

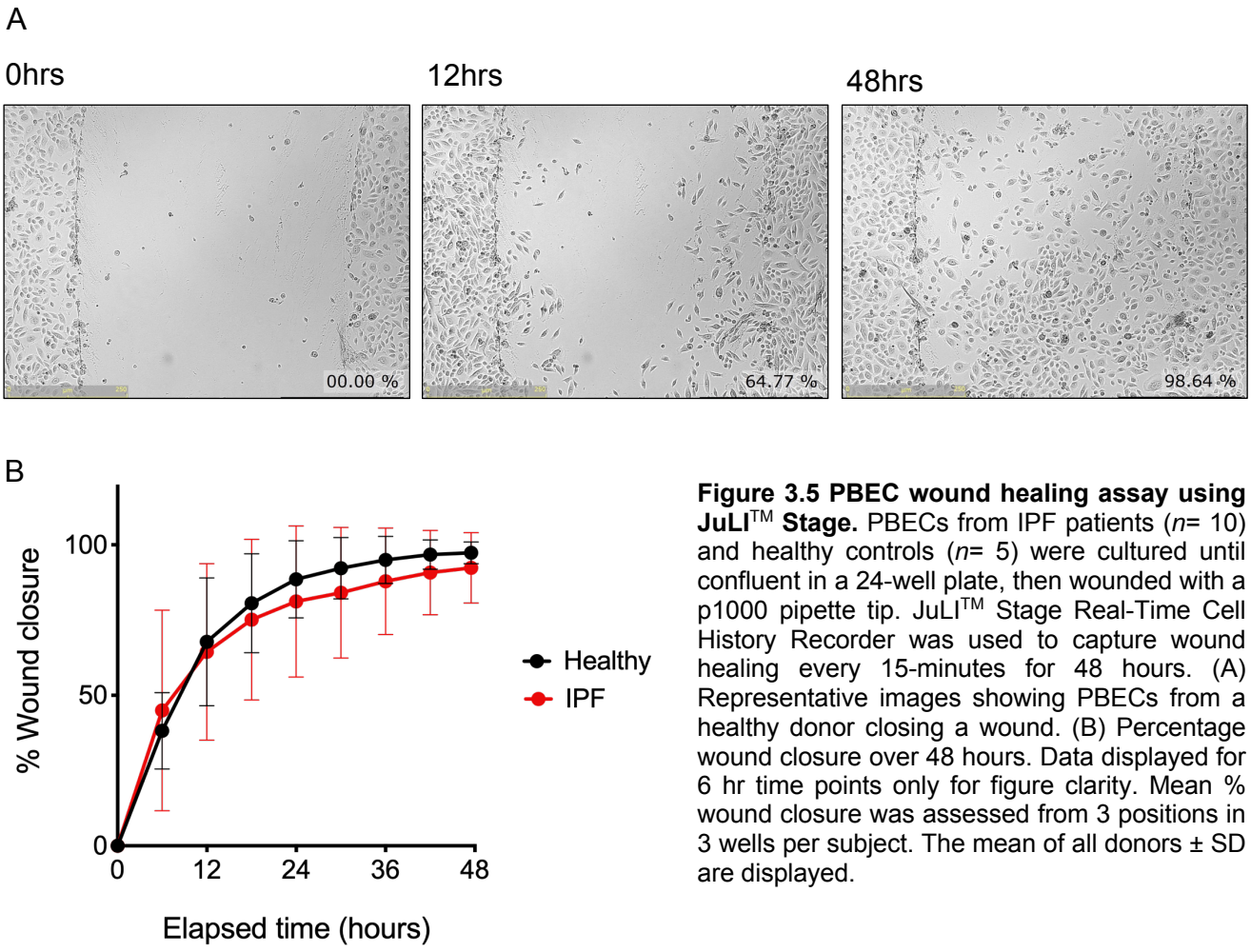


Figure 3.5 PBEC wound healing assay using JuLI™ Stage. PBECs from IPF patients (*n*= 10) and healthy controls (*n*= 5) were cultured until confluent in a 24-well plate, then wounded with a p1000 pipette tip. JuLI™ Stage Real-Time Cell History Recorder was used to capture wound healing every 15-minutes for 48 hours. (A) Representative images showing PBECs from a healthy donor closing a wound. (B) Percentage wound closure over 48 hours. Data displayed for 6 hr time points only for figure clarity. Mean % wound closure was assessed from 3 positions in 3 wells per subject. The mean of all donors ± SD are displayed.

3.4.4 IPF PBEC air-liquid interface (ALI) culture characteristics

To recapitulate *in vivo* phenotypic and functional characteristics of the human airway epithelial, PBECs were differentiated in air-liquid interface cultures. Mucus secretion was evident when performing apical washings, after around 7 days in culture, consistent with a previous report[293]. Epithelial integrity was measured using transepithelial electrical resistance (TEER), and at 21 days there was no significant difference between IPF and healthy PBECs (Fig. 3.6). Confocal microscopy showed normal appearances of tight junction protein, ZO-1, extensive cilia coverage and MUC5AC positive secretory cells (Fig. 3.7 A and B). These findings were further validated by H&E and PAS staining of ALI cross-sections showing mucociliary differentiation of the IPF PBECs (Fig. 3.7 C). Beating cilia could be visualised using a light microscope by day 21 in culture. Ciliary beat frequency, measured using digital high speed video imaging[229] was 11.85 Hz/CPS, comparable to adult reference values[294].

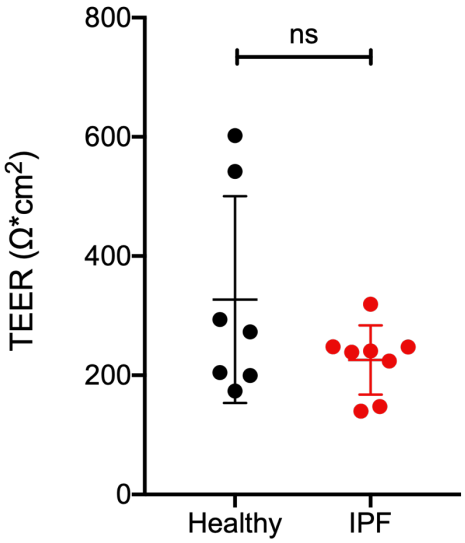


Figure 3.6 PBEC ALI transepithelial electrical resistance (TEER). IPF and healthy PBECs were differentiated in ALI cultures for 21 days. TEER was measured using an EVOM2 Epithelial Voltohmmeter**.** The mean TEER of 6 replicate wells was calculated. The final unit area resistance (Ω*cm^2) was calculated by multiplying the mean TEER value by the area of the insert membrane (0.33cm^2). Each dot represents a different subject. Mean ± SD are presented in this figure. A two-tailed, unpaired t-test was used for significance testing.

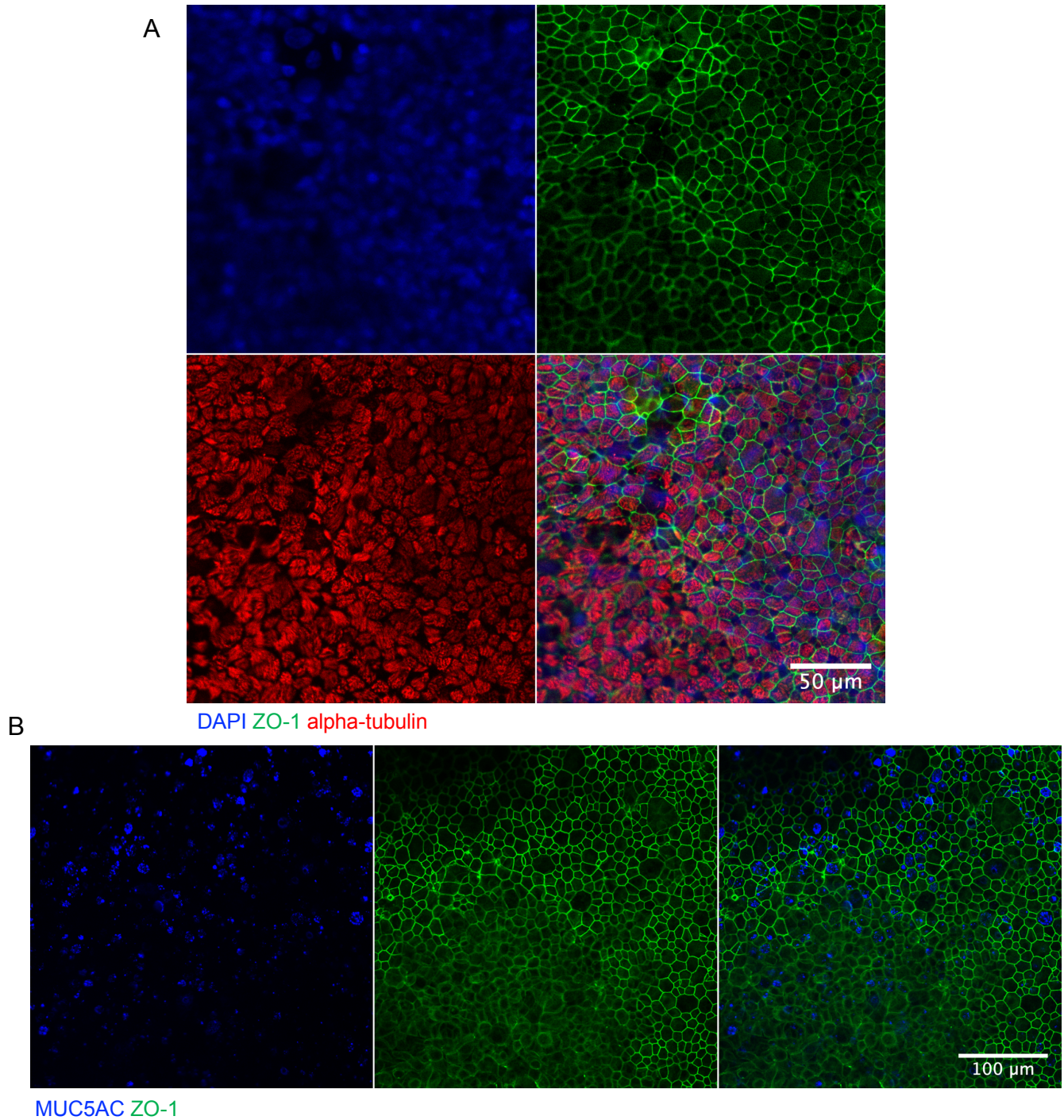


Figure 3.7 Characterisation of IPF PBEC ALI cultures. Representative images of IPF PBEC ALI cultures IF stained for (A) ZO-1 and alpha-tubulin (cilia), and (B) ZO-1 and MUC5AC. ALI cultures were agarose embedded, sectioned and stained with (C) H&E and (D) PAS by Lorraine Lawrence, NHLI. IF images acquired with a Leica SP5 inverted confocal microscope. Histology images obtained using a Leica DM2500 microscope.

3.4.5 Influenza virus challenge model

To test the barrier immune function and repair capacity of the fully differentiated IPF airway epithelium, PBEC ALI cultures were challenged with a common respiratory virus, influenza A. IPF (*n*= 8) and healthy PBECs (*n*= 7) were inoculated with a low-multiplicity of infection (MOI) (0.01)[231] of live influenza A virus/England/195/09 or UV-inactivated IAV (mock) (Fig. 3.8). Barrier integrity was assessed by performing TEER measurements every 24 hours. A reproducible fall in TEER was detected at 48 hours in all IAV- infected IPF and healthy ALI cultures, which recovered by 72 hours (Fig. 3.9 A). This finding was supported by immunofluorescence imaging demonstrating areas of discontinuity in the ZO-1 tight junction staining which were more pronounced at 48 hours (Fig. 3.9 B). Cytotoxicity, determined by measuring LDH release at 72 hours, was significantly increased in the IAV- infected ALI cultures compared to healthy controls (Fig. 3.9 C). There was no difference in the levels of cytotoxicity induced by IAV between IPF and healthy PBECs.

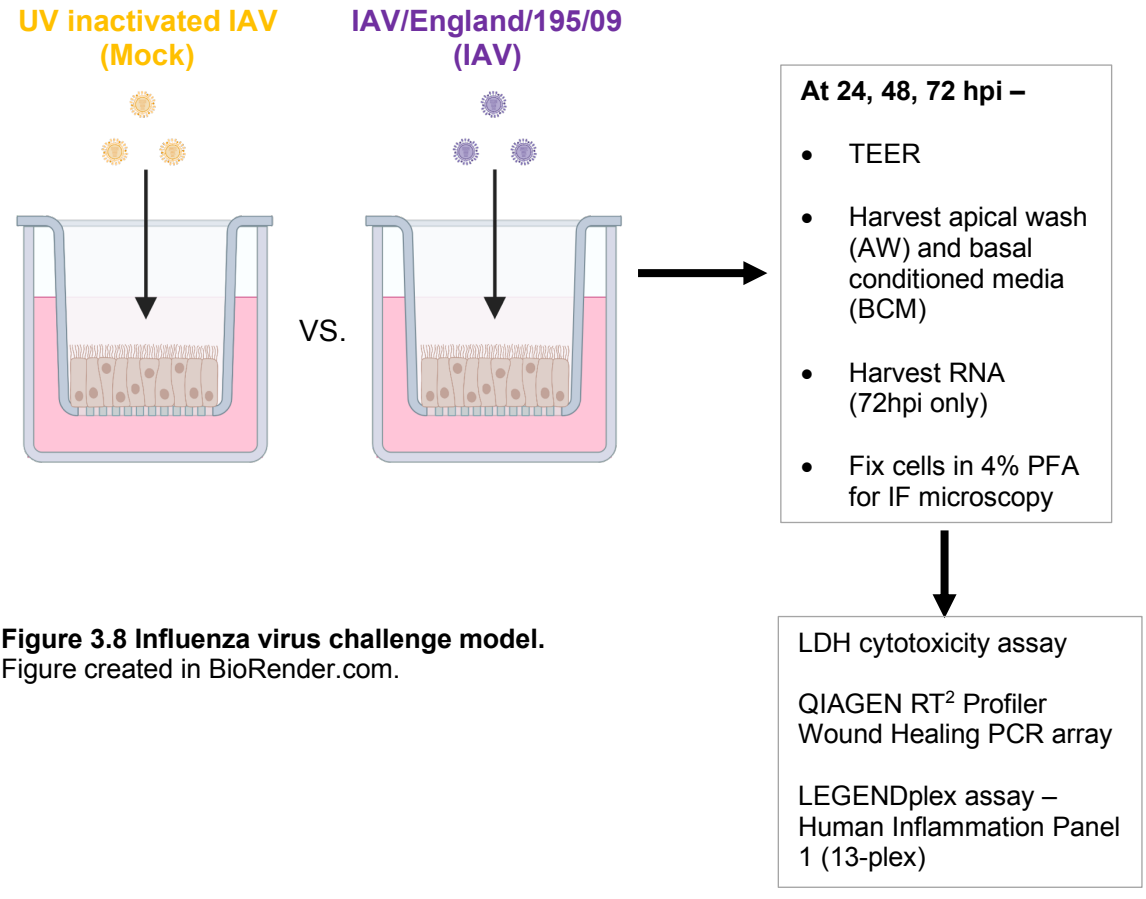


Figure 3.8 Influenza virus challenge model.
Figure created in BioRender.com.

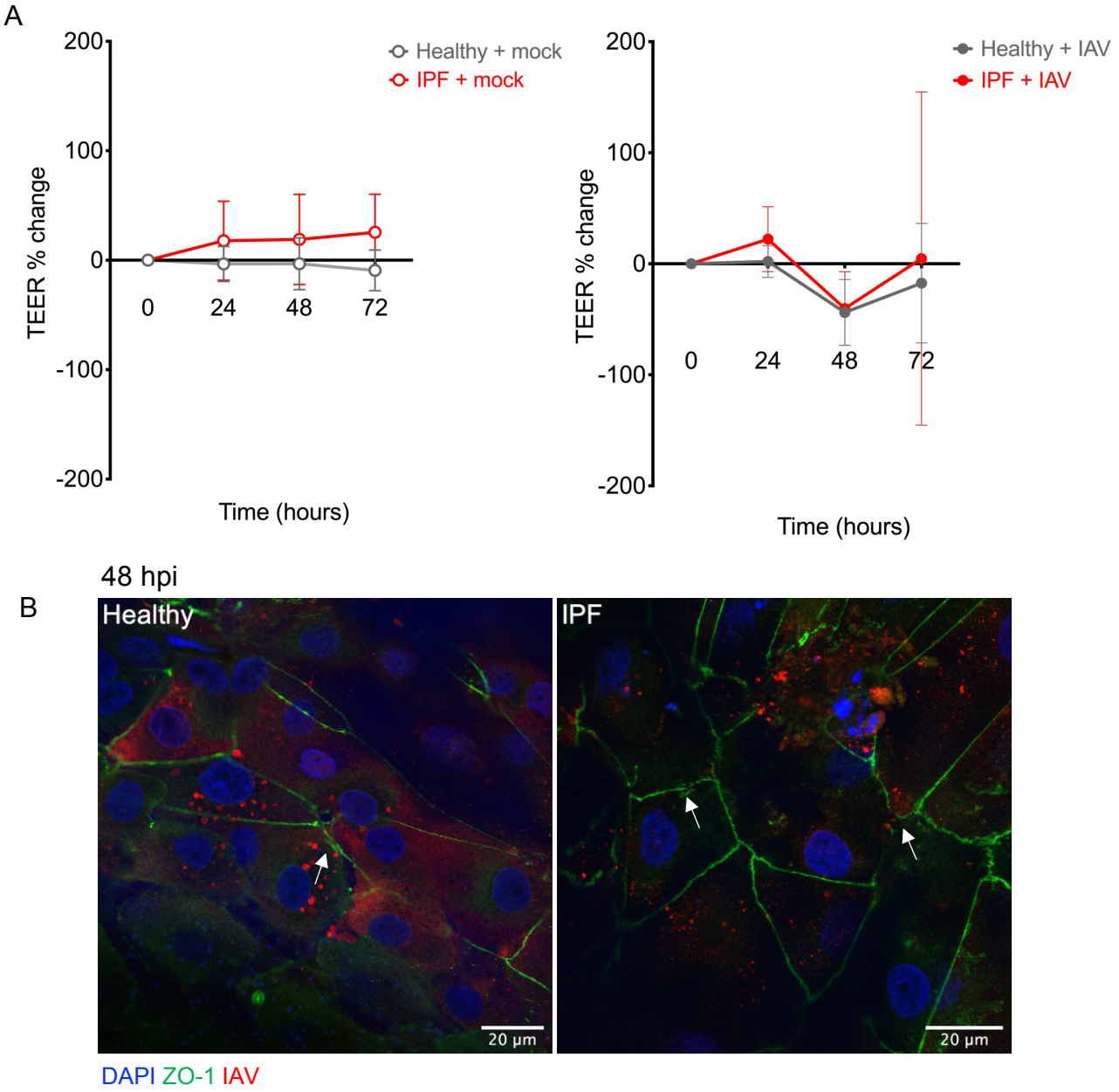


Figure 3.9 Influenza virus infection of IPF and healthy PBEC ALI cultures. PBECs from IPF patients (*n*=8) and healthy controls (*n*=7) were infected with IAV or mock and studied over 72 hours. (A) TEER presented as percentage change from baseline calculated using the mean of 6 replicate wells. (B) IF microscopy showing IAV- infected ALI cultures at 48 hours. White arrows indicated areas of discontinuous ZO-1 staining. (C) LDH cytotoxicity assay at 72 hours using basal conditioned media. Data presented as a percentage of 100% cell lysis control. Mean ± SD represented. Two-tailed, unpaired t-test used for significance testing. **P<0.01

A bead-based multiplex assay was used to quantify multiple inflammatory mediators released by PBECs into the apical washings (AW) and basal conditioned media (BCM) 72 hours post-infection. The LEGENDplex™ data analysis software, uses the PE signal fluorescence intensity detected by the flow cytometer to determine the concentration of a particular analyte based on a known standard curve (Appendix A, Figure 7.1). Mean cytokine concentration from triplicate wells for mock and IAV infected PBECs were reported. Very high levels of IL-8 were detectable in AW from mock and IAV- challenged cells, with no difference between IPF and healthy controls (Fig. 3.10 A). There was a slight increase in AW IL-6 in IAV infection compared to mock in IPF and healthy PBECs. These findings were mirrored in the BCM (Fig. 3.10 B) however BCM IL-6 was significantly increased in IPF and healthy PBECs ($P<0.05$). IL-8 release from IPF PBECs in BCM was higher than healthy PBECs with mock and IAV infection. There were relatively small increases in the concentrations of IL-1β, TNF-α, IL-33 and IL-18 with IAV infection in IPF and healthy PBECs. Although not statistically significant, IL-1β and TNF-α release in AW and BCM was attenuated in IPF compared to healthy PBECs 72 hours post-IAV infection. MCP-1 levels in AW were suppressed 72 hours post- IAV infection in IPF and healthy PBECs. With the exception of the IL-8 BCM results, these findings suggest the inflammatory mediator response to influenza is similar in IPF and healthy airway epithelial cells.

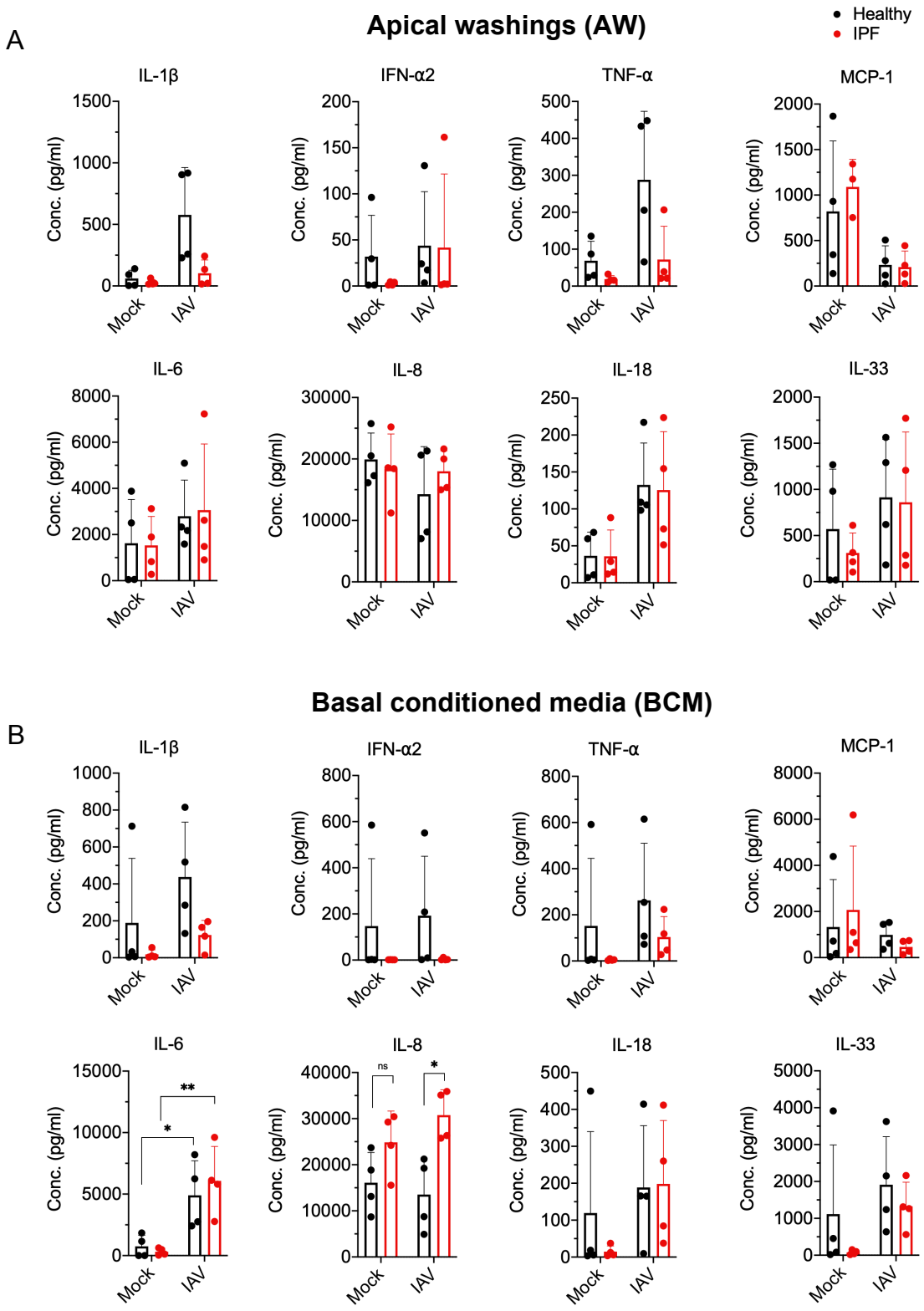


Figure 3.10 Inflammatory mediators produced by PBECs at 72 hours post-influenza A infection. A LEGENDplex™ immunoassay was used to measure 13 inflammatory mediators produced by IPF (*n*=4) and healthy (*n*=4) PBECs in ALI culture, 72 hours- post infection with IAV or mock. The concentration of analytes relevant to epithelial inflammation in (A) apical washings and (B) basal conditioned media from IPF and healthy control PBECs infected with mock or IAV. Each datapoint represents the concentration calculated from the average MFI of triplicate samples. Bars showing mean ± SD. A two-tailed, unpaired t-test was used for significance testing between IPF and control groups. *P<0.05

Next, wound healing pathways were interrogated with a PCR array using RNA isolated from IPF (*n*=6) and healthy (*n*=6) PBECs 72 hours post-IAV infection. There was no distinct clustering of IPF and healthy control samples based on the expression of 84-genes measured in the array (Fig. 3.11 A). There was a sub-cluster of 3 IPF samples (ID_5224, ID_1200 and ID_5359) which shared higher expression of genes encoding integrins and collagens (dashed box on Fig. 3.11 A). To help identify significant gene expression changes between IPF and healthy PBECs, fold change regulation and statistical significance a volcano plot was used (Fig. 3.11 B). This revealed significant downregulation of *CSF3,* encoding colony stimulating factor – 3 which controls granulocyte production, differentiation and function. Although not meeting the 2 fold-change threshold, integrin subunit alpha V, *ITGAV* was also significantly downregulated in IPF PBECs. There was a greater than 2-fold up-regulation of matrix remodelling genes; collagen type III alpha chain 1, *COL3A1* and matrix metallopeptidase 7, *MMP7* in IPF PBECs compared to healthy PBECs, although these changes did not reach statistical significance. Transgelin, *TAGLN*, a cross-linking protein involved in epithelial cell migration was also increased in IPF PBECs relative to healthy PBECs. There was a 2-fold down-regulation in cathepsin K, *CTSK,* a lysosomal cysteine proteinases with potent collagenolytic activity[295]; tumour necrosis factor, *TNF,* a key proinflammatory cytokine, and plasminogen, *PLG.*

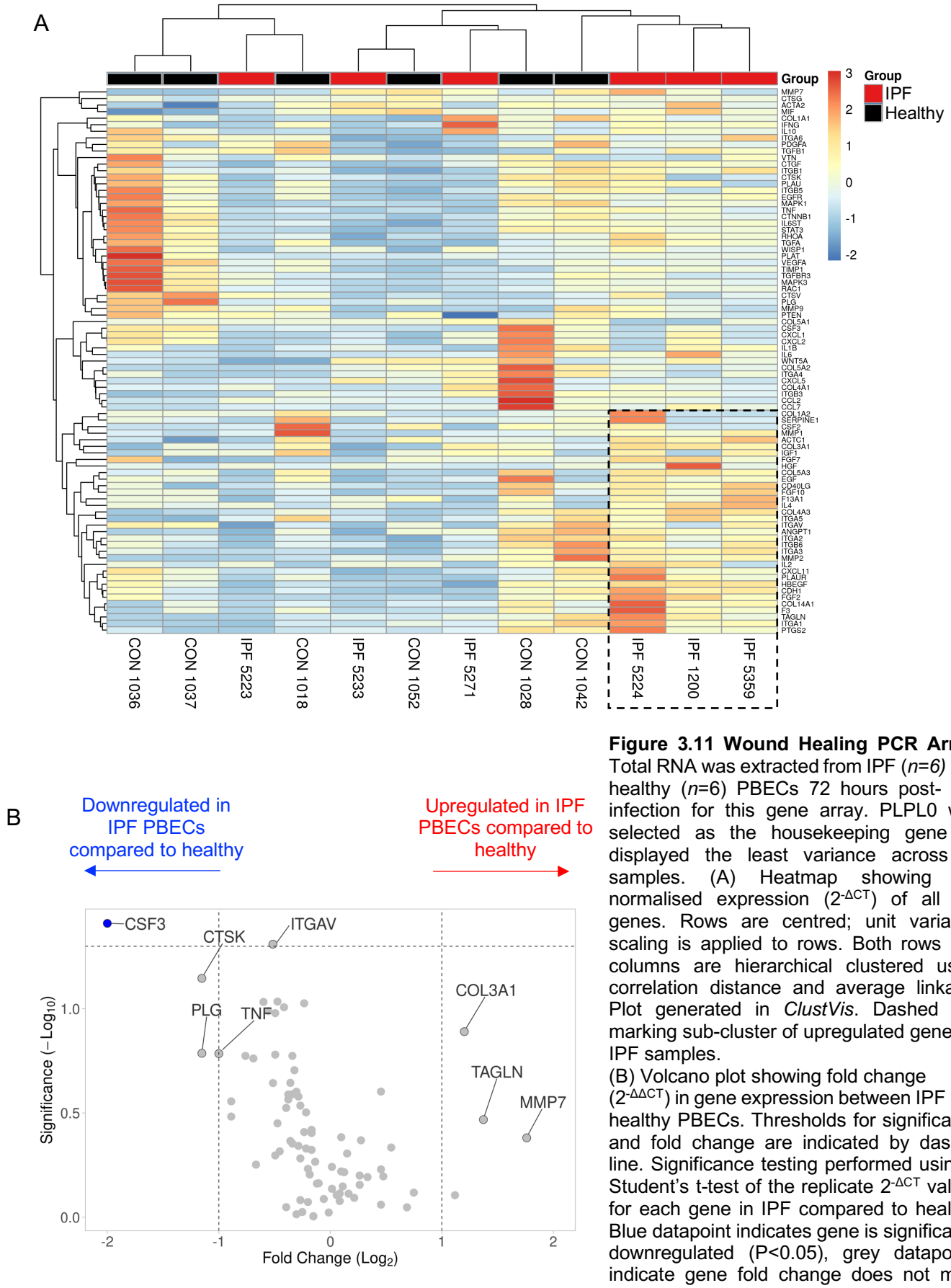


Figure 3.11 Wound Healing PCR Array. Total RNA was extracted from IPF (*n=6)* and healthy (*n*=6) PBECs 72 hours post- IAV infection for this gene array. PLPL0 was selected as the housekeeping gene as displayed the least variance across all samples. (A) Heatmap showing the normalised expression ($2^{-\Delta CT}$) of all 84-genes. Rows are centred; unit variance scaling is applied to rows. Both rows and columns are hierarchical clustered using correlation distance and average linkage. Plot generated in *ClustVis*. Dashed line marking sub-cluster of upregulated genes in IPF samples.
(B) Volcano plot showing fold change ($2^{-\Delta\Delta CT}$) in gene expression between IPF and healthy PBECs. Thresholds for significance and fold change are indicated by dashed line. Significance testing performed using a Student's t-test of the replicate $2^{-\Delta CT}$ values for each gene in IPF compared to healthy. Blue datapoint indicates gene is significantly downregulated ($P<0.05$), grey datapoints indicate gene fold change does not meet significance or fold change threshold. Plot generated in *VolcaNoseR*.

Dysregulated epithelial-fibroblast interactions in the distal lung have been implicated in fibrosis[296], however the precise nature of these interactions is unclear and have not been studied in the IPF bronchial epithelium. We therefore sought to determine if IAV-driven injury to IPF PBECs leads to dysregulated fibroblast activity. Conditioned media from IPF (n=3) and healthy (n=3) PBEC-ALI cultures infected with IAV was transferred to a wounded monolayer of normal primary HLFs (Fig. 3.12 A). Wound closure was monitored using the JuLI Stage system over 48 hours. In the presence of supernatant from IAV-infected IPF and healthy PBEC-ALI cultures, less than 50% fibroblast wound closure was achieved at 48 hours, compared to mock treated PBEC-ALI cultures where 50 – 85% was reached (Fig. 3.12 B and C). In the mock treated groups, there was a trend to greater wound closure with supernatant from healthy PBECs, however the reverse was true in the IAV treated groups with IPF PBEC supernatant promoting greater wound closure. The basal conditioned media used to treat wounded fibroblast monolayers were not UV-inactivated prior to use therefore it is not possible to discount a direct cytotoxic effect of IAV leading to impaired wound closure by HLFs.

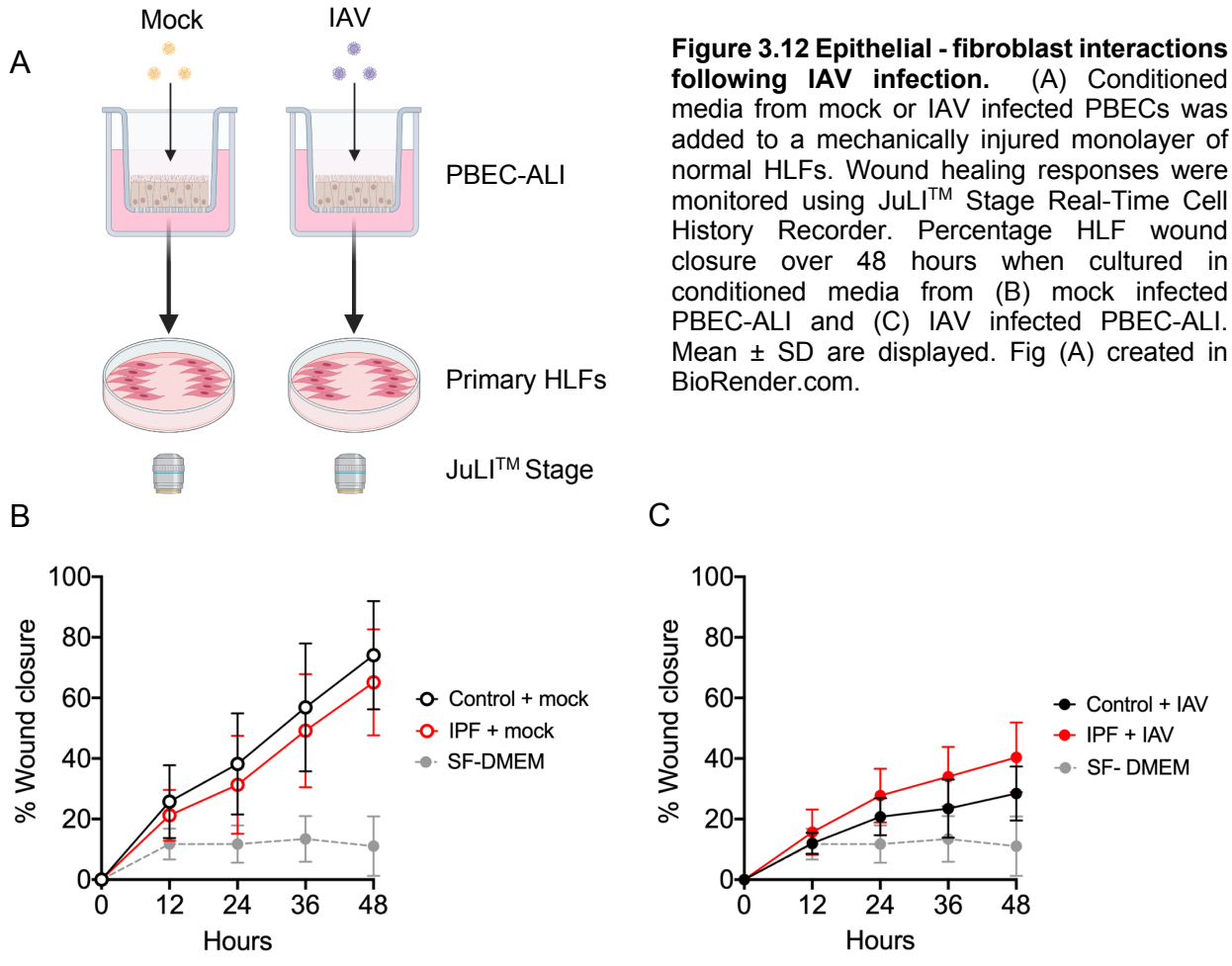


Figure 3.12 Epithelial - fibroblast interactions following IAV infection. (A) Conditioned media from mock or IAV infected PBECs was added to a mechanically injured monolayer of normal HLFs. Wound healing responses were monitored using JuLI™ Stage Real-Time Cell History Recorder. Percentage HLF wound closure over 48 hours when cultured in conditioned media from (B) mock infected PBEC-ALI and (C) IAV infected PBEC-ALI. Mean ± SD are displayed. Fig (A) created in BioRender.com.

3.4.6 Airway epithelial- fibroblast co-culture model

To recapitulate the close spatial relationship that exists between epithelial cells and fibroblasts in the alveoli and airways, and facilitate further studies of the direct cross-talk between these cell types, a co-culture model using primary cells was established. IPF PBECs were cultured in a Transwell insert, and normal primary HLFs underneath for 21 days (Fig. 3.13 A). PBECs were air exposed and the media was supplemented with retinoic acid to encourage mucociliary differentiation of the epithelium. The 0.4μm pores of the membrane allowed direct contact of PBECs and HLFs in culture. TEM was used to precisely characterise the epithelial and fibroblast layers (Fig. 3.13 B – E). This revealed normal organisation and polarisation of IPF PBECs in the Transwell inserts, with intercellular junctions. Microvilli and secretory granules were present but no cilia. Fibroblasts with a typical morphology[297] were arranged in compact layers adherent to the underside of the insert. HLFs could be seen extending into the pores of the membrane suggesting possible direct interactions with PBECs.

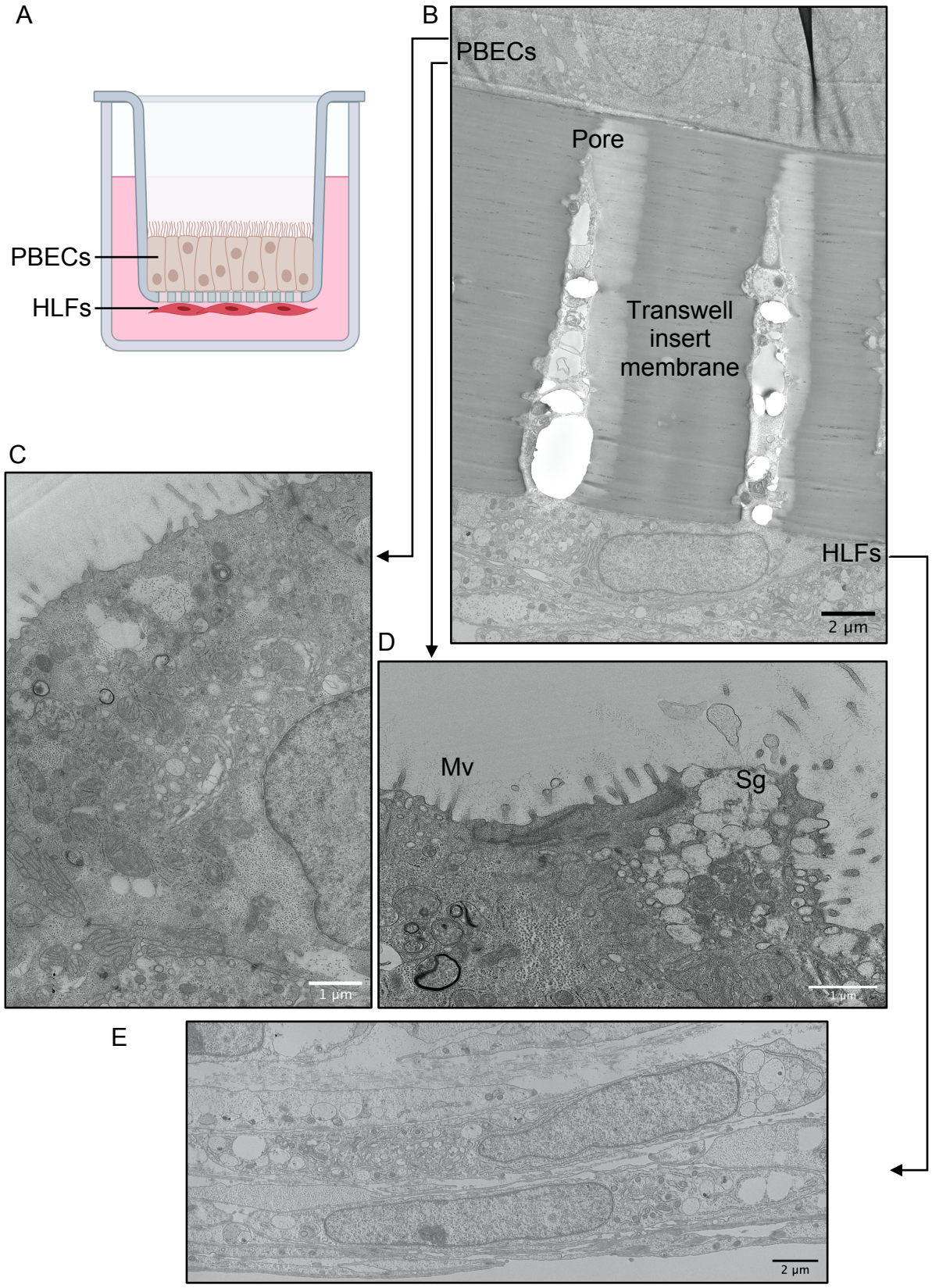


Figure 3.13 Transmission electron microscopy of PBEC-HLF co-culture system. PBECs were co-cultured with HLF at ALI as depicted in (A). Transmission electron microscopy (TEM) was used to characterise the co-culture. Micrographs were taken showing (B) the Transwell insert containing 0.4µm pores with PBECs cultured on the apical surface and primary HLFs cultured on the undersurface. PBECs on the apical surface showing (C) pseudostratified structure, normal polarisation and tight junctions (D) partial differentiation with secretory granules (Sg) and microvilli (Mv) but no cilia. (**E**) Distinct morphology of primary HLFs arranged in compact on the under surface of the insert. Fig. (A) created in BioRender.com.

3.4.7 Airway epithelial-macrophage interactions

Crosstalk between sessile alveolar macrophages and alveolar epithelial cells in the mouse lung was demonstrated to play an important role in regulating immune responses[181]. To establish the spatial relationship of macrophages with airway epithelial cells in the human lung, PCLS of human lung tissue was performed. The tissue was stained for a pan-epithelial marker, EpCAM and airway macrophage marker CD206. This revealed numerous airway macrophages closely associated with the airway and alveolar epithelium in the distal airways (Fig. 3.14).

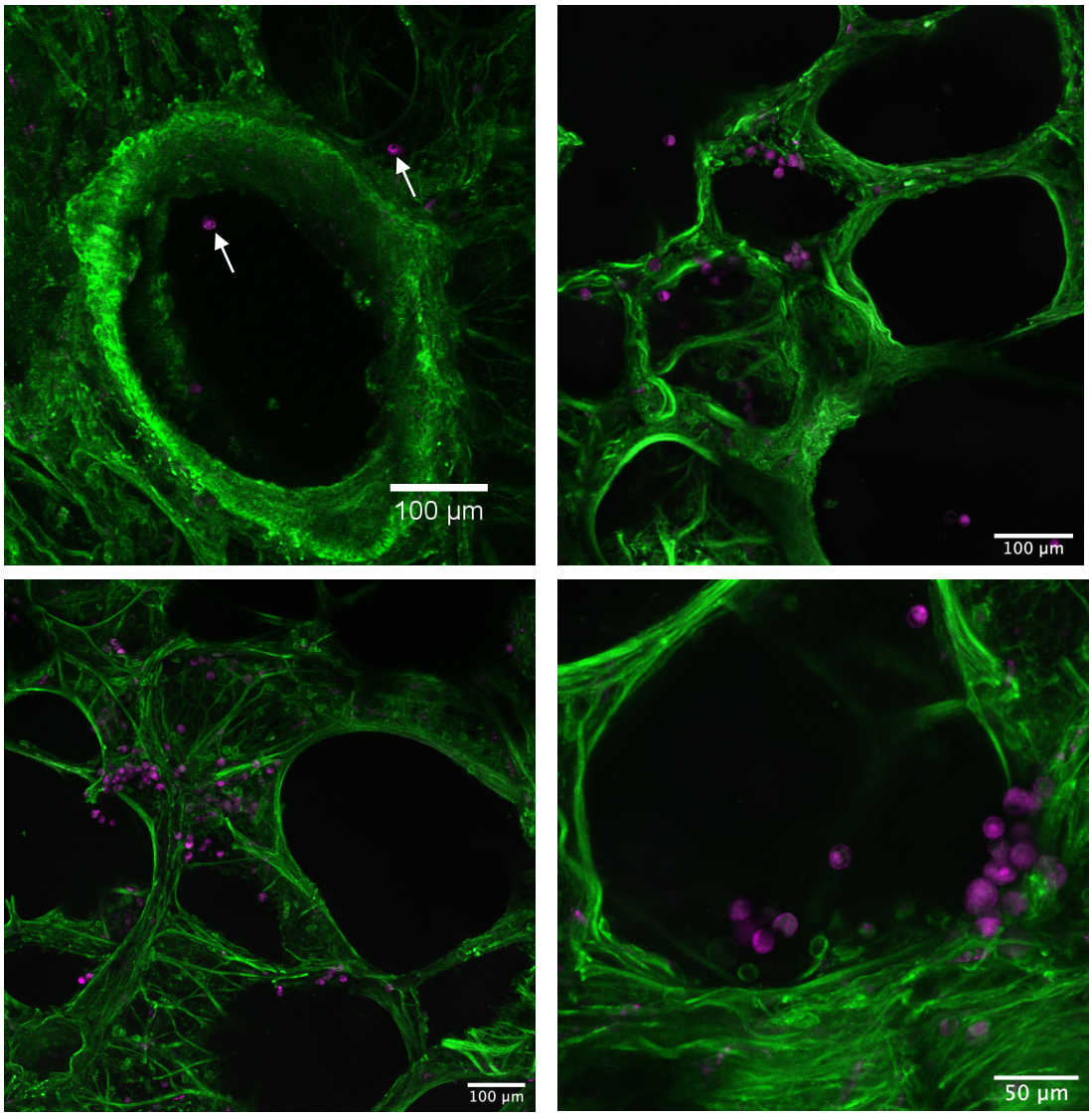


Figure 3.14 PCLS from human distal lung tissue. Normal lung tissue from an 81- year old female ex-smoker was obtained from the Royal Brompton Hospital. IF staining of PCLS for pan-epithelial marker, EpCAM (green) and airway macrophage marker, CD206 (magenta) were captured using a Leica SP5 confocal microscope. White arrows highlighting CD206+ macrophages associated with the epithelium in the airway.

3.5 Discussion

The data described in this Chapter address a critical knowledge gap in our understanding of IPF that is, the phenotype and function of IPF airway epithelial cells. With evidence accumulating in support of a role for the airway epithelium in IPF[53,97,145,298] it was essential to establish a functional IPF PBEC-ALI model for further investigations. In particular a pertinent question remains over whether IPF airway epithelial cells function appropriately to provide protection to the alveolar region or promote disease pathogenesis[298]. Overall, our *in vitro* studies revealed that compositionally the airway epithelium in IPF resembles that found in healthy subjects under homeostatic conditions. Flow cytometric and detailed transcriptomic analysis of cells isolated from airway brushings is provided in Chapters 4 and 5. There were some noteworthy observations from the functional PBEC-ALI experiments which will be elaborated on here.

Transmission electron microscopy and immunofluorescence microscopy confirmed bronchial brushings sampled all major epithelial subtypes in the airway. To maximise cell usage per donor, PBECs were expanded to passage 3 (P3) prior to use in functional experiments. Low passage cells were used to preserve the intrinsic phenotype of IPF airway epithelial cells in culture. Reeves et al. have demonstrated that beyond P3 the stability of gene expression by PBECs declines[299]. For our experiments the cells were not modified in any way to promote propagation in culture.

IPF PBECs cultured directly from brushings took significantly longer to reach 80% confluency levels than healthy control cells, indicating a possible reduction in their intrinsic proliferative capacity. However, when attempting to precisely track cell proliferation using real-time imaging over 48-hours there was no demonstrable difference between IPF and healthy PBECs. These discordant findings may be attributable to differences in the passage number; PBECs were studied directly from brushings in the first experiment, but in the second experiment were used at P3 to allow multiple donors to be studied as a batch under consistent conditions. The discrepancy may also be explained by the shorter duration of the real-time imaging experiment. Nevertheless, this data raises the possibility that IPF PBECs may proliferate at a slower rate than healthy PBECs.

In submerged culture, PBECs isolated from airway brushings revert to a basal cell phenotype and luminal ciliated and secretory cells are lost[274,300,301]. Since airway basal cells play a vital role in the response to epithelial injury and regeneration[194], a submerged culture system was used to test the wound healing capacity of IPF basal cells. This revealed following mechanical injury, there was a modest reduction in wound closure by IPF PBECs (P3) compared to healthy PBECs over 48 hours, which did not reach statistical significance due to inter-donor variability.

In this Chapter I have characterised the first ALI culture model of IPF PBECs derived from airway brushings. There is evidence that airway epithelial barrier dysfunction is a contributory factor in other chronic lung diseases such as COPD[302]. Genetic variants in *DSP,* which encodes desmoplakin - a desmosomal protein that links adjacent cells to maintain epithelial barrier integrity - were associated with fibrotic interstitial pneumonias in a GWAS study[83]. Increased *DSP* mRNA expression was discovered in IPF lung tissue compared to healthy controls, and immunohistochemical staining localised DSP to the airways and honeycomb cysts[303]. Intriguingly, the disease-associated minor allele (G) of intron 5 variant rs2076295 was associated with lower *DSP* expression in IPF and control lungs. The authors postulated this could indicate impaired epithelial repair in IPF patients with G/G genotype at rs2076295[303].

Under homeostatic culture conditions at air-liquid interface, I found IPF airway epithelial cells formed continuous ZO-1 tight junctions with normal transepithelial electrical resistance (TEER) values. This suggests epithelial integrity is preserved in the IPF airway under steady-state conditions. Mucociliary differentiation was observed in all IPF ALI cultures and CBF was comparable to that measured in healthy adults[294], suggesting effective mucociliary clearance.

The airway epithelial surface is continually exposed to environmental insults including air pollutants, cigarette smoke and pathogens which have the potential to cause damage[52], and have been implicated in the aetiology of IPF[304]. Evidence for a role of viral infections in IPF is conflicting[69,71,305]. To study IPF airway epithelial responses to a natural form of injury, a PBEC ALI culture model of influenza A virus infection was employed. Human influenza viruses preferentially infect the human airway epithelium

via α2,6-linked sialic acid receptors[231,306]. Influenza A virus H1N1 infection of immortalised human bronchial epithelial cells resulted in αvβ6 integrin-mediated TGF-β activation and cell death, and in mice led to an increase in total lung collagen deposition mediated by TGF-β signalling[307]. In our model, IAV infection resulted in a transient reduction in epithelial integrity in IPF and healthy PBEC ALI cultures at 48 hours which recovered. Cytotoxicity associated with IAV infection was slightly increased in IPF compared with control PBECs, but this was not statistically significant. Consistent with existing data[306], IL-6, IL-1β and TNF-α were secreted by healthy PBECs in response to IAV infection. Inflammatory mediator release at 72 hours by IPF PBECs was comparable to healthy controls however there was some attenuation in TNF-α and IL-1β responses. Notably increased levels of IL-8 (CXCL-8), a major neutrophil chemoattractant, were discovered in basal CM of IPF PBECs in response to mock and live IAV compared to healthy controls, the latter reaching statistical significance. Consistent with our findings, IL-8 has been shown to be constitutively produced by human airway epithelial cells in culture[282,293,308]. Elevated levels of IL-8 have been reported in the serum and BAL of IPF patients, and these were positively correlated with BAL neutrophil counts[309]. Increased serum IL-8 was also associated with impaired lung function. IL-8 promoted proliferation and migration of IPF fibrogenic mesenchymal progenitor cells, and stimulated migration of macrophages[310].

The microenvironmental milieu created by the cell culture media has been shown to differentially alter responses of cells from the same donor[311]. The composition of the media used may have been responsible for baseline IL-8 release from PBECs cultured at ALI. Epidermal growth factor (EGF) - a ligand for the EGF receptor - can enhance IL-8 release from PBECs[312], and is a supplement used in other airway epithelial cell media, such as that manufactured by PromoCell. The constituents of the supplement used in PneumaCult™-ALI Medium to promote mucociliary differentiation were not readily available upon request from the manufacturer, STEMCELL Technologies. Inc. To validate my findings, it will be necessary to perform further experiments with a larger cohort of donors, studying epithelial-derived cytokine release at multiple time-points following IAV infection.

Several notable genes were upregulated in IPF PBECs compared to healthy controls following IAV challenge. Although these gene changes did not meet statistical significance – perhaps due to insufficient patient numbers - they are candidates which have been previously implicated in IPF. The first, matrilysin or matrix metalloproteinase 7 (*MMP7*) was increased in the lungs of IPF patients and was localised by IHC to the alveolar and bronchiolar epithelium[126]. Levels of serum MMP7 distinguished IPF from other chronic lung diseases and were associated with disease severity, so it has been intimated as possible biomarker of disease[313]. A study using *Mmp7*-null mice showed MMP7 suppressed ciliated cell differentiation in the airway epithelium during wound repair[314]. The second upregulated gene - transgelin, *TGLN* – encodes a cytoskeleton- associated protein which has been localised to ATII cells, and is induced by TGF-β1/Alk5/Smad3 signaling[315]. Higher *TGLN* mRNA expression was reported in IPF lung tissue compared to controls, and it promoted migration of A549 cells[315]. Notably, there was upregulation of *COL3A1,* a gene highly expressed in a pathogenic subpopulation IPF fibroblasts[146], which encodes type III collagen, a key extracellular matrix protein constituent of the fibroblastic foci in IPF[34].

A powerful neutrophil chemoattractant, colony stimulating factor 3, *CSF3* (also known as granulocyte- colony stimulating factor, G-CSF) was significantly downregulated in IPF PBECs compared to controls following IAV infection. In studies using a bleomycin mouse model of pulmonary fibrosis, treatment with low dose G-CSF was protective and reduced fibrosis[316,317]. This was possibly mediated by increased homing of bone marrow- derived mesenchymal stem cells (BMSC) to injured lung tissue where they are thought to promote repair[316]. An attenuated CSF-3 response in IPF airway epithelial cells following injury may therefore promote fibrosis by reducing regenerative responses by BMSCs. There is a paucity of human data on CSF-3 in IPF, however an early study showed G-CSF was detectable in the BAL of IPF patients and correlated with neutrophil count, but was undetectable in healthy controls[318].

A fundamental limitation of the PBEC ALI culture model is that it does not take into account the complex cell-cell interactions which occur in the airway microenvironment *in vivo*[274]. Crosstalk between epithelial cells and the closely associated mesenchyme in the airway and alveolar regions is critical to lung repair and regeneration[99,319]. Evidence suggests epithelial-fibroblast interactions are pivotal to the progression of

IPF[296]. To study these interactions, co-culture models of normal or IPF primary lung fibroblasts and A549 alveolar epithelial cell lines have been used[285,320]. IPF airway epithelial cells are yet to be studied in a co-culture model. A co-culture system using tracheal guinea-pig epithelial cells and fibroblasts demonstrated following mechanical epithelial injury, myofibroblasts expressing α-smooth muscle actin, procollagen I and III mRNA transiently induced on day 4[321].

In this Chapter I explored interactions between primary IPF airway epithelial cells and lung fibroblasts using a conditioned media transfer approach and co-culture model. Conditioned media from IAV- infected PBECs attenuated the wound healing response of normal primary HLFs. It is likely that the effect on fibroblast function was mediated by inflammatory cytokines or cytotoxic by-products released by epithelial cells during IAV infection. Using a co-culture model, Boyd et al. demonstrated infection of normal human bronchial epithelial cells with seasonal influenza (H3N2) induced inflammatory genes and ECM proteases in human lung fibroblasts[322]. Although not statistically significant, CM from mock- infected IPF PBECs caused modest suppression in fibroblast wound healing, and from IAV-infected IPF PBECs caused a modest increase in fibroblast wound healing compared to control. This suggests that injury to IPF airway epithelial cells may differentially modulate interactions with underlying fibroblasts.

A close spatial association exists between human alveolar epithelial cells and fibroblasts[297], and bronchial epithelial cells and myofibroblasts[323], therefore I developed a co-culture model to recapitulate this. Separated by a microporous membrane, IPF PBECs were co-cultured at ALI with primary normal HLFs. Transmission electron microscopy confirmed harmonious growth of the two cell types which displayed normal morphology. Notably, PBECs did not ciliate adequately in co-culture. This may be due to a direct repressive effect of HLFs or the ALI media mix composition which was developed to accommodate both epithelial and fibroblast growth. Further refinement of this proof of principle co-culture model will be required for future experiments.

In addition to bi-directional signalling with stromal cells, epithelial cells interact with immune cells in the airway niche to maintain homeostasis and orchestrate responses to environmental stimuli[52]. Macrophages are abundant in the airspace and reported to

play a significant role in IPF[163,164]. Murine data supports direct epithelial-macrophage interactions in the airway[181], but human data is lacking. For the first time I have demonstrated, using confocal microscopy, co-localisation of macrophages with the airway and alveolar epithelium using PCLS from human distal lung tissue. The close spatial relationship between these cell types in the human airway supports likely crosstalk which will be explored using single-cell RNA-seq in Chapter 5.

In summary, the *in vitro* data presented in this Chapter suggests there are no significant structural or functional differences between cultured IPF and healthy PBECs under steady state conditions. On this basis, the Chapter 3 hypothesis was rejected. IPF-specific changes that were seen following influenza challenge and in co-culture with fibroblasts were modest and there was inter-individual variability therefore would require further validation. Despite remarkable similarities in the transcriptional profile of primary airway epithelial cells cultured at ALI and the *in vivo* airway epithelia[324], and the persistence of cell- intrinsic disease phenotypes in culture[325,326], cells in these models may differ from those in their native microenvironment[327]. In a study comparing transcriptional differences between cultured and directly brushed primary airway epithelial cells, 19% of the genes detected by microarray were differentially expressed[276]. Gene Ontology analysis revealed genes upregulated in brushed airway epithelial cells relative to *in vitro* cultured epithelial cells were involved in cytoskeletal organisation and the immune response, reflective of the *in vivo* environment. To capture differences in the molecular phenotype of epithelial cells freshly isolated from the IPF airway that may not be detectable in the *in vitro* milieu, RNA-sequencing was performed and will be discussed in the subsequent chapters.

CHAPTER 4

Bulk transcriptomic profiling of the IPF airway epithelial niche

4 Bulk transcriptomic profiling of the IPF airway epithelial niche

4.1 Background and rationale

The *in vitro* studies presented in Chapter 3, failed to demonstrate any significant functional differences between IPF and healthy airway epithelial cells under homeostatic culture conditions. However, PBECs were studied in isolation, without the influence of immune cells which play an integral role in shaping the *in vivo* airway microenvironment. Airway mucosal immunity is fine-tuned through the co-operation of airway epithelial cells and resident macrophages[52,169,181]. In Chapter 3 I have shown using human precision-cut lung slices, that airway macrophages, expressing the mannose receptor CD206, reside in close association with $EPCAM^+$ epithelial cells in the lung. Therefore, next-generation sequencing was used to explore IPF-associated changes in gene expression and molecular pathways in both airway epithelial cells and macrophages freshly isolated from the lungs via bronchoscopy.

RNA-sequencing (RNA-seq) generates genome-wide expression data facilitating the unbiased identification of genes and molecular pathways linked to the pathogenesis of disease. The first RNA-seq study in IPF, utilised whole lung tissue samples from IPF patients, COPD patients and histologically normal lung from cancer resections, acquired through Lung Tissue Research Consortium (NHLBI) in the US[328]. The study discovered that the p53/ hypoxia pathway was enriched in both IPF and COPD. In addition to identifying disease-relevant molecular pathways, transcriptional profiling of lung tissue has been used to develop a gene signature predictive of usual interstitial pneumonia (UIP), the histopathological lesion associated with IPF[329].

IPF transcriptomic studies have typically sequenced lung tissue from those undergoing lung transplant[330], thereby reflecting an advanced stage in the disease process. The samples studied were often of peripherally located, fibrotic parenchymal lung tissue. Although the focus of IPF disease pathology is the alveolar region, there is evidence that the nasal epithelium and other macroscopically normal areas of IPF lung tissue also demonstrate transcriptomic changes[331,332]. We reasoned that utilising bronchoscopy to obtain airway samples at an earlier stage during the evolution of the disease may provide new mechanistic insight into the molecular changes occurring in

epithelial cells and macrophages, which are pertinent to IPF disease pathogenesis[30]. In this Chapter I have used a novel, paired sampling strategy to expose dysregulated pathways in both airway epithelial cells and airway macrophages which may be driving fibrotic lung disease. It is important to note that airway macrophages were isolated using BAL fluid, which collects macrophages from both the alveolar region and conducting airways.

4.2 **Hypothesis**

Airway epithelial cells and macrophages isolated from IPF patients by bronchoscopy, herald distinct transcriptomic changes which contribute to molecular pathways driving the development of fibrosis.

4.3 **Aims**

1. Obtain high-quality RNA for transcriptomic analysis from freshly isolated bronchial brushings and sorted CD206+ AMs from IPF patients and healthy controls.
2. Perform unbiased transcriptomic profiling of brushings and AMs in IPF and healthy controls.
3. Identify and functionally annotate key differentially expressed genes between IPF and healthy controls.

4.4 **Results**

4.4.1 Clinical characterisation of subjects recruited

Twenty-one subjects, including 10 healthy controls and 11 IPF patients, were recruited and underwent bronchoscopy to obtain bronchial brushings and BAL samples for bulk RNA-seq studies (Fig. 4.1). The clinical characteristics of the recruited cohort are outlined in Table 4.1. The mean age of healthy control subjects was 48 ± 16.4 years, lower than that of the IPF cohort which was 71.5 ± 4.9 years. Gender was balanced in the healthy control group but, reflecting the demographics of the disease, 7 out of 11 IPF subjects were male. The majority (73%) of IPF patients were ex-smokers. Mean FVC was 83.5% predicted and TLCO was 52.4% predicted, reflecting relatively preserved lung function in the early stages of the IPF. There was heterogeneity in

disease severity amongst IPF subjects as indicated by the range in values for compositive physiological index (CPI); a validated measure of disease extent derived from pulmonary function tests[19]. There was no evidence of airway obstruction in any of the IPF patients or healthy controls recruited (FEV1/FVC >0.7). IPF ID_5414 was taking steroids at the time of bronchoscopy. There was no evidence of active respiratory infection in any of the patients recruited. Heterozygous carriage of the minor *T* allele of *MUC5B* variant rs35705950[90] – an established genetic risk factor for IPF - was found in 30% of healthy controls and 64% of IPF subjects recruited.

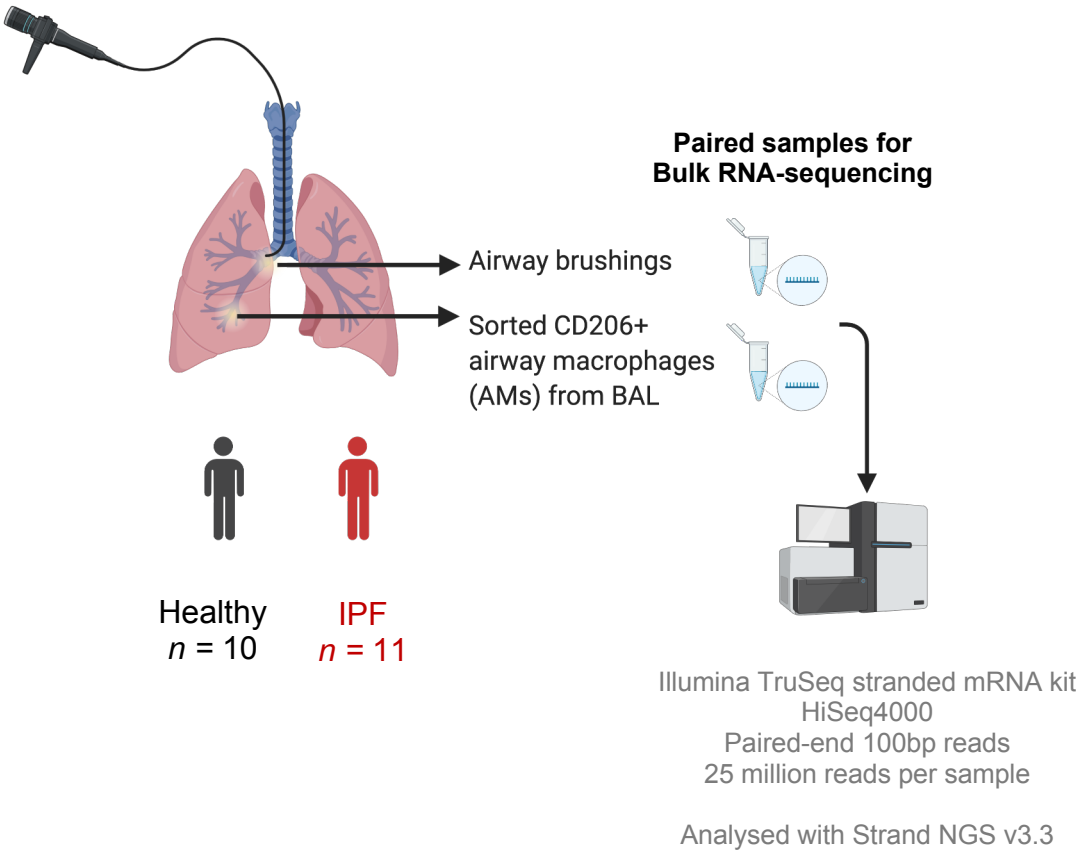


Figure 4.1 Bulk transcriptomic profiling of airway brushings and sorted airway macrophages from IPF patients and healthy controls. Created with BioRender.com.

Table 4.1 Clinical characteristics of recruited subjects for bulk RNA-seq

Sample ID	Diagnosis	Sample	Age	Sex	Smoking status	FEV1 (%pred)	FVC (%pred)	FEV1/FVC ratio	TLCOc (%pred)	CPI*	CT scan			*MUC5B* (rs35705950) genotype
											FS	**TB**	**HC**	
1044	Healthy	BR, AM	60	F	Never	2.82 (98.9)	3.61 (98.9)	0.78	n/a	n/a	n/a	n/a	n/a	G/G
1045	Healthy	BR, AM	61	M	Never	3.6 (101.1)	4.89 (105.8)	0.74	n/a	n/a	n/a	n/a	n/a	G/T
1046	Healthy	BR, AM	59	M	Current	3.62 (99.7)	4.55 (97)	0.80	n/a	n/a	n/a	n/a	n/a	G/T
1047	Healthy	BR, AM	59	M	Never	3.26 (100.6)	3.98 (96.1)	0.82	9.37 (103.56)	n/a	n/a	n/a	n/a	G/G
1048	Healthy	BR, AM	30	M	Never	3.63 (80.6)	4.9 (89.7)	0.74	n/a	n/a	n/a	n/a	n/a	G/G
1049	Healthy	BR, AM	44	F	Never	2.52 (93)	3.12 (93.7)	0.81	n/a	n/a	n/a	n/a	n/a	G/G
1050	Healthy	BR, AM	26	F	Never	4.24 (112.4)	5.05 (113.5)	0.84	n/a	n/a	n/a	n/a	n/a	G/T
1051	Healthy	BR, AM	22	F	Never	3.33 (103.1)	3.55 (95.7)	0.94	n/a	n/a	n/a	n/a	n/a	G/G
1052	Healthy	BR, AM	52	F	Never	2.83 (107.6)	3.64 (111)	0.78	n/a	n/a	n/a	n/a	n/a	G/G
1053	Healthy	BR, AM	67	M	Never	3.04 (85.6)	4.08 (87.2)	0.75	n/a	n/a	n/a	n/a	n/a	G/G
5392	IPF	BR, AM	72	M	Ex	1.81 (62.54)	2.20 (58.04)	0.82	5.55 (68.11)	37	20	7	0	G/G
5400	IPF	BR, AM	69	F	Never	1.81 (83.14)	2.17 (83.17)	0.83	4.56 (67.88)	31	15	9	10	G/T
5413	IPF	BR	75	M	Ex	2.20 (95.90)	2.62 (84.81)	0.86	2.93 (42.06)	51	25	10	0	G/T
5414	IPF	BR, AM	75	F	Ex	2.49 (123.59)	3.04 (124.92)	0.82	4.81 (73.86)	19	10	6	0	n/a
5415	IPF	BR, AM	78	M	Ex	3.35 (111.96)	4.32 (107.88)	0.78	6.02 (70.87)	26	5	2	0	G/T
5418	IPF	BR, AM	64	F	Never	1.38 (74.01)	1.62 (72.18)	0.85	2.98 (50.69)	45	20	6	0	G/T
5421	IPF	BR, AM	62	M	Ex	3.24 (92.89)	4.32 (96.69)	0.75	3.91 (41.26)	45	20	12	15	G/T
5571	IPF	AM	76	M	Ex	2.31 (83.24)	2.65 (72.03)	0.87	3.57 (45.03)	52	15	10	0	G/G
5573	IPF	AM	72	F	Never	1.53 (100.29)	1.67 (89.57)	0.92	2.15 (37.33)	53	15	6	10	G/T
5575	IPF	BR, AM	70	M	Ex	2.44 (86.62)	2.70 (73.33)	0.90	4.05 (46.47)	51	10	3	0	G/G
5604	IPF	BR	73	M	Ex	1.87 (60.9)	2.30 (56.7)	0.81	2.97 (33)	60	15	6	5	G/T

BR – brushings, AM – CD206+ airway macrophages, FS – fibrotic score, TB – traction bronchiectasis, HC – honeycombing

CPI - Composite physiological index = 91 – (0.65 x % predicted TLCO) – (0.53 x % predicted FVC) + (0.34 x % predicted FEV1)

4.4.2 Sample collection and processing for RNA-seq

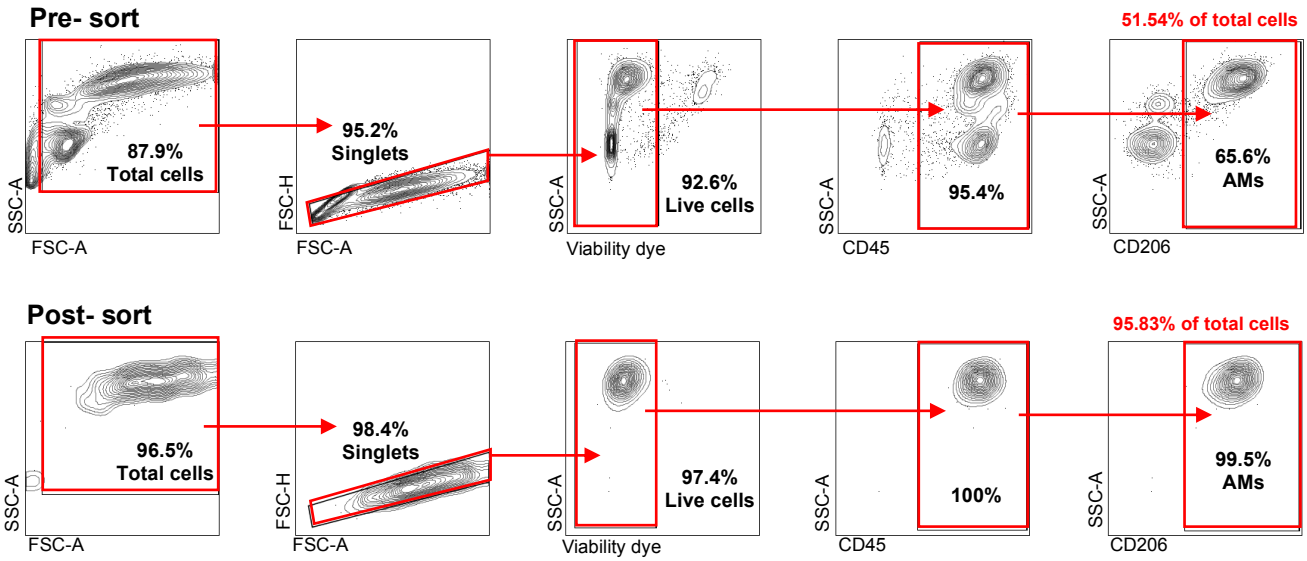
Airway brushings

Paired samples (airway brushings and CD206$^+$ AMs) were successfully obtained from healthy controls (n=10) and IPF patients (n=7). In the IPF cohort it was not possible to obtain paired samples for a total of 4 patients; CD206$^+$ AMs were not obtained from 2 patients due to insufficient BAL return, and brushings were not obtained from 2 patients due to poor patient tolerance of the procedure. Fluorescence- activated cell sorting (FACS) of epithelial cells for sequencing was not feasible due to the low and variable cell yield from brushings (~40,000 cells per brush).

Airway macrophages

Flow cytometry studies have validated CD206 as a cell surface marker of alveolar macrophages[166,333,334]. Airway macrophages expressing CD206 were enriched from IPF and healthy donor BAL using FACS (n=10) or, if unavailable due to the time of sample collection, magnetic-activated cell sorting (MACS) (n=9) system (Fig. 4.2 A and B). Both approaches to cell sorting resulted in a significant enrichment in CD206$^+$ cells as a percentage of total cells, determined by flow cytometry (Fig. 4.2 C). Post-sorting the purity for IPF ID_5400 was less than 80% so this sample was excluded from the downstream RNA-seq analysis. After exclusion of this sample, mean CD206$^+$ purity post-sorting was 93% for all samples.

A



B

FACS sorted samples Sample ID	MACS sorted samples Sample ID
1048	1044
1049	1045
1050	1046
1051	1047
5414	1052
5415	1053
5418	5392
5571	5400
5573	5421
5575	

C

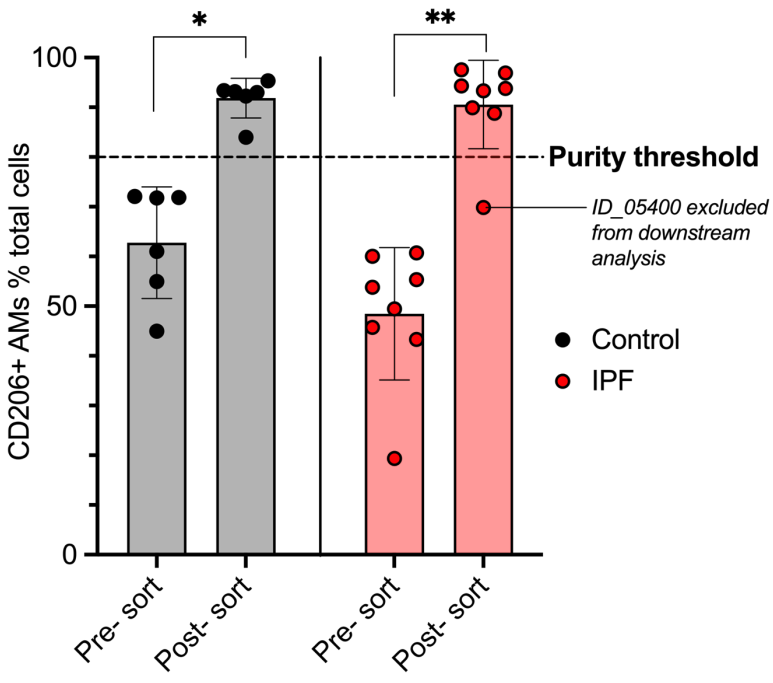


Figure 4.2 Sorting airway macrophages (AM) from human bronchoalveolar lavage fluid. (A) AM flow sorting strategy (example: control subject 1050): live, CD45+CD206+ cells were selected as airway macrophages. % of parent population indicated on each plot, % total cells indicated in red. *SSC-A* – side scatter area, *FSC-A* – forward scatter area, *FSC-H* – forward scatter height. (**B)** Sample IDs listed according to method used for sorting CD206+ AMs. (**C**) Proportion of CD206+ AMs as a percentage of total cells, pre- and post- sorting. IPF ID_5400 was excluded on account of a post-sort purity of <80%. Statistics: Wilcoxin matched-pairs signed rank test. *p<0.05, **p<0.01

4.4.3 Quality control of RNA

The mean total RNA yield was 2874.9ng (89.9 – 8904ng) for airway brushings, compared to 944ng (152 – 3712ng) for AMs (Fig. 4.3 A). From each sample 200ng RNA was used for cDNA library preparation, except for samples with lower yields than this where the entire sample was used (brushing ID_1044, AM ID_1053, brushing ID_5413 and AM ID_5573). RNA quality can significantly influence the reliability of RNA-seq data[335,336]. A validated metric of RNA integrity can be obtained using a software algorithm - the RNA Integrity Number (RIN) - which evaluates features of the electrophoretic trace of RNA samples; a RIN value of 1 indicates a degraded sample, and 10 indicates an intact sample[337]. High quality RNA was obtained from airway brushings and sorted CD206+ AMs (Fig. 4.3 B). The mean RIN value for RNA extracted from sorted AMs was higher (9.4) than for brushing samples (8.0).

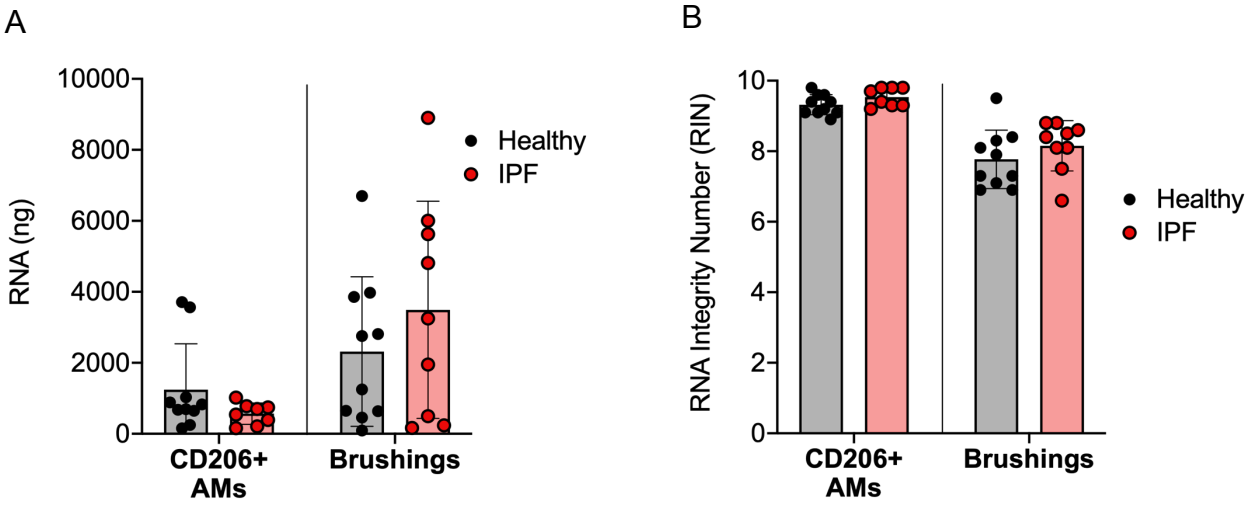


Figure 4.3 Quality control of RNA samples used for sequencing. Scatter/ bar plots showing **(A)** RNA quantity, and **(B)** RNA integrity number as determined by the Agilent 2100 Bioanalyser System. Data for RNA extracted from CD206+ AM samples and airway brushings shown.

4.4.4 Quality control of cDNA libraries and RNA-seq data

To prepare libraries for sequencing, poly-A containing mRNA molecules were purified, fragmented and randomly primed for cDNA synthesis. Indexing adapters were added to the ends of the double-stranded cDNA and PCR was used to enrich DNA fragments with adapters and amplify DNA in the library. Library size was validated using the Agilent DNA 1000, prior to normalisation and pooling for sequencing in the Advanced Sequencing Facility at the Francis Crick Institute, using the Illumina HiSeq 4000 platform. The raw paired-end RNA-seq data for all 37 samples underwent quality control using FastQC (Babraham Bioinformatics, Babraham, UK) and MultiQC v1.6[236]. Phred quality scores estimate the probability of a base being called in error[239]. In this RNA-seq experiment, mean Phred scores across each base position in the read were greater than 35 indicating good quality base calls in all samples sequenced (Fig. 4.4 A). Similarly, all samples passed on mean sequence quality scores (Phred score >28) (Fig. 4.4 B) and the GC content in the library conformed to a roughly normal distribution as would be expected in a normal random library (Fig. 4.4 C). Overrepresented sequences were highest in airway brushing sample ID_5392 (4 – 5% of total sequences). In this sample, untrimmed 'TruSeq Adapter sequence, Index 27' and 'Illumina Single End PCR Primer 1' were identified as possible contributing sources.

Strand NGS v3.3 was used to align the raw reads against the human transcriptome (Genome assembly GRCh38) using Ensembl (Genes and Transcript) annotations (2016.12.01). Reads were filtered out if they mapped with less than 90% identity to the transcriptome. Sequencing generated between 66 and 83 million, 100 base pair, paired- end reads per sample, and there was no difference in sequencing depth between IPF and healthy control samples (Fig. 4.5 A). Alignment metrics revealed the mean percentage of uniquely matched reads per sample was 80.2% (Fig. 4.5 B). Two airway brushing samples (ID_5392 and ID_1044) had less than 70% uniquely matched reads on account of a greater number of multiply matched and unaligned reads and were therefore excluded from downstream analysis. Therefore the mapped and aligned reads of 17 airway brushing samples (healthy *n*=9, IPF *n*=8) and 18 AM samples (healthy *n*=10, IPF *n*=8) underwent quantification and normalisation using the DESeq method[237].

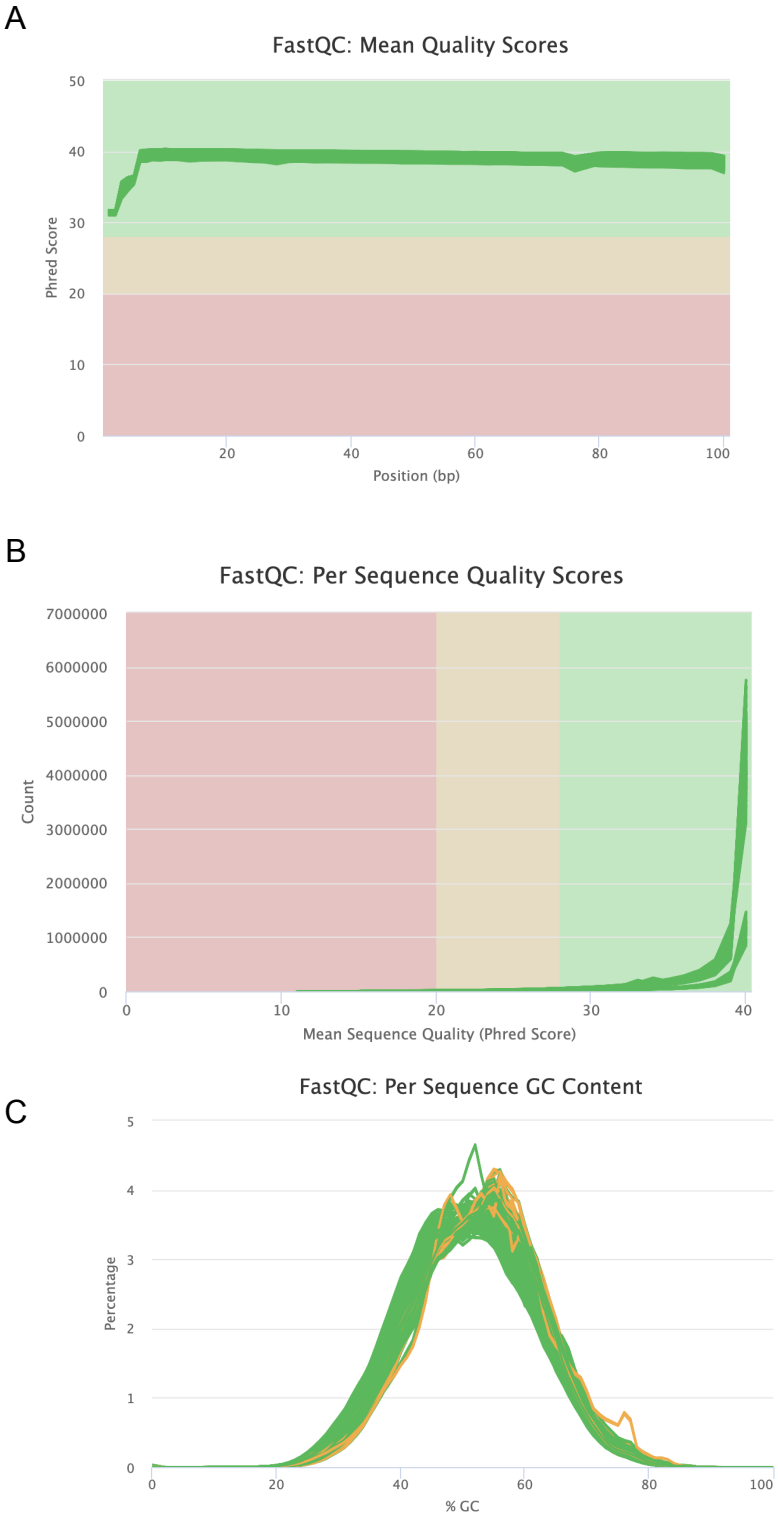


Figure 4.4 Quality control metrics of raw paired-end sequencing libraries. MultiQC v1.6 was used to aggregate and visualise FastQC results across all RNA-seq samples. (A) Mean quality Phred scores for each RNA-seq sample are plotted across each base position in the read. Quality of base calls indicated by shading; good quality (green), reasonable quality (orange), poor quality (red). (B) Per sequence quality scores showing the number of reads according to mean sequence quality scores. (C) Average GC content across each sequence, compared to modelled normal distribution of GC content.

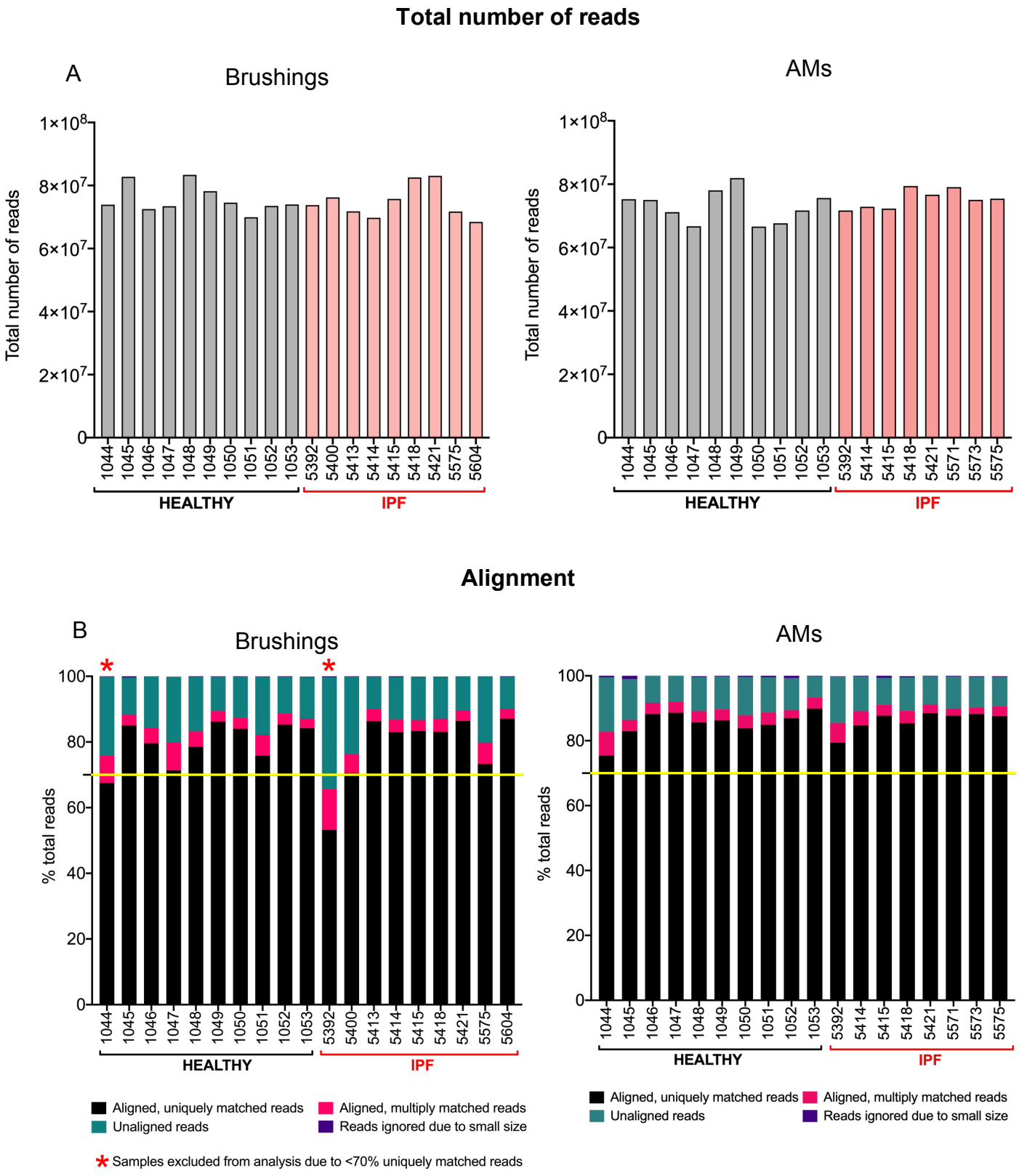


Figure 4.5 Alignment statistics for sequencing files. Bar charts representing (A) total read counts per sample for airway brushings and AMs (B) alignment statistics per sample for airway brushings and AMs generated in Strand NGS v3.3.

4.4.5 Gene filtering

Low abundance genes were filtered using a raw read count threshold prior to downstream analysis (Fig. 4.6). For the airway brushings, only genes with greater than 5 raw counts were retained if present in at least 7 out of the 17 samples *(n=25073* entities*)*. For the AM dataset, only genes with greater than 5 raw counts were retained if present in at least 7 out of the 18 samples *(n=19723* entities*)*. The sample number used for filtering was determined by number of samples in the smallest group minus one.

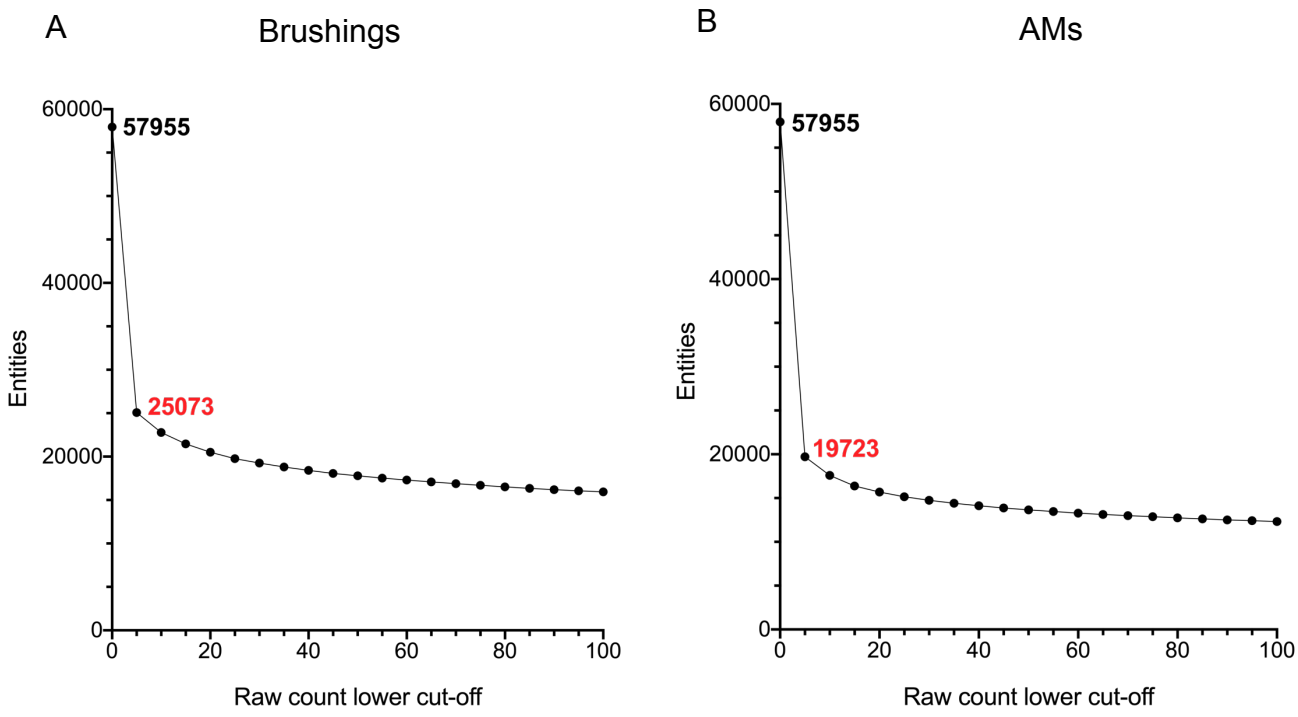


Figure 4.6 Gene filtering by raw read count. Strand NGS v3.3 was used to determine an appropriate read count threshold to filter low abundance genes in (A) airway brushings, and (B) AMs. Only genes (entities) with raw read counts ≥5 (total number indicated in red) were taken forward for downstream analysis.

4.4.6 Transcriptomic analysis of airway brushings

4.4.6.1 Highly abundant genes

To validate the transcriptomic data obtained by sequencing, the 500 most highly expressed genes (ranked by median expression values across samples) in airway brushings from IPF and healthy controls were evaluated. This confirmed the presence of well-established marker genes of major epithelial cell types such as secretory cells (*SCGB1A1, WFDC2, MUC5AC),* ciliated cells (*DNAH5, DNAH9, CETN2)* and basal cells (*KRT5*). When comparing the top 500 most highly abundant genes between IPF patients and healthy controls, there were 471 shared genes, and 29 which were unique to IPF patients or healthy controls (Fig. 4.7). Exclusive to the top 500 highly abundant genes in IPF patients was extracellular matrix glycoprotein, fibronectin, *FN1,* and the cell-surface receptor mediating cell-cell interactions, adhesion and migration, *CD44.*

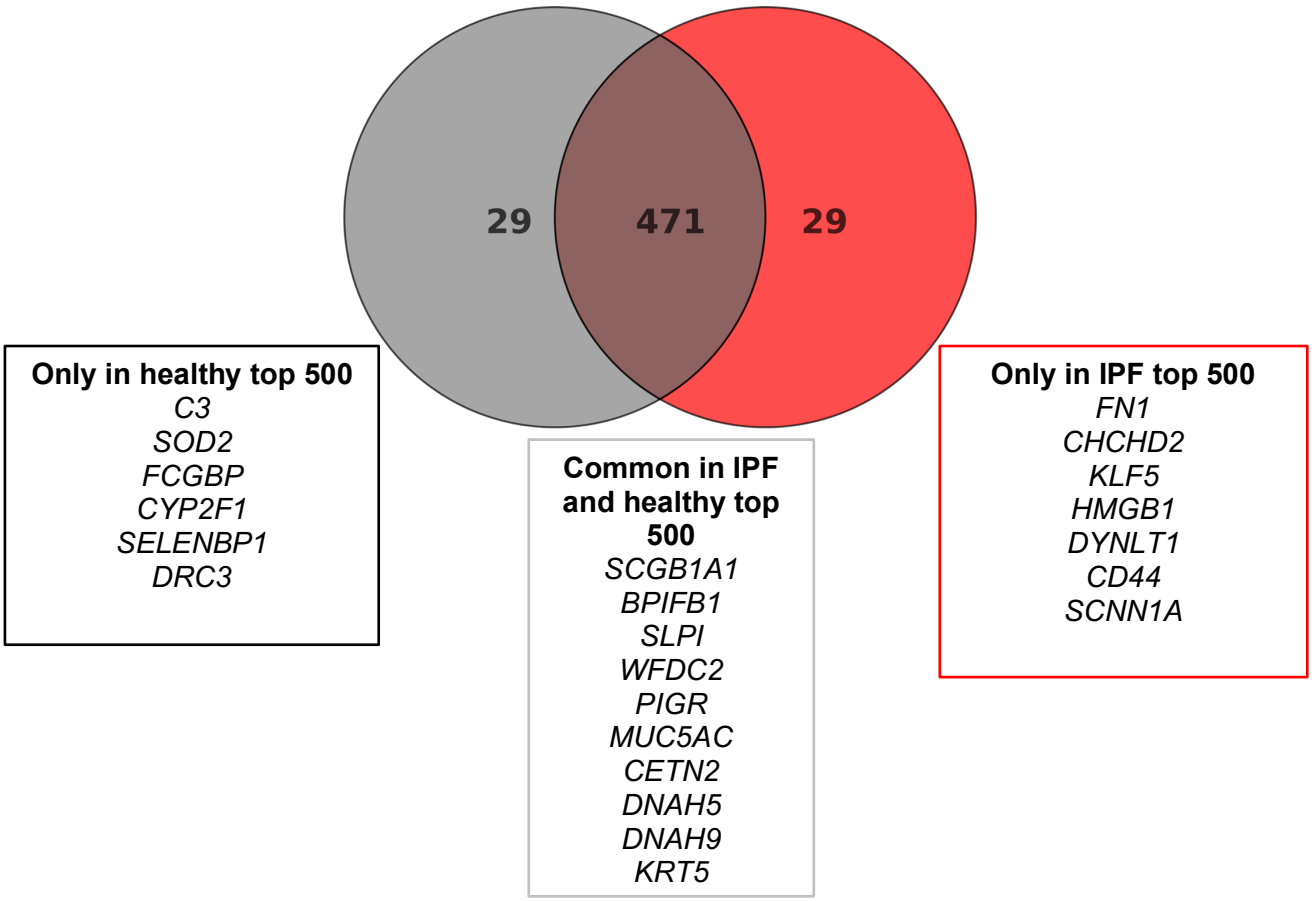


Figure 4.7 Venn diagram showing overlapping genes between the top 500 most abundant genes in airway brushings from IPF patients and healthy controls. Selected genes listed in boxes.

4.4.6.2 Dimensional reduction

To explore high-dimensional RNA-seq data, principle component analysis (PCA) projects the features of the data onto dimensions or principle components (PCs), the first few of which capture most of the variance in the dataset[238]. PCA facilitates the identification of outliers in the dataset. PCA of all filtered genes (*n=25073*) from the airway brushing samples demonstrated no natural clustering of samples by condition, or outlier samples (Fig. 4.8). Principle components 1 (PC1) and 2 (PC2) accounted for only 15.79% and 11.95% of variance in the dataset, respectively.

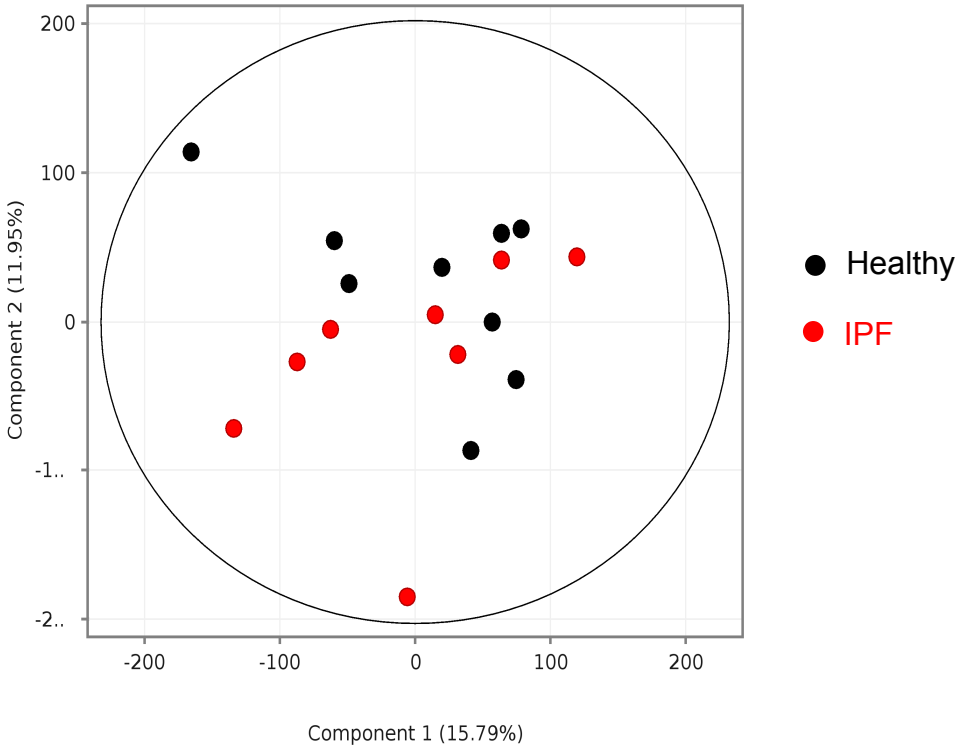


Figure 4.8 Principal component analysis (PCA) of bulk RNA-seq of airway brushings. All filtered genes (*n* = 25073*)* from IPF (*n* = 8) and healthy control (*n* = 9) samples were used for PCA. The major principal components (PC1 and PC2) accounting for the variance in the dataset are shown. A plot ellipse represents the 95% confidence interval using Hotelling's T2 statistics.

4.4.6.3 Clustering

Unsupervised clustering approaches can be used to identify – in an unbiased way - groups of samples in an RNA-seq dataset with similar expression profiles[239]. Low expression genes and those showing limited variability in the dataset add statistical noise to the analysis, and do not aid identification of clusters with similar transcriptional profiles. Therefore, two approaches were adopted to select genes displaying the most variance in the dataset for hierarchical clustering. The first method clustered the top 5% most variable genes in the dataset (*n* = 1254) (Fig. 4.9 A). The second method, a modification of that used by Bloom et al.[338], clustered genes exhibiting greater than 4-fold deviation in expression up or down to the median expression calculated across all samples (*n* = 1054) (Fig. 4.9 B). No clear clustering of transcriptional profiles by condition was identified by either approach. Subject gender did influence the clustering of some samples; in the hierarchical clustering heatmaps, female subjects could be identified by gene modules of intensely downregulated Y-linked genes for example *USP9Y, UTY* and *ZFY* (highlighted on Fig. 4.9 A). No clear gene modules could be identified through these clustering approaches.

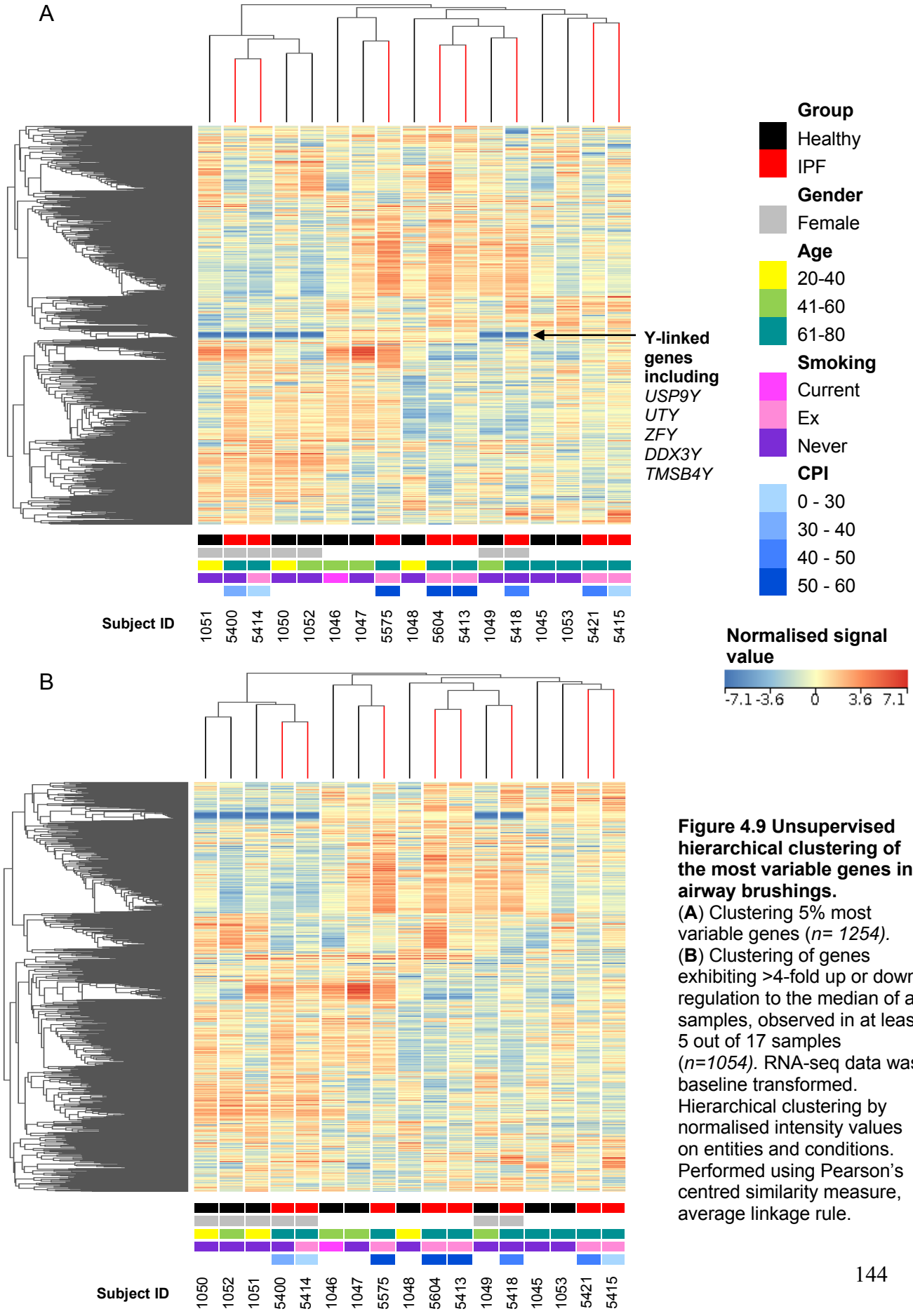


Figure 4.9 Unsupervised hierarchical clustering of the most variable genes in airway brushings. (**A**) Clustering 5% most variable genes (*n= 1254).* (**B**) Clustering of genes exhibiting >4-fold up or down-regulation to the median of all samples, observed in at least 5 out of 17 samples (*n=1054).* RNA-seq data was baseline transformed. Hierarchical clustering by normalised intensity values on entities and conditions. Performed using Pearson's centred similarity measure, average linkage rule.

4.4.6.4 Differential expression analysis

To identify differentially expressed genes between airway brushings from IPF and healthy controls, the DESeq2 method[240] was used within Strand NGS. The DESeq2 package is tailored to perform RNA-seq differential expression analysis where there is high variance (overdispersion) in gene count data and often experiments involve small sample numbers. This revealed a total of 58 differentially expressed genes that were statistically significant using a Wald test, and the Benjamini – Hochberg procedure[241] for multiple-testing adjustment (pAdj <0.05) (Fig. 4.10).

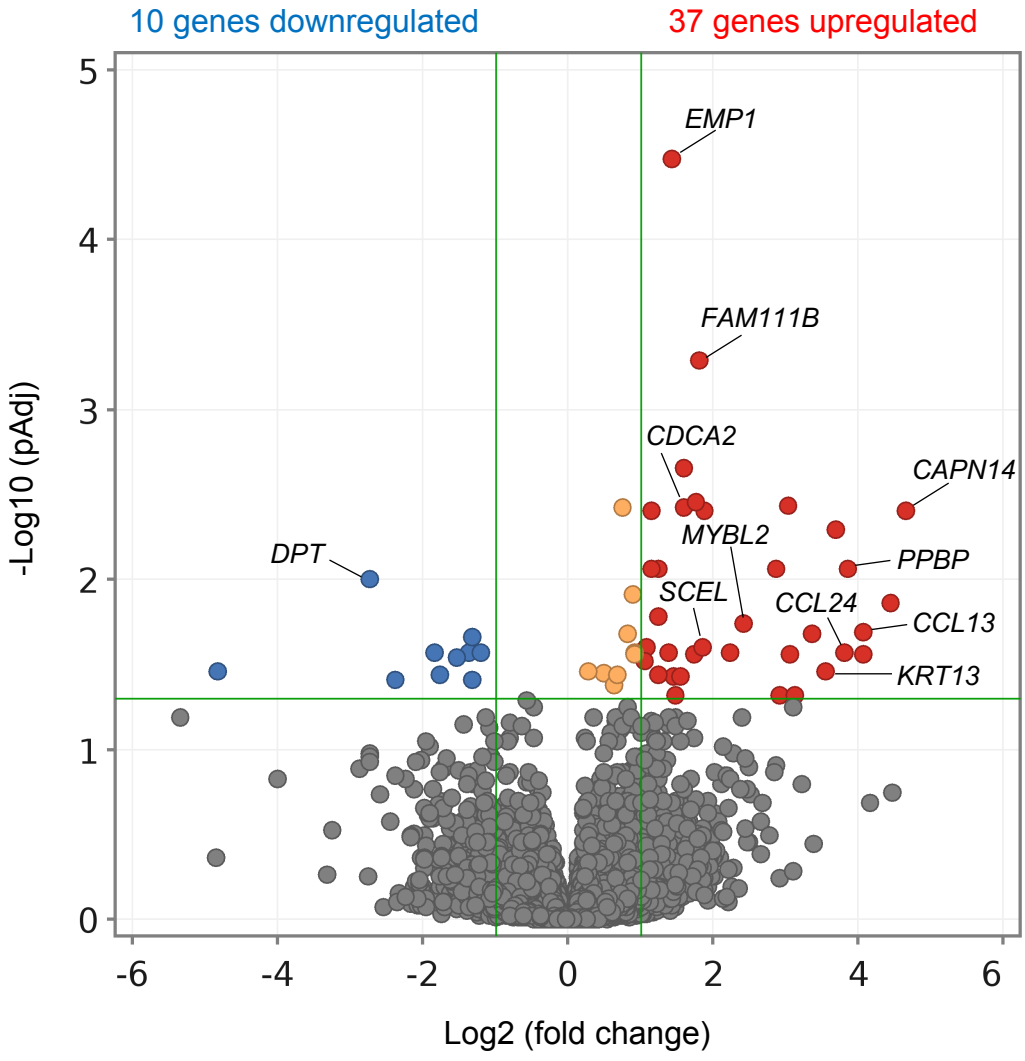


Figure 4.10 Volcano plot showing differentially expressed genes between IPF and healthy airway brushings. DESeq2 was used for differential expression testing. There were 58 differentially expressed genes that were statistically significant using a Wald test and the Benjamini – Hochberg procedure for multiple-testing adjustment (pAdj <0.05). The number of statistically significant genes exhibiting an absolute fold change ≥2 are indicated at the top of the plot. Up- and down- regulated genes are indicated in red and blue respectively. Selected genes of interest are marked on the plot.

The most highly overexpressed gene in airway brushings from IPF patients compared to healthy controls was *CAPN14,* calpain 14 a cysteine protease which has been linked to impaired epithelial barrier function in the oesophagus[339]. Interestingly, *KRT13* a gene encoding a type 1 cytokeratin which has been associated with high-turnover, squamous epithelial structures in the murine trachea[200] and human nasal[179] epithelium, was upregulated 3.5- fold in IPF compared to control. Other genes associated with a squamous epithelial cell phenotype and upregulated in IPF were genes *SCEL,* sciellin and *LY6D,* lymphocyte antigen 6 family member D. Monocyte and lymphocyte chemoattractant *CCL13,* and T cell chemoattractant *CCL24* were upregulated in IPF airway brushings compared to healthy controls. There was higher expression of genes associated with cell division *(MYBL2, SHCBP1, TOP2A, CDCA2)* in IPF compared to healthy controls. The transmembrane, epithelial membrane protein 1, *EMP1,* has been shown to play a role in formation of bronchial epithelial tight junctions and was significantly elevated in IPF compared to controls.

There were too few differentially expressed genes to perform meaningful pathway analysis. Hierarchical clustering of differentially expressed genes revealed two IPF subgroups which clustered subjects by disease severity measured by CPI; a higher CPI subgroup and lower CPI subgroup (Fig. 4.11). Driving these subgroups were two modules of genes, 'Module A' containing genes linked to cell division and proliferation, and 'Module B' containing genes encoding immune cell chemoattractants. Notably, pro-platelet basic protein, *PPBP,* and chemokines *CCL13* and *CCL24* showed a significant positive correlation with CPI of IPF subjects (Fig. 4.12).

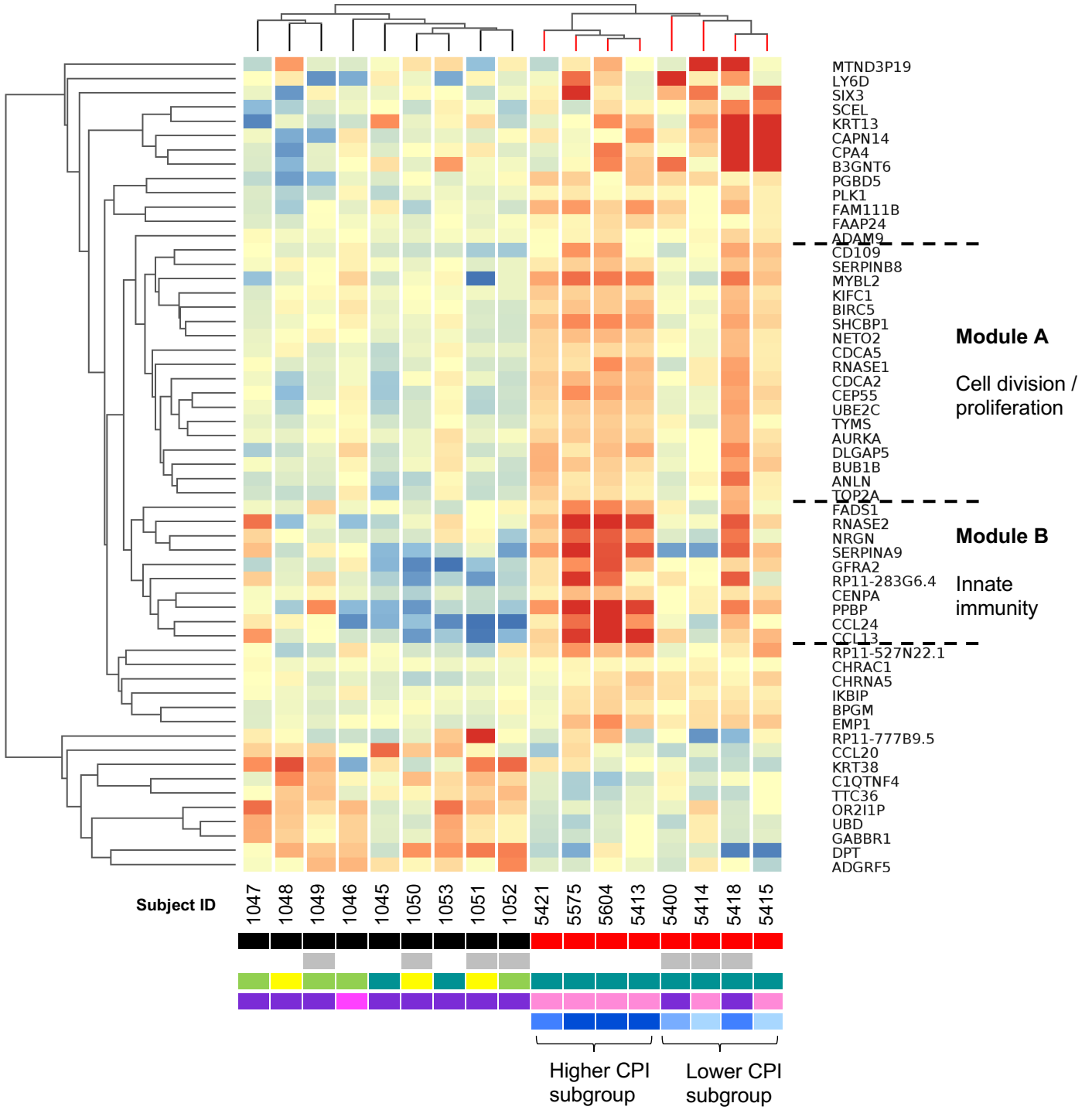


Figure 4.11 Hierarchical clustering of differentially expressed genes between IPF and healthy airway brushings. Significant differentially expressed genes (pAdj <0.05, fold change ≥2) identified by DESeq2 were used to generate the heatmap (*n*=58). RNA-seq data was baseline transformed. Hierarchical clustering by normalised intensity values on entities and conditions. Performed using pearson's centred similarity measure, average linkage rule.

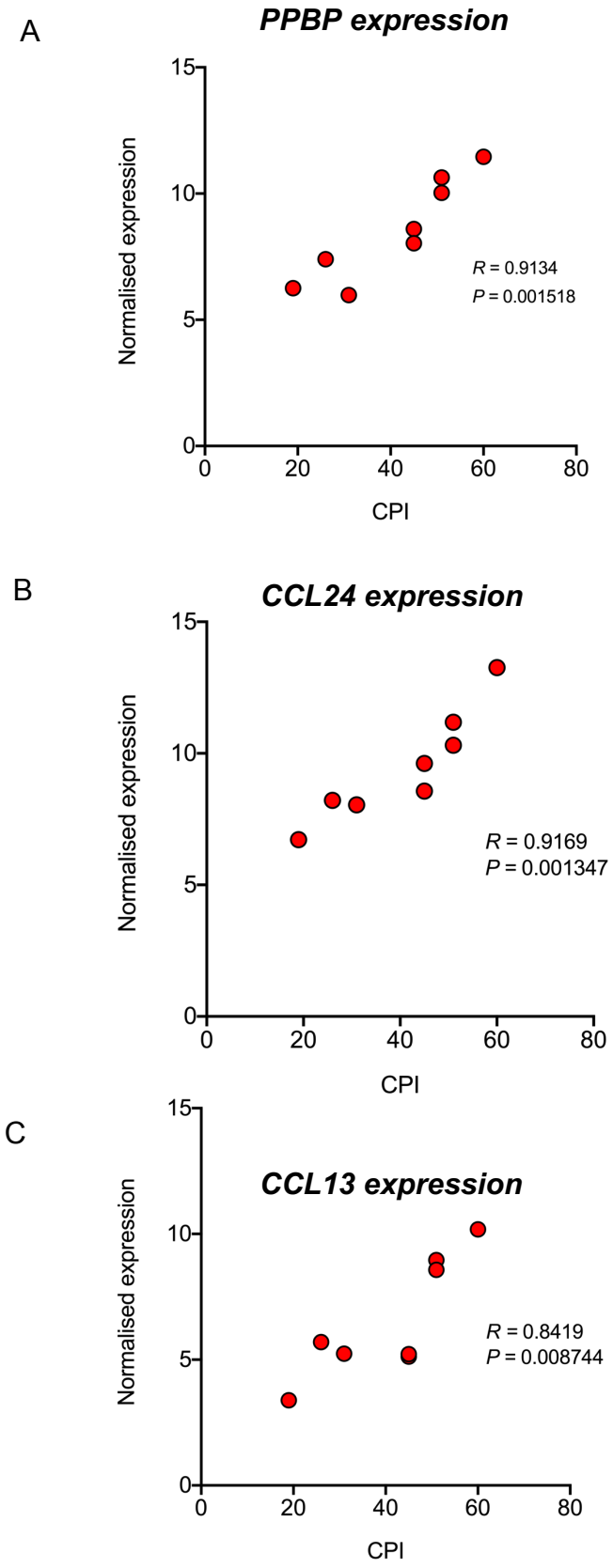


Figure 4.12 Correlation between key upregulated chemokines and IPF composite physiological index (CPI). Correlation between (A) *PPBP, (B) CCL24,* and (C) *CCL13,* and CPI, an indicator of IPF disease severity

4.4.6.5 Immunostaining human samples for KRT13

The upregulation of *KRT13* in IPF airway samples was intriguing given its association with high-turnover squamous structures recently discovered in the murine[200] and human airway[179]. To validate this discovery at the protein level, immunofluorescence staining for KRT13 was carried out on cytospins from airway brushings (Fig. 4.13 A and B), and proximal airways in lung tissue sections from IPF and controls (Fig. 4.13 C – E). This confirmed that rare cells expressing KRT13 were present on IPF airway brushings, and areas of apparent epithelial damage in the proximal airways of IPF lung tissue, but not controls.

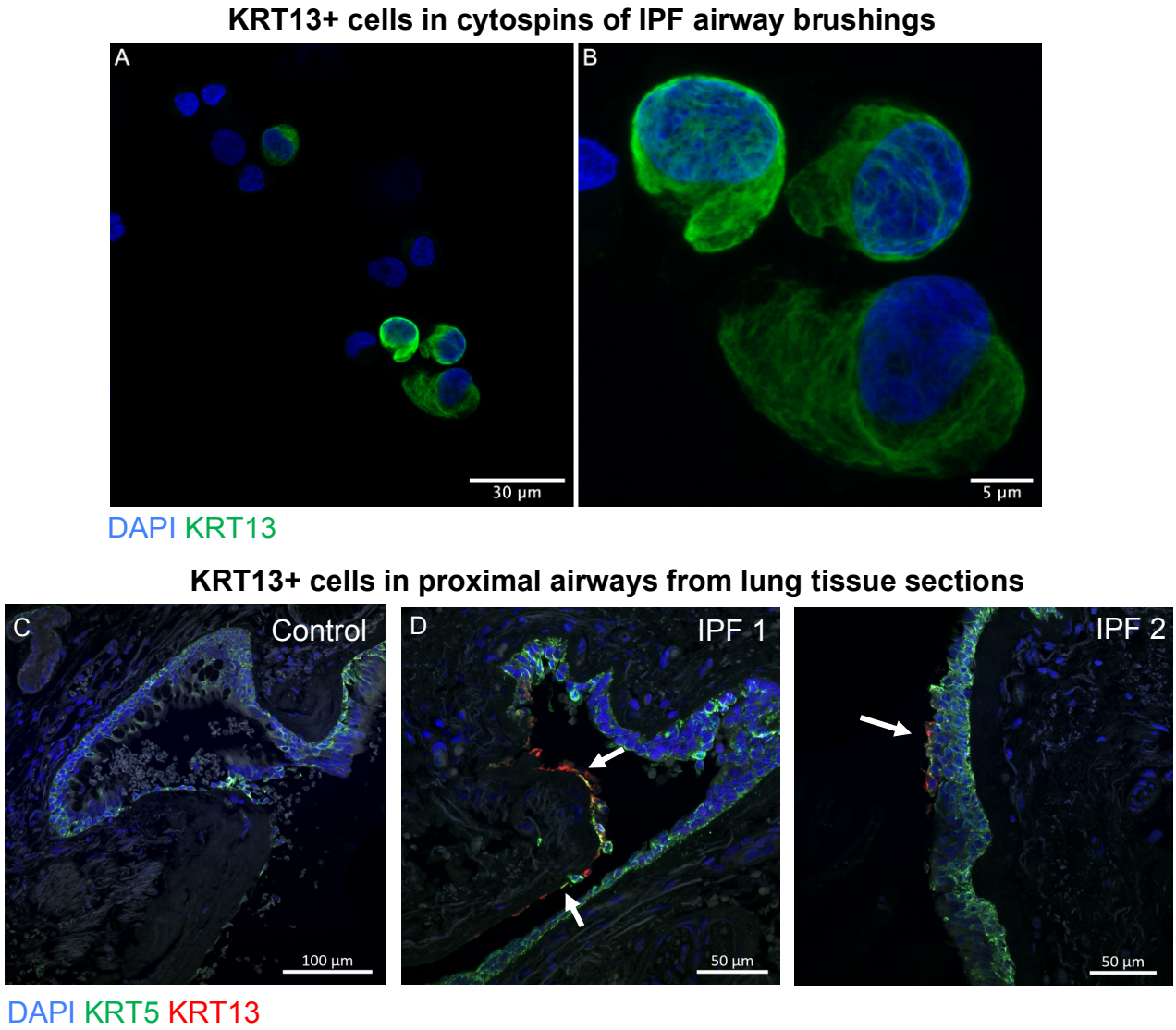


Figure 4.13 KRT13-expressing cells on airway brushings and tissue sections. Immunofluorescence microscopy showing epithelial cells expressing KRT13 on airway brushings (A and B; IPF subject ID_5749) and in proximal airway tissue sections (C – E). Images A and B captured using a Leica SP8 inverted confocal microscope. Images C – E acquired with Dr Fred Fercoq (Cancer Research UK Beatson Institute, Glasgow, UK) using a Zeiss 880 2-photon microscope with spectral detection.

4.4.6.6 Immune cell contaminants

Due to the practical limitations of performing FACS on relatively low numbers of epithelial cells (~40,000 cells collected per brushing), the brushings collected were not purified prior to RNA-seq. In addition to the expected epithelial marker genes in the top 10% most abundantly expressed genes in airway brushings, macrophage marker genes including *TYROBP, APOC1, MRC1* and *MARCO* were also noted. Therefore, to estimate the relative proportions of contaminating immune cells which may be present on airway brushings from our bulk transcriptomic data, a computational tool, CIBERSORTx[250], was used. Using normalised expression values of all filtered genes (n=25073) and LM22[251], a signature matrix of 22 immune cell types, cell fractions were imputed. This revealed the predominant immune cell populations also present in the airway brushings were T cells and macrophage subsets (Fig. 4.14). This finding will be elaborated further using scRNA-seq in Chapter 5.

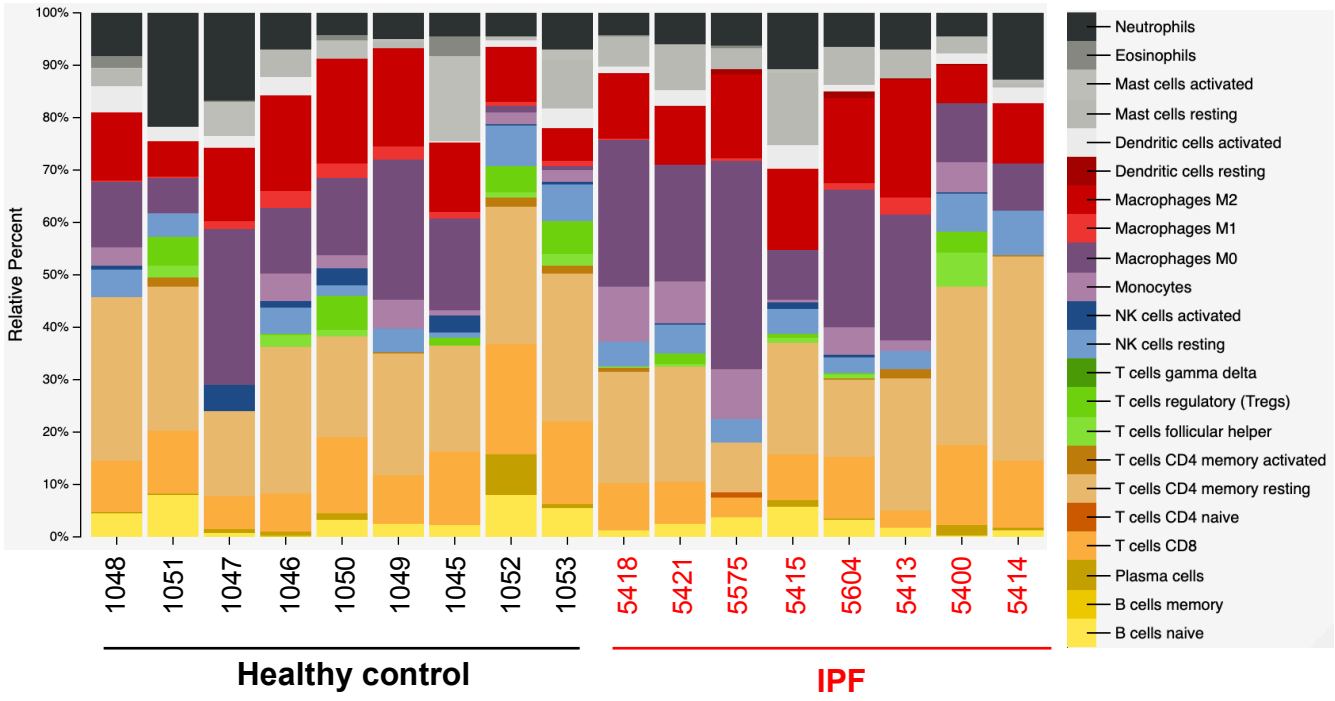


Figure 4.14 Immune cell deconvolution of bulk-RNA seq of airway brushings using CIBERSORTx. All filtered genes from airway brushings (*n* = 25073) were deconvoluted using LM22, a signature matrix of 22 immune cell types. Stacked bar chart showing the relative immune cell fractions present on brushings per subject. B-mode batch correction was performed and 100 permutations were used for statistical analysis.

4.4.7 Transcriptomic analysis of airway macrophages (AMs)

4.4.7.1 Highly abundant genes

To validate the RNA-seq data obtained from AMs, the top 500 most highly abundant genes from IPF and healthy controls were compared. As anticipated, macrophage markers were enriched in all samples and included scavenger receptor genes (*MARCO, MSR, CD163*), fatty acid transport (*FABP4*) and lipoprotein metabolism genes (*APOC1, APOE*). Macrophage-associated genes such as *MRC1,* complement genes including *C1QA,* and cathepsins such as *CTSD* were expressed and have been previously reported to be conserved in macrophages across species (mouse and human)[340]. Exclusive to IPF were genes associated with cell adhesion (*ITGA5, ITGAM, CTNNA1, TLN1),* and proliferation and differentiation of macrophages (*CSFR1, MYADM*[341]*).* Exclusive to the top 500 most abundant genes in healthy controls were genes involved in iron homeostasis (*FBXL5*) and the TCA cycle (*ACO1, MDH1).*

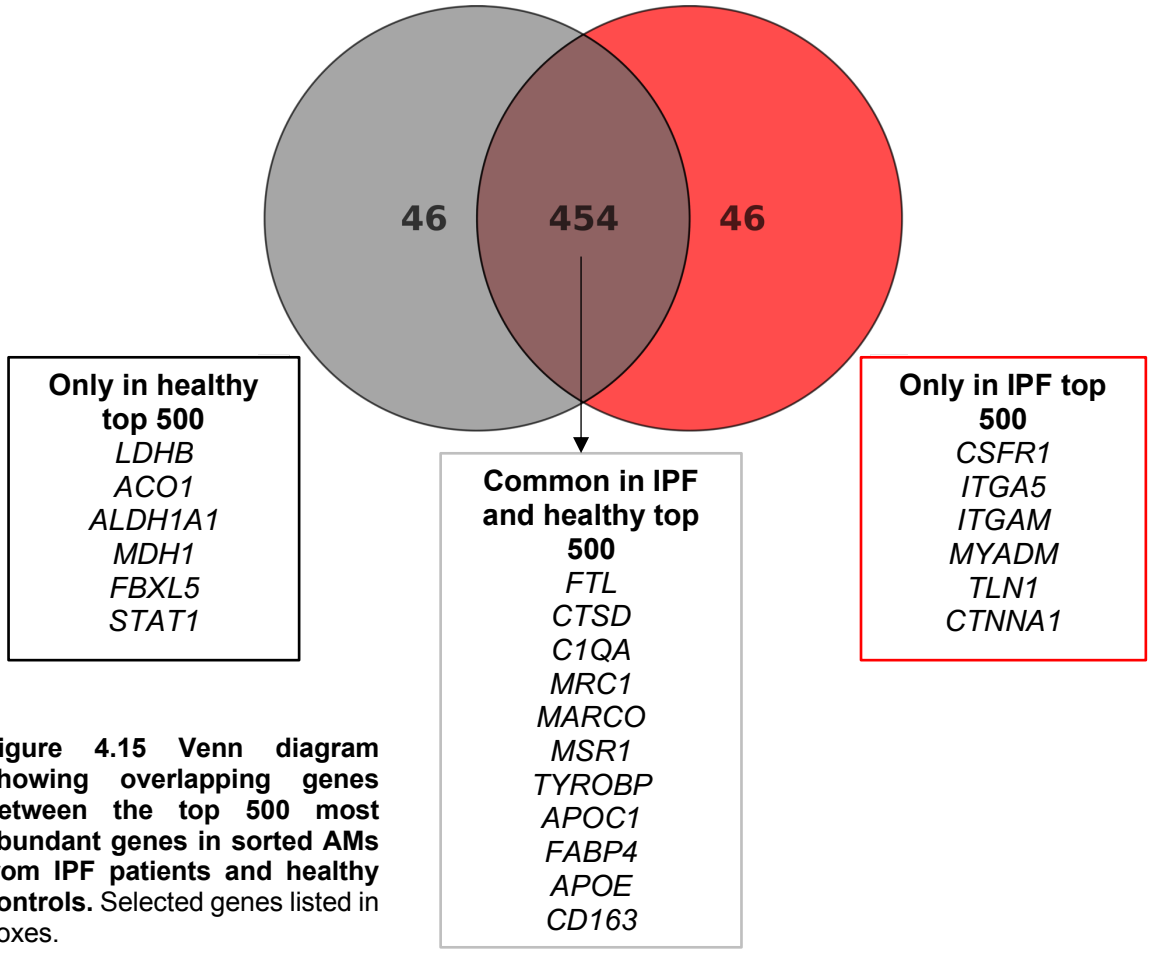


Figure 4.15 Venn diagram showing overlapping genes between the top 500 most abundant genes in sorted AMs from IPF patients and healthy controls. Selected genes listed in boxes.

4.4.7.2 Deconvolution using CIBERSORTx

Mannose receptor, CD206+, expressed by alveolar[166] and other tissue macrophages/ monocytes[333], and enhanced in alternatively activated (M2) macrophages[342], was used to enrich BAL samples prior to sequencing by either flow cytometry or MACS-bead purification to over 93% purity. To provide context to the transcriptomic data obtained and gain insight into the macrophage phenotypes present, I used CIBERSORTx, a method for 'digital cytometry', which can estimate cell type abundance from bulk RNA-seq data. Applying this approach to the bulk AM dataset demonstrated macrophages and monocytes comprise a mean 83% of all cells in each sample (Fig. 4.16). This validated the sample purities calculated using flow cytometry and demonstrated minimal contamination from other immune cell types. CIBERSORTx cell deconvolution was achieved using LM22[251], a signature matrix of 22 immune cells, and highlighted a spectrum of macrophage differentiation and polarisation states[343] were present; monocytes, uncommitted 'M0' and alternatively activated 'M2' macrophages.

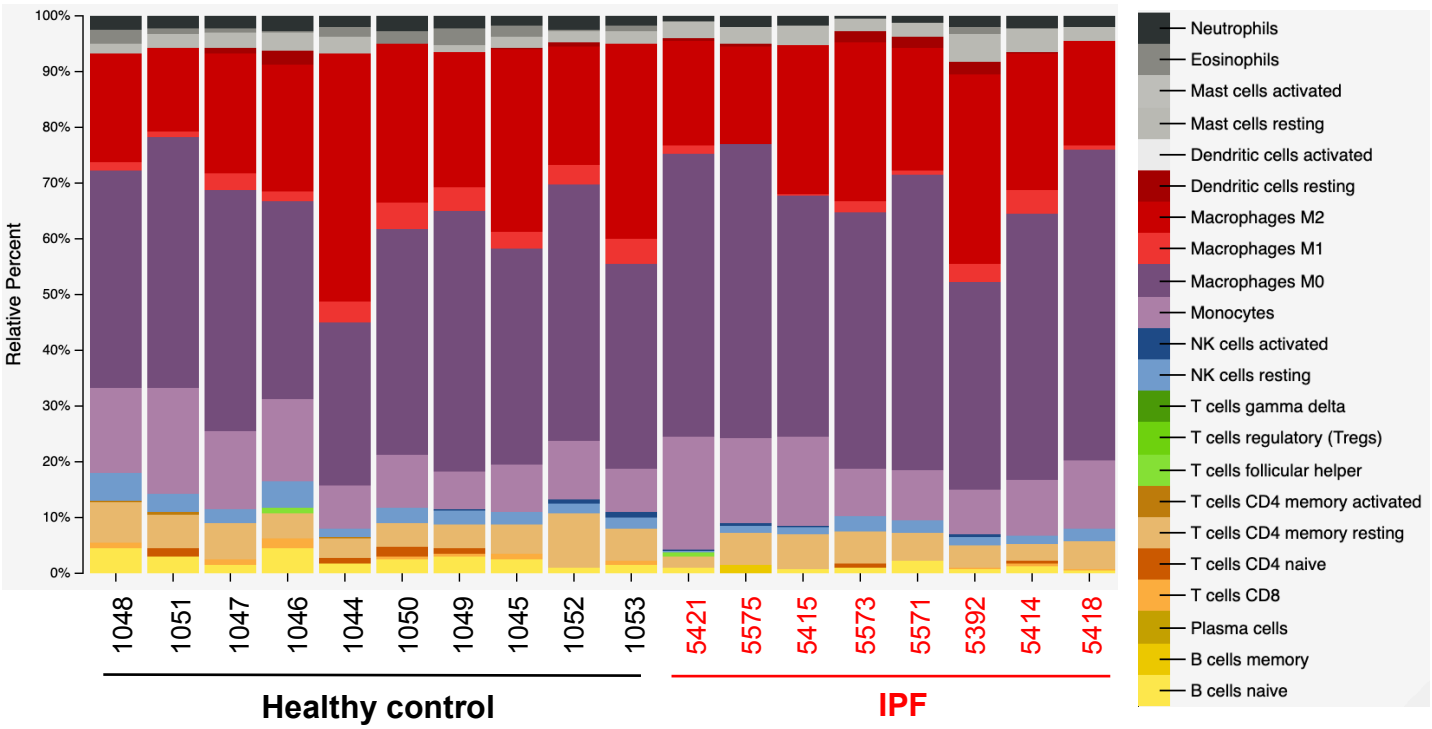


Figure 4.16 Immune cell deconvolution of bulk-RNA seq of sorted AMs using CIBERSORTx. All filtered genes from sorted AMs (*n* = 19723) were deconvoluted using LM22, a signature matrix of 22 immune cell types. Stacked bar chart showing the relative immune cell fractions present in each sample per subject. B-mode batch correction was performed and 100 permutations were used for statistical analysis.

4.4.7.3 Dimensional reduction

PCA of all filtered genes (*n=19723*) from the AM samples did not show natural clustering by condition (Fig. 4.17). There was a trend to healthy samples congregating, however IPF samples collected in the same PCA 'space' (centre of plot). After exclusion of a low purity sample described in 4.4.2, there were no additional outlier samples demonstrated by PCA. The sum of the first two principal components described less than 40% of the variability in the dataset indicating the filtered genes could not easily separate the two groups. It is likely that differences between groups were small and to some extent masked by additional statistical 'noise' from low abundance and low variance genes.

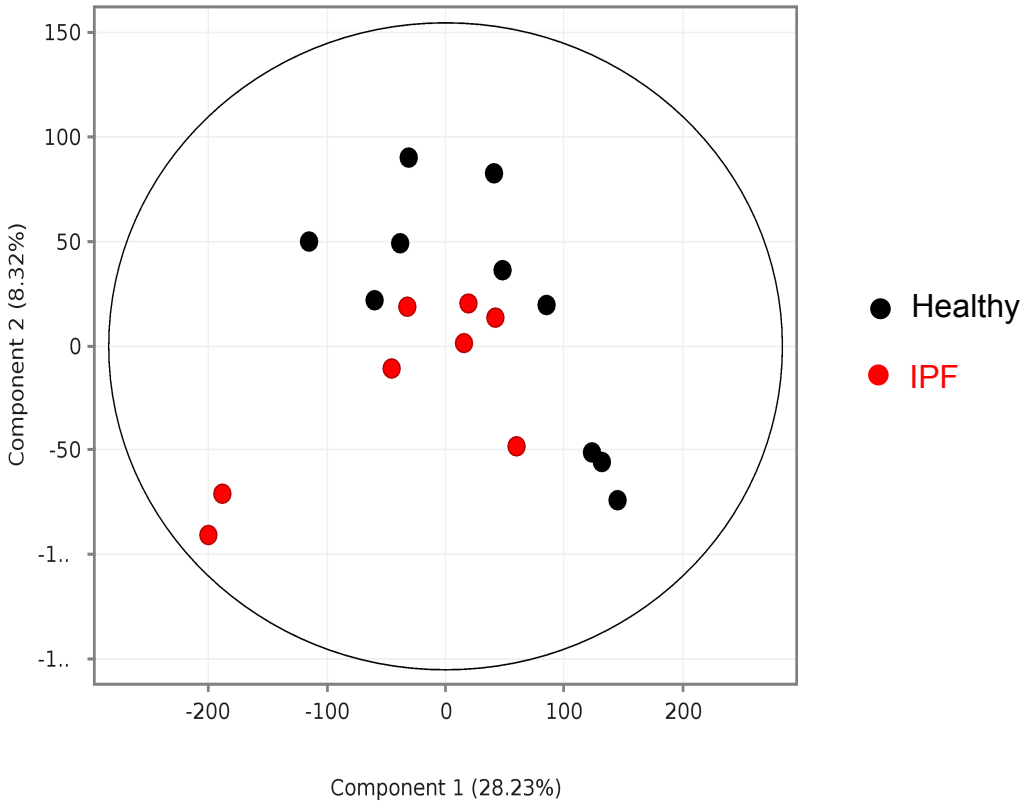


Figure 4.17 Principal component analysis (PCA) of bulk RNA-seq of AMs. All filtered genes (*n* = 19723*)* from IPF (*n* = 8) and healthy control (*n* = 10) samples were used for PCA. The major principal components (PC1 and PC2) accounting for the variance in the dataset are shown. A plot ellipse represents the 95% confidence interval using Hotelling's T2 statistics.

4.4.7.4 Clustering

Two approaches were again adopted to select and cluster the genes displaying the greatest variation in gene expression; the first using the top 5% most variable genes in the AM dataset (*n* = 986) (Fig. 4.18A), and the second taking genes exhibiting greater than 4-fold deviation in expression up or down to the median expression calculated across all samples (*n* = 1317) (Fig. 4.18 B). Clustering using normalised intensity values was performed on entities (genes) and conditions. Whilst there was a cluster of six healthy controls with similar transcriptomic profiles, other healthy controls clustered with IPF samples. Age, gender or smoking status were not responsible for the main clusters observed. However, amongst the cluster of six healthy controls, further separation by age was appreciable, with two subjects in the 60 – 80 year age category clustering separately from four subjects spanning 20 – 60 years. IPF subjects with severe disease - defined by a higher CPI - such as ID_5571 and ID_5573, generally clustered separately to those with milder disease, with the caveat that there were healthy controls amongst these clusters. By clustering 5% of the most variant genes it was possible to identify 3 modules which were used for GO enrichment analysis (Fig. 4.19). Genes in module 'A' (*n* = 167) showed better separation by condition and were generally upregulated in IPF subjects compared to controls. Interestingly this module was enriched in GO biological processes such as 'extracellular matrix organisation' and 'myeloid leukocyte migration'. Module B (*n* = 342) did not show clear separation by condition and was enriched in only one GO biological process; 'golgi organisation'. Genes in module 'C' were generally upregulated in control (*n* = 201) and included process such as 'response to biotic stimulus' and 'regulation of immune system process'.

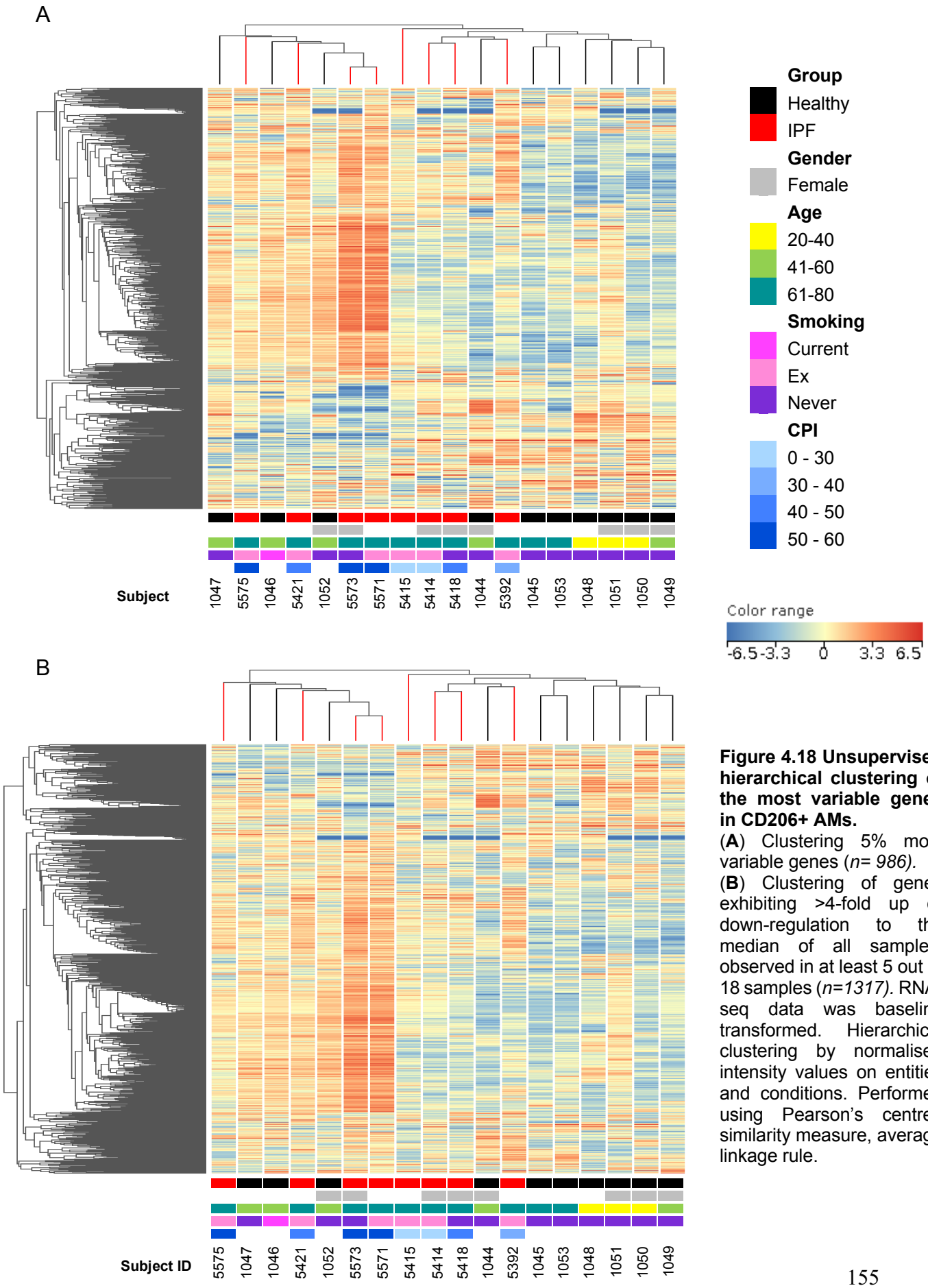


Figure 4.18 Unsupervised hierarchical clustering of the most variable genes in CD206+ AMs.
(**A**) Clustering 5% most variable genes (*n= 986).* (**B**) Clustering of genes exhibiting >4-fold up or down-regulation to the median of all samples, observed in at least 5 out of 18 samples (*n=1317).* RNA-seq data was baseline transformed. Hierarchical clustering by normalised intensity values on entities and conditions. Performed using Pearson's centred similarity measure, average linkage rule.

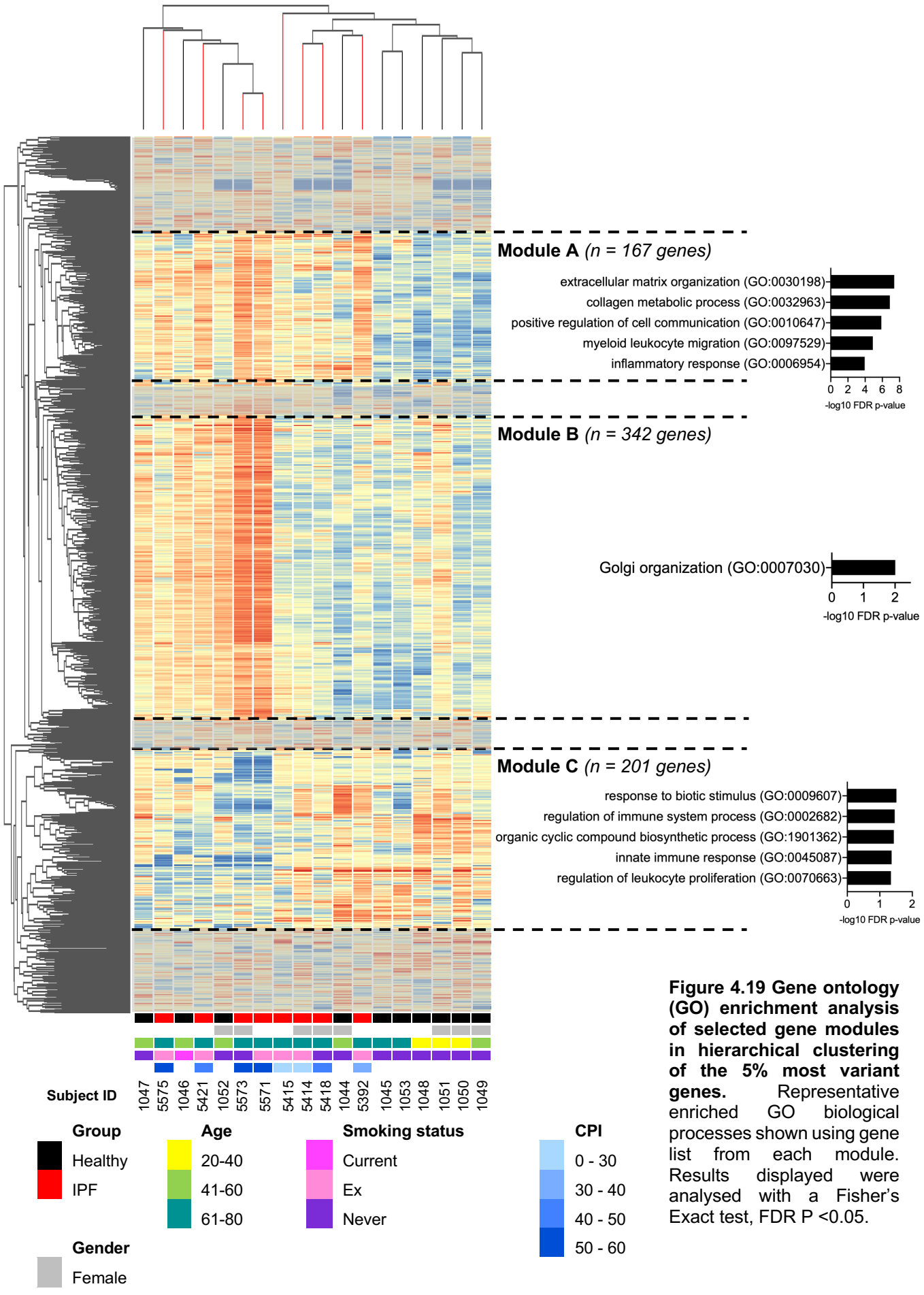


Figure 4.19 Gene ontology (GO) enrichment analysis of selected gene modules in hierarchical clustering of the 5% most variant genes. Representative enriched GO biological processes shown using gene list from each module. Results displayed were analysed with a Fisher's Exact test, FDR P <0.05.

4.4.7.5 Differentially expressed genes between IPF and healthy controls

A total of 929 differentially expressed genes (pAdj <0.05) were revealed by DESeq2[240]. To focus on the most biologically relevant expression changes, genes exhibiting an absolute fold change of ≥2 between IPF and control (*n* = 566), were used for further downstream annotation and pathway analysis. There were 393 upregulated genes and 173 downregulated genes (≥2 fold change, pAdj <0.05) in IPF AMs compared to controls (Fig. 4.20).

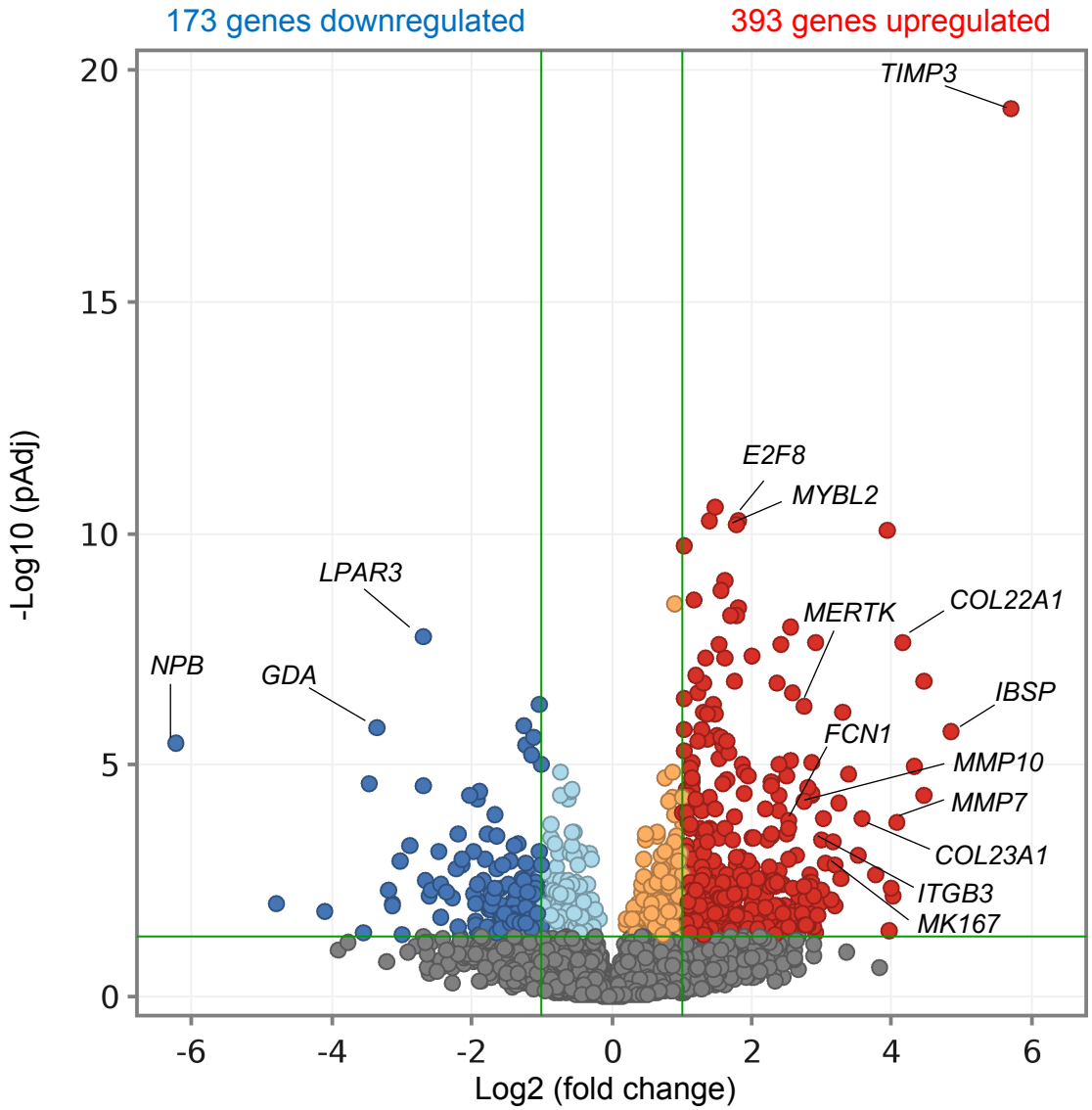


Figure 4.20 Volcano plot showing differentially expressed genes between IPF and healthy sorted AMs. DESeq2 was used for differential expression testing. There were 929 differentially expressed genes that were statistically significant using a Wald test and the Benjamini Hochberg procedure for multiple-testing adjustment (pAdj <0.05). The number of statistically significant genes exhibiting an absolute fold change ≥2 are indicated at the top of the plot. Up- and down- regulated genes are indicated in red and blue respectively. Selected genes of interest are marked on the plot.

The most significantly upregulated gene was tissue inhibitor of metalloproteinase 3, *TIMP3,* which can irreversibly inactivate metalloproteinases. Metallopeptidases responsible for breakdown of the extracellular matrix including *MMP8, MMP7, MMP10, MMP2* and *MMP25* were significantly upregulated in IPF AMs compared to controls. Also relevant to matrix remodelling, collagens *COL22A1, COL23A1* and *COL5A3* were also overexpressed in IPF AMs. There was increased expression of integrins mediating cell-matrix interactions including integrin subunit beta-3, *ITGB3,* and integrin subunit alpha 9, *ITGA9* in IPF. Notable was the overexpression of monocyte associated genes *FCN1*[340] and *IL1R2*[164]*,* and *MERTK,* a marker gene previously found to be co-expressed in profibrotic SPP1-high macrophages[164]*.* Genes linked to cell cycle progression (*E2F8, MYBL2*) and proliferation (*MKI67)* were also increased in IPF AMs compared to control. Differentially expressed genes were subjected to hierarchical clustering (Fig. 4.21). Unlike the RNA-seq data from airway brushings, there were no disease severity-driven subgroups in the IPF cluster. Interestingly control subject ID_1046, a current smoker with a 20- pack year history of smoking clustered with the IPF subjects.

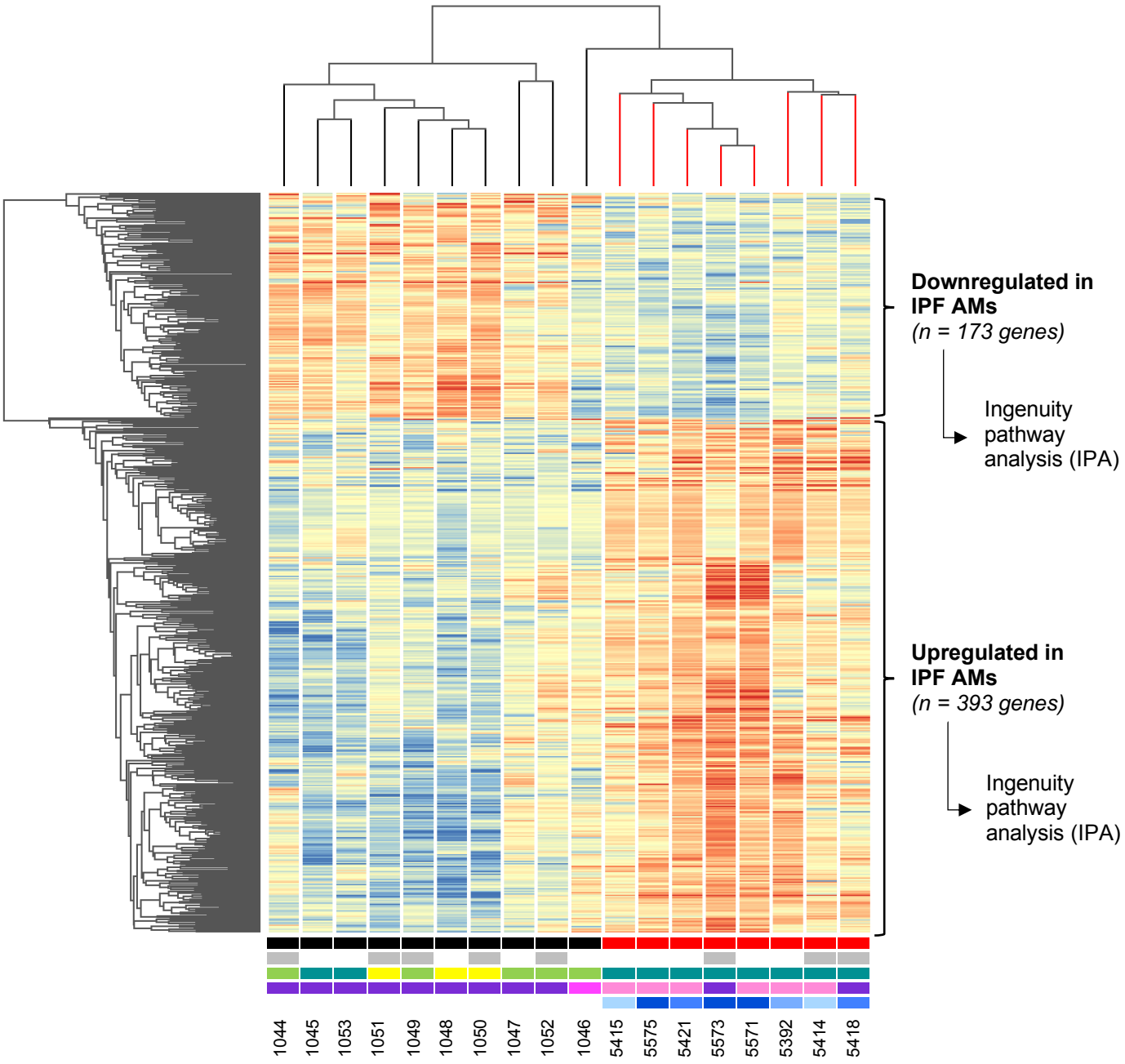
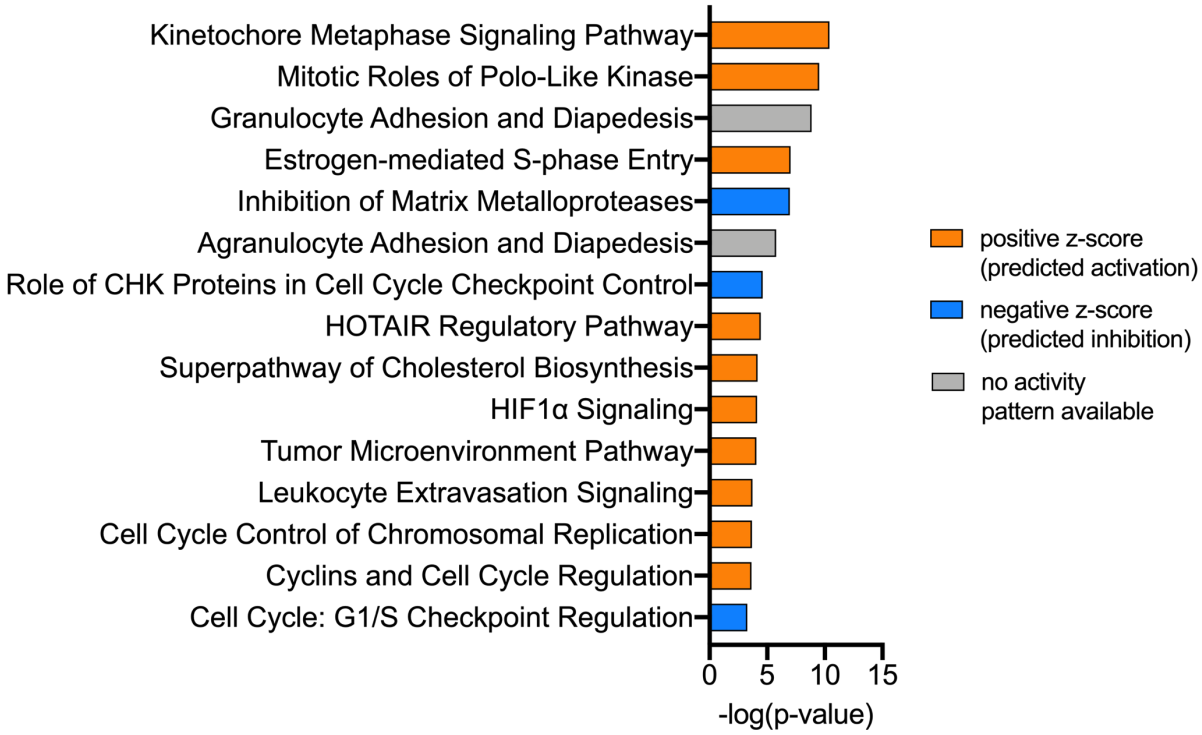


Figure 4.21 Hierarchical clustering of differentially expressed genes between IPF and healthy sorted AMs. Significant differentially expressed genes (pAdj <0.05, fold change ≥2) identified by DESeq2 were used to generate the heatmap (*n* = 566). RNA-seq data was baseline transformed. Hierarchical clustering by normalised intensity values on entities and conditions. Performed using pearson's centred similarity measure, average linkage rule.

4.4.7.6 Ingenuity Pathway Analysis (IPA)

To provide biological context to the transcriptomic changes observed in the IPF AMs, the data were analysed with use of Ingenuity Pathway Analysis (IPA) (QIAGEN Inc.). Based on gene expression changes in the dataset, this software package provides insight into signalling and metabolic pathways activated, upstream regulators, interacting gene networks and biological function using curated information from the QIAGEN Knowledge Base[245]. Genes significantly upregulated by an absolute fold change ≥2 in IPF AMs compared to control AMs were submitted with fold change regulation and adjusted p-values for IPA. The significance of the overlap of these genes with well-established, canonical pathways in the QIAGEN Knowledge Base was calculated in the software by the right-tailed Fisher's Exact Test. The most significant canonical pathways in the dataset were associated with cell cycle regulation including 'mitotic roles of polo- like kinases' and 'role of CHK proteins in cell cycle checkpoint control', and cellular immune response including 'granulocyte adhesion and diapedesis' and 'leukocyte extravasation signalling' (Fig 4.22 A, Table 4.2). Canonical pathways in these broader categories shared common genes, in particular those encoding matrix metalloproteinases (MMPs) (Fig. 4.22 B, Table 4.2). 'HOTAIR regulatory pathway' which was enriched and predicted to be activated in IPF AMs has been reported to play a role in epigenetic regulation and cancer[344]. The enrichment of 'superpathway of cholesterol biosynthesis' in IPF AMs adds to emerging evidence that dysregulated lipid homeostasis in macrophages may contribute to chronic lung disease[345].

A



B

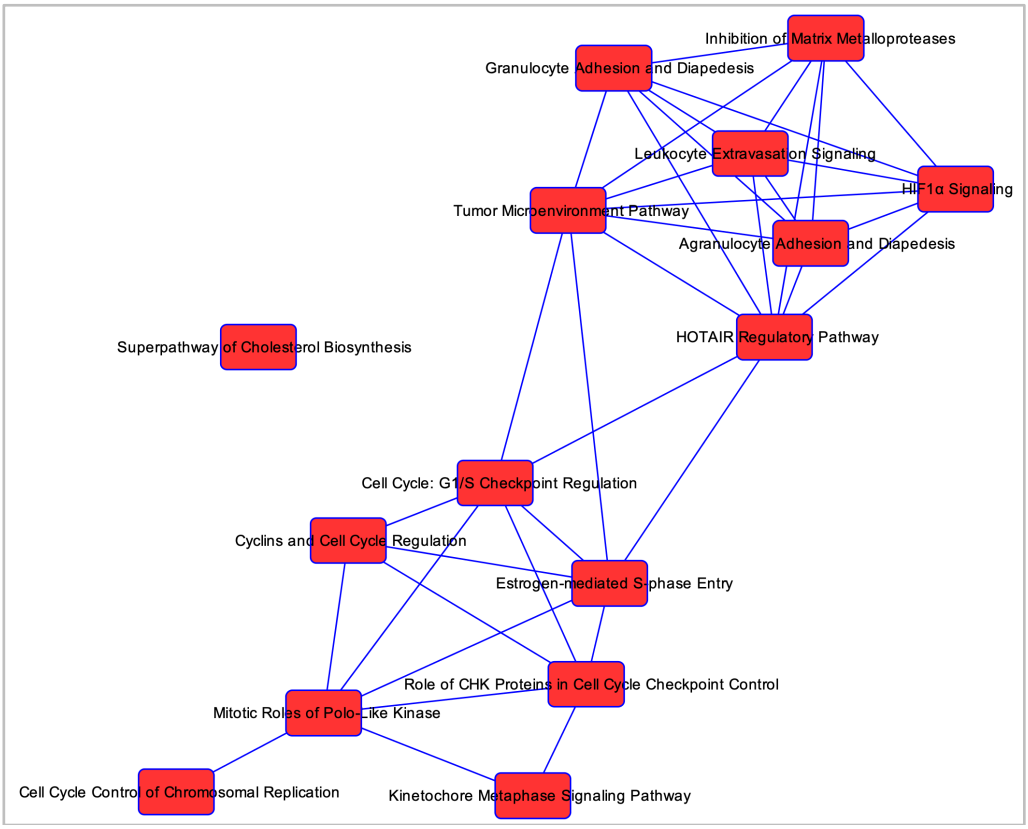


Figure 4.22 Most significant IPA canonical pathways in 393 genes upregulated in IPF AMs compared to healthy controls. (A) Bar chart displaying canonical pathways with significance values based on the right-tailed Fisher's exact test and coloured according to predicted pathway activation (orange bars) or predicted inhibition (blue bars), based on the z-score statistic. Pathways ineligible for a prediction are coloured grey. (B) Network showing overlapping canonical pathways which share sets of genes.

Table 4.2 Top canonical pathways for genes upregulated in IPF AMs

Ingenuity Canonical Pathway	Pathway category	Genes in pathway	Genes in pathway (%)
Kinetochore Metaphase Signalling Pathway	Cellular growth, proliferation and development	*AURKB,BIRC5,BUB1, BUB1B, CDC20, CENPE,ESPL1, KIF2C, NDC80, NEK2,NUF2, PLK1,SKA1, SPC24, TTK*	15/101 (15%)
Mitotic Roles of Polo-Like Kinase	Cell cycle regulation	*CCNB2,CDC20,CDC25A, CDC25C,CDC7, ESPL1, KIF11, KIF23,PKMYT1, PLK1, PLK4, PRC1*	12/66 (18%)
Granulocyte Adhesion and Diapedesis	Cellular immune response	*CCL13,CCL17,CCL22, CCL7, CLDN1,CXCL5, HRH2,IL1R2, MMP10, MMP17,MMP2, MMP25, MMP7, MMP8, PPBP, SDC2,SDC3*	17/173 (10%)
Estrogen-mediated S-phase Entry	Cell cycle regulation	*CCNA2,CCNE2,CDC25A, E2F2, E2F7,E2F8,MYC*	7/26 (27%)
Inhibition of Matrix Metalloproteases	Organismal growth and Development	*MMP10,MMP17,MMP2, MMP25, MMP7, MMP8,SDC2, TIMP3*	8/39 (21%)
Agranulocyte Adhesion and Diapedesis	Cellular immune response	*CCL13,CCL17,CCL22, CCL7, CLDN1,CXCL5, MMP10,MMP17,MMP2, MMP25,MMP7,MMP8, PODXL,PPBP*	14/193 (7%)
Role of CHK Proteins in Cell Cycle Checkpoint Control	Cell cycle regulation	*CDC25A,CDC25C,CLSPN, E2F2,E2F7,E2F8,PLK1*	7/57 (12%)
HOTAIR Regulatory Pathway	Cancer Cellular growth, proliferation and development	*DNMT3B,EZH2,FOXM1, MMP10,MMP17,MMP2, MMP25,MMP7,MMP8, MYC,PIK3R6*	11/160 (7%)
Superpathway of Cholesterol Biosynthesis	Sterol biosynthesis	*DHCR7,HMGCS1,MVD, SQLE,TM7SF2*	5/29 (17%)
HIF1α Signalling	Cellular stress and injury	*CAMK1,CAMK1G,FLT1, HGF,MMP10, MMP17, MMP2, MMP25 ,MMP7, MMP8, PIK3R6,PRKCB*	12/205 (6%)
Tumour Microenvironment Pathway	Cancer	*HGF,ITGB3,MMP10, MMP17, MMP2,MMP25, MMP7, MMP8,MYC, OSM,PIK3R6*	11/176 (6%)
Leukocyte Extravasation Signalling	Cellular immune response	*CLDN1,MMP10,MMP17, MMP2,MMP25, MMP7, MMP8,PIK3R6,PRKCB, PXN,TIMP3*	11/193 (6%)
Cell Cycle Control of Chromosomal Replication	Cell cycle regulation	*CDC45,CDC6,CDC7,CDT1, ORC1, TOP2A*	6/56 (11%)
Cyclins and Cell Cycle Regulation	Cell cycle regulation	*CCNA2,CCNB2,CCNE2, CDC25A,E2F2, E2F7,E2F8*	7/81 (9%)
Cell Cycle: G1/S Checkpoint Regulation	Cell cycle regulation	*CCNE2,CDC25A,E2F2, E2F7,E2F8,MYC*	6/67 (9%)

To understand the role of molecules in the airway microenvironment – particularly cytokines and growth factors - that may be responsible for the observed downstream gene expression changes, upstream regulator analysis was performed in IPA. Upstream cytokines or growth factors with a predicted activation z-score ≥2 were ranked and the top upstream regulator was *CSF2,* colony stimulating factor 2 (also known as GM-CSF, granulocyte-macrophage colony-stimulating factor) (Table 4.3). CSF2 (GM-CSF) regulates the proliferation[346], differentiation and function[347,348] of AMs and was predicted to activate 50 genes in the dataset which are displayed as a network according to cellular location in Fig. 4.23. Transforming growth factor-beta 1, *TGFB1,* was also predicted as a top upstream regulator and has been reported to play a role in AM development and homeostasis[349,350], and fibrosis[351].

The connectivity of genes upregulated by IPF AMs was explored using gene interaction networks (Figs. 4.24 – 4.26). These were ranked according to a network score which is based on the number of the dataset focus molecules that a network contains; a higher score indicates a lower probability of finding the genes in a given network by random chance alone. Networks of interest in the top 20 were those containing genes involved in 'Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking' (ranked #3, network score 42), 'Cancer, Connective Tissue Disorders, Organismal Injury and Abnormalities' (ranked #8, network score 22), and 'Connective Tissue Disorders, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders' (ranked #17, network score 17).

Table 4.3 Top 5 activated upstream regulators in IPF

Upstream Regulator	Fold change*	Molecule type	Predicted Activation z-score**	log10(p-value)	Number of target molecules in dataset
CSF2	-	cytokine	6.26	24.03	50
HGF	2.02	growth factor	6.06	23.42	48
AREG	-	growth factor	4.46	17.34	20
TGFB1	-	growth factor	2.70	13.46	74
IL6	-	cytokine	2.98	13.16	45

**Fold change noted if upstream regulator is present in dataset.*
***Predicated activation z-score predicts the activation state of the upstream regulator based on the expression of genes downstream. Only absolute z-scores ≥2 were considered significant and included here.*

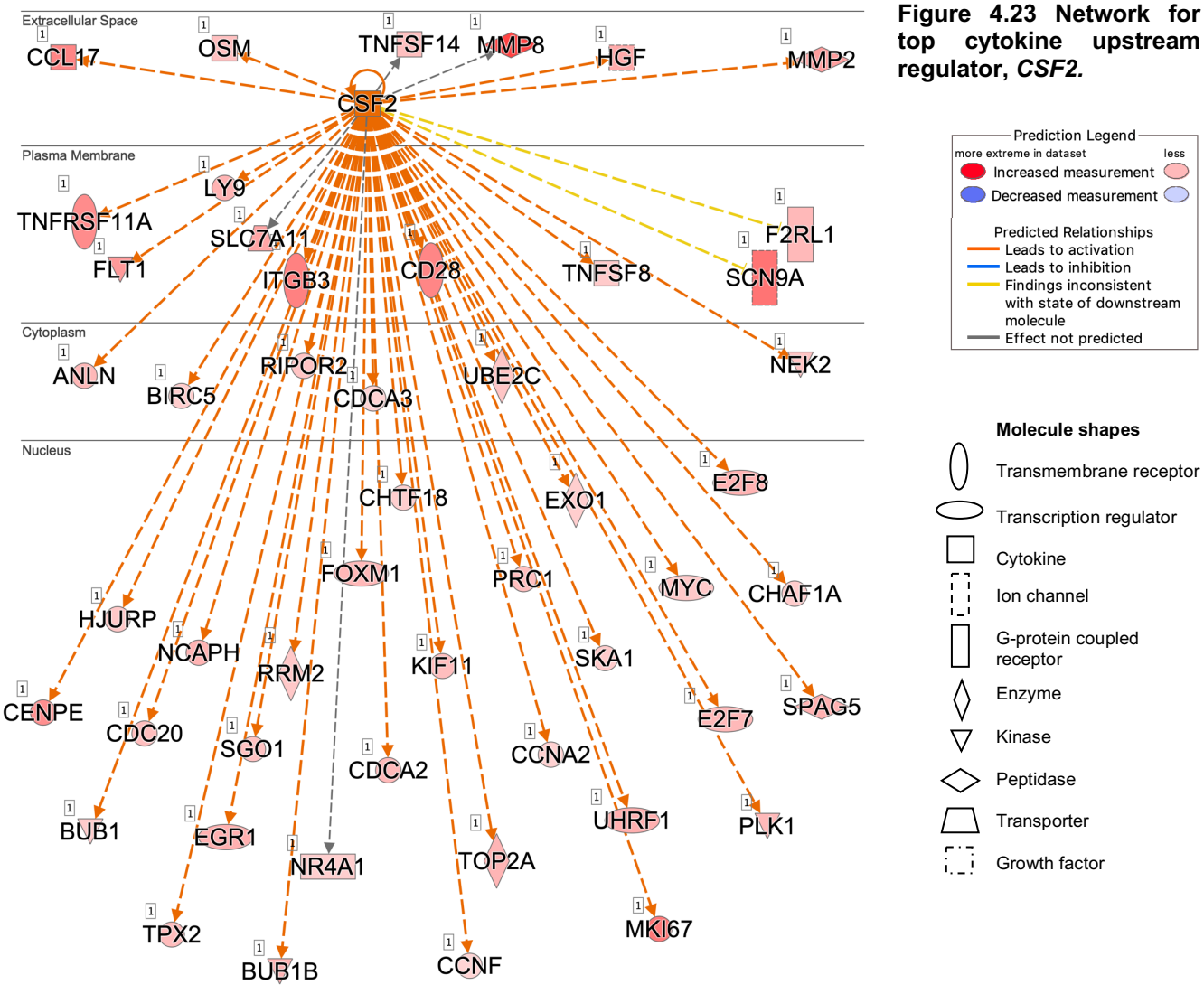


Figure 4.23 Network for top cytokine upstream regulator, *CSF2*.

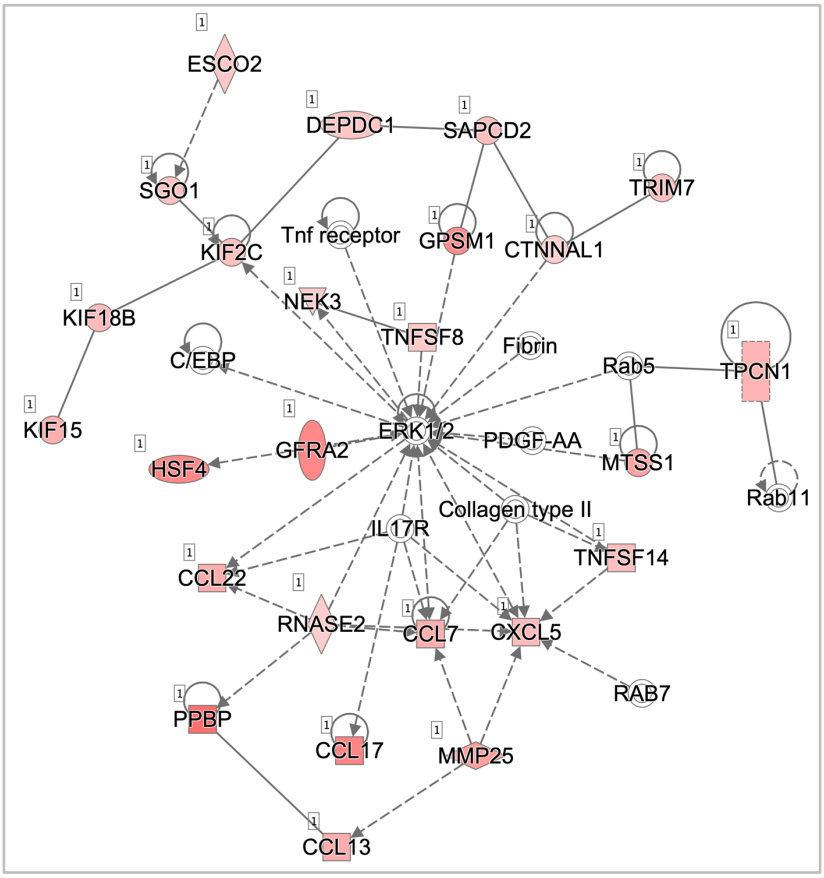


Figure 4.24 Gene interaction network: 'Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking'. Non-directional gene interaction network generated by IPA bioinformatic suite, based on genes upregulated in IPF AMs. Network is described by the top 3 most significant biological functions. Network score, calculated as the negative exponent of the right-tailed Fisher's exact test result = 42. Network contains 25 focus molecules.

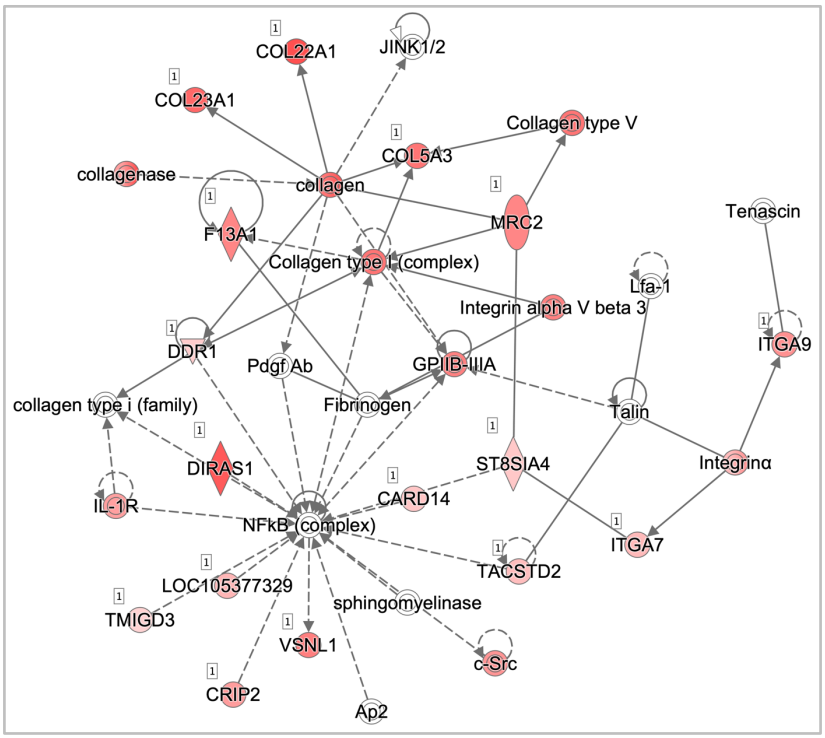


Figure 4.25 Gene interaction network: 'Cancer, Connective Tissue Disorders, Organismal Injury and Abnormalities.' Non-directional gene interaction network generated by IPA bioinformatic suite, based on genes upregulated in IPF AMs. Network is described by the top 3 most significant biological functions and diseases. Network score = 22. The network score is calculated as the negative exponent of the right-tailed Fisher's exact test result. Network contains 16 focus molecules.

Molecule fill colours

- Gene upregulated in dataset
- Gene from the Knowledgebase – not part of dataset

Relationship line descriptions

- Direct interaction
- Indirect interaction

Molecule shapes

- Transmembrane receptor
- Transcription regulator
- Cytokine
- Ion channel
- G-protein coupled receptor
- Enzyme
- Kinase
- Peptidase
- Complex/ Group
- Transporter
- Growth factor

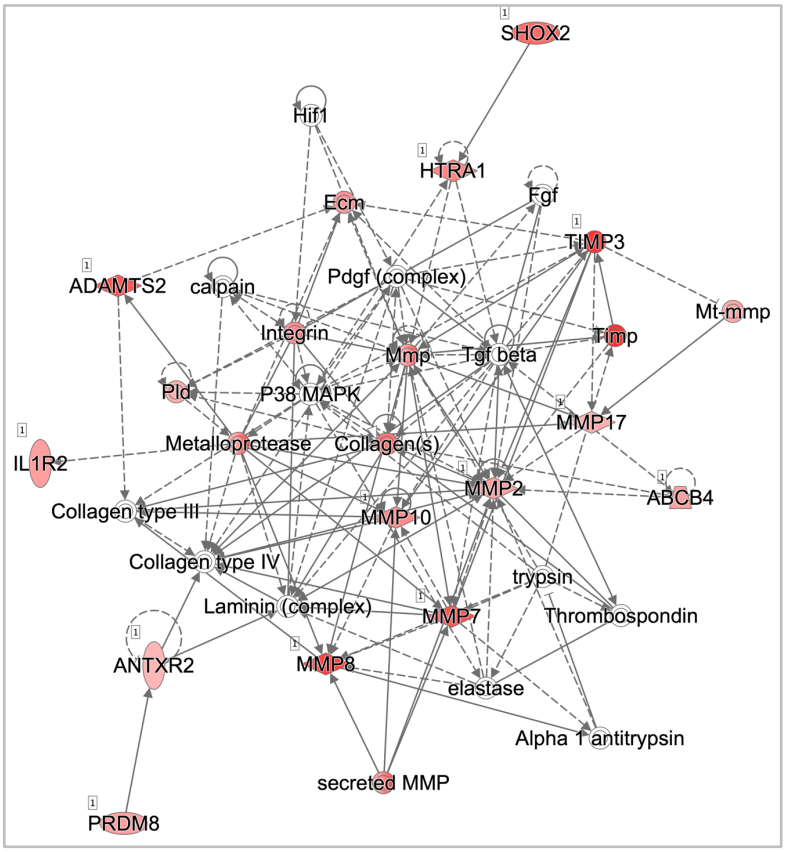


Figure 4.26 Gene interaction network: 'Connective Tissue Disorders, Organismal Injury and Abnormalities, Skeletal and Muscular Disorders.' Non-directional gene interaction network generated by IPA bioinformatic suite, based on genes upregulated in IPF AMs. Network is described by the top 3 most significant biological functions and diseases. Network score = 17. The network score is calculated as the negative exponent of the right-tailed Fisher's exact test result.

IPA was less informative with downregulated genes from the DESeq2 analysis given the relatively low number of genes (*n* = 173). The data was used to facilitate the functional annotation of downregulated genes in IPF AMs. Only canonical pathways in which ≥3 genes were present in the dataset accounting for ≥5% of the pathway were considered. Although there were too few genes per pathway to predict a z-score, it was clear that downregulated genes in IPF AMs were involved in detoxification, fatty acid degradation, the complement system and xenobiotic metabolism (Fig. 4.27, Table 4.4).

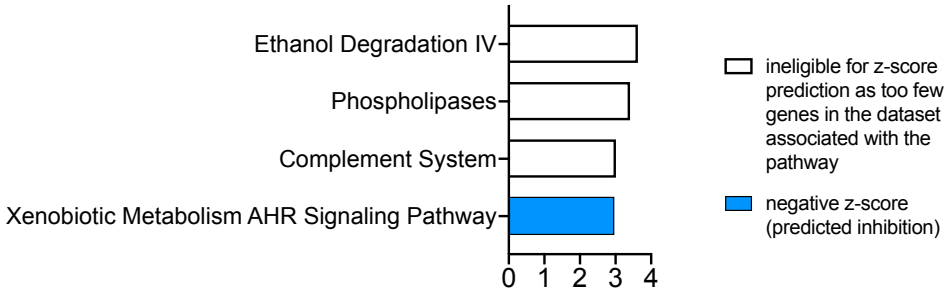


Figure 4.27 Most significant IPA canonical pathways in 173 genes downregulated in IPF AMs compared to healthy controls. Bar chart displaying canonical pathways with significance values based on the right-tailed Fisher's exact test and coloured according to the z-score statistic.

Table 4.4 Top canonical pathways for genes downregulated in IPF AMs

Ingenuity Canonical Pathway	Pathway category	Genes in pathway	Genes in pathway (%)
Ethanol Degradation IV	Ethanol degradation	*ACSS3,ALDH3A1,CYGB*	3/23 (13%)
Phospholipases	Fatty acid and lipids degradation	*PLA1A,PLA2R1,PLAAT1,PLAAT5*	4/65 (6%)
Complement System	Humoral immune response	*C8A,C8B,CFI*	3/37 (8%)
Xenobiotic Metabolism AHR Signaling Pathway	Xenobiotic metabolism	*ALDH3A1,GSTO2,IL1A,NQO1*	4/85 (5%)

4.4.7.7 Gene set enrichment analysis (GSEA)

Gene Set Enrichment Analysis (GSEA)[247] was performed on all filtered genes in the AM RNA-seq dataset (*n* = 19723), using 947 curated genes associated with pulmonary fibrosis (MESH ID: D011658) in the Comparative Toxicogenomics Database[248]. GSEA determines whether members of an *a priori* defined gene set tend to occur toward the top or bottom of a list of all genes ranked according to their differential expression between two classes[247]; IPF and healthy controls in the case of this experiment. GSEA demonstrated significant enrichment for genes related to pulmonary fibrosis (NES 1.5, FDR q-value 0.045) (Fig. 4.28) in IPF AMs compared to healthy controls. The maximum enrichment score was achieved at position 3457 in the ranked list of genes. From the list of 393 genes that were significantly upregulated in IPF AMs compared to control, 60 genes were present in the curated Comparative toxicogenomics database pulmonary fibrosis gene set (Table 4.5).

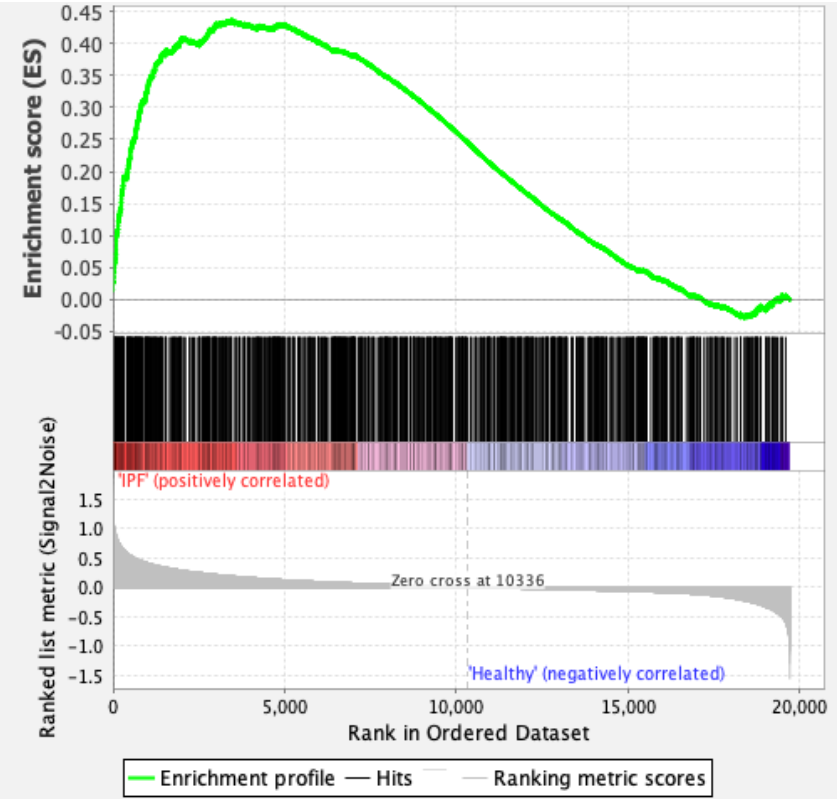


Figure 4.28 Gene Set Enrichment Analysis (GSEA) using curated pulmonary fibrosis gene set. GSEA was performed using all filtered genes (*n* = 19723) from IPF AMs compared to controls. GSEA enrichment plot with table displaying key statistics.

Enrichment score (ES)	0.44
Normalised enrichment score (NES)	1.50
FDR q-val	0.045
Rank at max	3457

Table 4.5 Genes associated with pulmonary fibrosis upregulated in IPF AMs. List of 60 genes from the curated Comparative Toxicogenomics Database Pulmonary Fibrosis Gene Set which were significantly upregulated in IPF AMs.

ABCB4	*CCNB2*	*E2F8*	*IL1R2*	*MYC*	*PRC1*
ANGPTL4	*CDC20*	*EGR1*	*KIF23*	*NCAPG*	*PRKCB*
AQP1	*CDC25A*	*FASN*	*LDLR*	*NDRG2*	*RRM2*
AURKB	*CDC25C*	*FLT1*	*MEF2C*	*NEK2*	*SDC2*
BIRC5	*CDCA7*	*GADD45G*	*MELK*	*NR4A1*	*SLC40A1*
BUB1	*CENPE*	*GINS1*	*MKI67*	*NUF2*	*SLC7A11*
CCL17	*CH25H*	*GPT*	*MMP10*	*OLFML3*	*SQLE*
CCL22	*CLDN1*	*GTSE1*	*MMP2*	*OSM*	*STIL*
CCL7	*CXCL5*	*HGF*	*MMP7*	*PLK1*	*TIMP3*
CCNA2	*E2F2*	*HMGCS1*	*MMP8*	*PLK4*	*TOP2A*

4.4.7.8 Comparison with published datasets

Single-cell RNA-seq has been used to characterise macrophage populations from IPF lung tissue[53,163,164] and healthy BAL[164]. Morse et al. described three discrete subpopulations of macrophages from lung explant tissue; SPP1^{hi}, FABP4^{hi} and FCN1^{hi} macrophages. SPP1^{hi} macrophages demonstrated marked upregulation of *SPP1* and *MERTK* in the lower lobes of the lungs in IPF[164]. The expression of the marker genes used to define these macrophage subsets was examined in our bulk RNA-seq dataset from AMs purified from BAL (Fig. 4.29). Consistent with Morse et. al.[164], *SPP1* and *MERTK* expression was increased in IPF AMs compared to controls. Thus, SPP1^{hi} macrophages from the interstitial compartment around fibroblastic foci in the Morse et. al. dataset, were likewise detected in our RNA-Seq data from CD206+ airway macrophages. Monocyte marker gene *FCN1,* encoding lectin ficolin-1[164], was also elevated in IPF AMs compared to controls in our dataset.

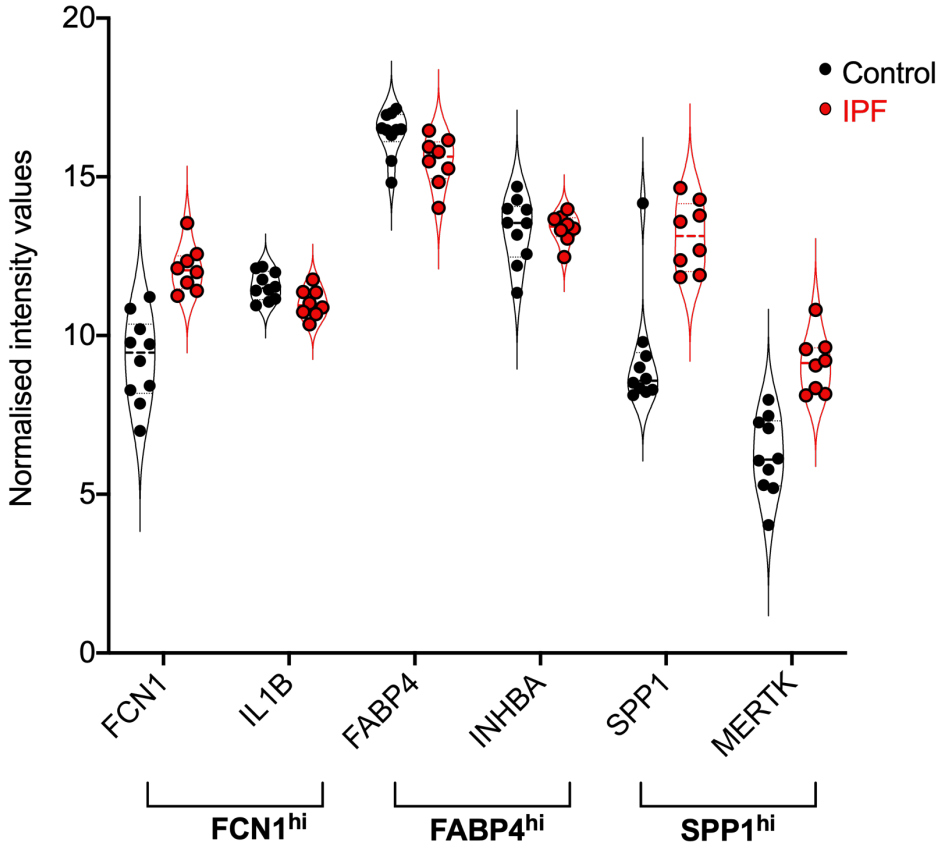
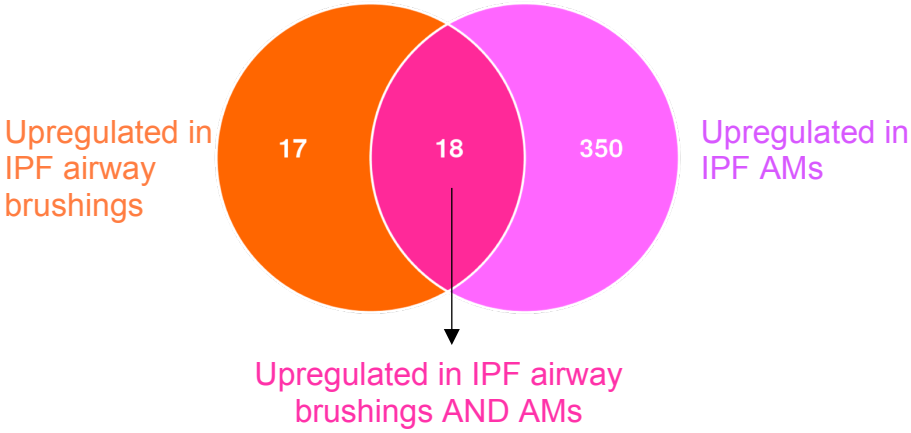


Figure 4.29 Violin plot showing expression of macrophage subset marker genes. Normalised expression of marker genes defining macrophage subpopulations described by Morse et al.[164] were examined in our RNA-seq dataset of CD206+ AMs from IPF subjects (*n* = 8) and healthy controls (*n* = 10). Dashed line indicates median expression value.

4.4.7.9 Comparison of differentially expressed genes in paired airway brushings and sorted AMs

Next, I sought to understand the overlap in gene expression between the paired airway brushing and airway AM samples obtained in our study. There were 18 genes common to both airway brushings and sorted AMs that were upregulated in IPF compared to controls (Fig. 4.30). These included genes involved in cell cycle regulation (*CDCA2, DLGAP5, MYBL2, TOP2A)* and chemokines *PPBP/ CXCL7* and *CCL13* which were upregulated in both airway brushings and CD206+ AMs in IPF compared to controls. The expression of these chemokines in IPF AMs did not exhibit a significant positive correlation with disease extent estimated by CPI, as shown in airway brushings (Fig. 4.21 A and C). *FAM111B* was also upregulated in the paired samples in IPF compared to controls in our dataset. Of note, mutations in this gene have been reported to cause hereditary fibrosing poikiloderma (HFP) characterised by poikiloderma, tendon contractures, and progressive pulmonary fibrosis[352]. There was no overlap in genes that were significantly downregulated in airway brushings and AMs.



ANLN	*DLGAP5*	*NRGN*
BIRC5	*FADS1*	*PPBP*
BUB1B	*FAM111B*	*RNASE2*
CCL13	*GFRA2*	*SHCBP1*
CDCA2	*MTND3P19*	*TOP2A*
CEP55	*MYBL2*	*UBE2C*

Figure 4.30 Venn diagram showing overlapping upregulated genes in IPF airway brushings and airway macrophages. Number of upregulated genes in IPF compared to control (DESeq2; pAdj <0.05, fold change ≥2) for each sample type indicated. The number of genes shown on this Venn diagram is lower than that listed previously for differential expression testing as there were some entities which did not match NCBI symbols, such as pseudogenes.

4.5 **Discussion**

4.5.1 Airway brushings

Airway epithelial cells and AMs are crucially positioned at lung-environment interfaces and play a pivotal role in immune defence of the lower airspace. Bulk RNA-seq of samples collected from IPF patients and healthy volunteers during bronchoscopy has provided novel insight into transcriptomic changes occurring in the airway environment during the evolutionary stages of the lung fibrosis. The findings presented in this Chapter provide evidence of immunomodulatory changes in the airway epithelium in IPF and reinforce the significant role played by airway macrophages in disease pathogenesis. On this basis, the Chapter 4 hypothesis was accepted.

A core aim of this work was to sequence the transcriptome of key structural and immune cells in the IPF airway. Multi-omics profiling of 12 organs from mice, emphasised the importance of immune regulation by structural cells such as epithelial cells[353]. In our study, two prominent populations regulating immune homeostasis were studied from the same human subjects; epithelial cells collected by bronchial brushings and airway macrophages enriched from BAL. This novel approach also provided spatial gene expression information as cells from bronchial brushings were sampled from a more proximal location in the conducting airways than the AMs which were sampled from BAL retrieved from the distal airspace.

This is the first study to evaluate the transcriptome of cells obtained from proximal airway brushings in IPF patients. A bulk RNA-seq approach was used which averages the transcriptomes of all cells in a sampled population, providing a global unbiased perspective of gene expression changes. An important advantage of bulk RNA-seq is the detection of differential gene expression for low- abundance transcripts[354]. An unsupervised clustering approach of the most variable entities from the RNA-seq dataset, did not reveal any grouping of samples by condition but sex-linked genes were responsible for some clustering by gender. There were relatively few differentially expressed genes between IPF and healthy airway brushings, but interestingly there was significant upregulation of *KRT13*, encoding a type 1 cytokeratin. A basal- to- club cell transitional population of *Krt13* expressing cells was first described in the murine

tracheal epithelium[200]. *Krt13*+ cells were arranged in stratified, high-turnover structures called “hillocks” and co-expressed gene markers of squamous cell differentiation, cell adhesion and immunomodulation. In a parallel study, *Krt4/Krt13*+ cells emerged in the murine tracheal epithelium after polidocanol-induced injury[201]. In the healthy human airways *KRT13* was expressed by a population of cycling cells defined by expression of proliferation marker genes such as *MKI67* and *TOP2A*[168,179]. *KRT13* was also detected in a cluster of nasal epithelial cells which co-expressed markers of squamous cell differentiation such as sciellin, *SCEL*[179,355]. Hierarchical clustering of differentially expressed genes in my dataset demonstrated *SCEL* was significantly upregulated and clustered with *KRT13.* There was also a module of genes (Module A) associated with cell proliferation (*TOP2A, CDCA2).* Given the nature of bulk RNA-seq it was not possible to ascribe the expression of these genes to a particular epithelial cell subset as has been achieved with single-cell RNA-seq[168,179]. The presence of KRT13- expressing cells on IPF airway brushings was validated by immunofluorescence microscopy. Rare groups of KRT13- positive cells were demonstrated by immunostaining IPF proximal airways in tissue sections. These cells were often in areas of epithelial damage and were absent from control airway tissue sections.

Hierarchical clustering of significant differentially expressed genes in airway brushings identified two distinct transcriptional profiles in IPF patients which corresponded to disease extent estimated by the CPI. Focusing on individual genes discriminating between the IPF subgroups, I discovered a significant positive correlation between chemokines *CCL13, CCL24* and *PPBP/ CXCL7* and CPI. CC family chemokine, CCL13 is chemoattractant to monocytes, eosinophils, basophils and T cells[356] and is upregulated in the bronchial epithelium in asthma[357]. High expression of *CCL13* was discovered in IPF lung tissue[358]. This chemokine was also evaluated as a candidate blood plasma biomarker for IPF in a post hoc analysis of the ASCEND trial but did not reach statistical significance[359]. The identification of increased *CCL24* expression in the IPF airway brushings compared to controls, is indirectly supported by a study which demonstrated - using a targeted mass spectrometry-based proteomic assay - that CCL24 was upregulated in IPF BAL compared to controls[360]. CCL24 / Eotaxin-2 is produced by epithelial cells[361] and immune cells including monocyte derived-

macrophages[362]. It promotes immune cell trafficking of eosinophils via CCR3[363]. *In vitro* CCL24 has been reported to induce fibroblast proliferation and collagen synthesis[364], and enhanced airway epithelial cell chemotaxis, proliferation, and wound closure[365]. An anti-CCL24 monoclonal antibody reduced immune cell infiltration and collagen deposition in a bleomycin mouse model of lung fibrosis[366]. Pro-platelet basic protein (PPBP) which belongs to the CXC chemokine family (CXCL7), is synthesised and released by activated platelets, and is strongly chemotactic for neutrophils[367]. Levels of CXCL7 protein were significantly increased in IPF BAL[360].

Taken together, transcriptomic data derived from IPF bronchial brushings suggest the airway epithelium may orchestrate immune cell trafficking in IPF through the release of specific chemokines, and this may have prognostic relevance. Higher expression of *KRT13*, markers of proliferation and squamous cell differentiation, suggest alterations to airway epithelial cell phenotype and barrier function may also be a feature of IPF. Linking these observations to pathology in the distal lung will require further mechanistic studies. The role of KRT13 in epithelial integrity and differentiation could potentially be investigated using *in vitro* air-liquid interface culture models or *Krt13* knockout mice[368].

A key drawback of a bulk RNA-seq approach is the lack of information regarding the heterogeneity of cellular populations sampled by the airway brushings. This sequencing approach does not identify gene expression changes specific to each epithelial cell type, but provides an average readout from all cells in the airway mucosa. Overall, a small number of differentially expressed genes were detected between IPF and healthy airway brushings, and this may be due to cellular 'impurities' in the unsorted samples. The expression of macrophage marker genes in the dataset, combined with results of CIBERSORTx cellular deconvolution, raised the possibility that contaminating immune cells may have masked a distinct epithelial transcriptomic signature. Crucially, the presence of closely associated immune cells on the brushings may be reflective of important cell interactions *in vivo* which may contribute to disease pathogenesis. Therefore, in Chapter 5, a single-cell RNA sequencing approach was used to comprehensively chart epithelial and immune cell populations present on the brushings, and unravel interactions between these cell types.

4.5.2 Airway macrophages

Most transcriptomic studies in IPF have analysed lung tissue homogenates or peripheral blood[328,330]. Studies specifically characterising IPF alveolar macrophages with single-cell and bulk RNA-seq approaches have generated cell suspensions from dissociated lung tissue of patients with end-stage disease[163,164], rather than using BAL obtained by bronchoscopy. The cold ischaemic time of harvested lung tissue[369] and the process of enzymatic tissue dissociation[370] can introduce artefacts in the transcriptomic data obtained from such specimens. There has been one previously published bulk RNA-seq study of BAL cells isolated via bronchoscopy from patients with IPF and healthy controls[371]. Lee et al. did not purify macrophages from BAL, but instead evaluated macrophage recovery (>80%) using cell differential counts. Also notable was the lower RIN values of the RNA samples sequenced (average 6.3 for controls, 5.4 for IPF patients) compared to those in my study. Regardless, consistent with our dataset, a sentinel finding from the RNA-seq analysis by Lee et al. was enrichment in chemokine signalling pathways, with upregulation of CC chemokine *CCL13,* CXC chemokines *CXCL5* and *CXCL7/ PPBP*[371]*.* Upregulation of additional chemokine genes in our AM dataset, such as macrophage- derived chemokine, *CCL22* may perpetuate AM recruitment and activation in IPF[372].

Computational approaches - GO enrichment analysis[243,244] and ingenuity pathway analysis (IPA)[245] - were used to capture the function of groups of genes acting in concert in a particular biological or cellular process. GO biological processes including ‘extracellular matrix organisation’, ‘positive regulation of cell communication’ and ‘myeloid leucocyte migration’ were enriched in gene module ‘A’ which better discriminated CD206+ AMs from IPF subjects compared to healthy controls in the unsupervised hierarchical clustering. Differentially expressed genes between IPF and controls in AMs were unified by pathway themes including regulation of cell cycle and cell proliferation, cellular adhesion, movement and immune cell trafficking. Matrix metalloproteinases including *MMP2, MMP7, MMP8,* and *MMP10* were also significantly upregulated ($\geq$ 2 log fold change) in IPF AMs. MMPs degrade components of the ECM and have other diverse roles which promote the development of pulmonary fibrosis[373]. For example matrilysin/ MMP7, a serum biomarker of IPF[374], was

previously shown to localise to hyperplastic epithelial cells and alveolar macrophages in IPF lungs[375], and may mediate E-cadherin ectodomain shedding and cell migration and activation in the injured epithelium[376]. In murine *in vitro* and *in vivo* models of airway epithelial cell wound repair, MMP7 suppressed ciliated cell differentiation[314]. These studies suggest that in addition to matrix degradation, the expression of *MMP7* by AMs that we detect in IPF may influence epithelial cell function. The most strongly upregulated gene in IPF AMs was tissue inhibitor of metalloproteinase 3, *TIMP3* which has been demonstrated in fibroblastic foci in IPF and shown to be upregulated by the profibrotic mediator, TGF-β1[377]. Emphasising the close relationship between IPF AMs and the fibrotic microenvironment of the distal lung, genes linked to cell- matrix interactions were upregulated such as integrin subunit beta 3, *ITGB3* and syndecan 2 SDC2. Our GSEA confirmed a significant enrichment of genes linked to pulmonary fibrosis in IPF AMs.

Colony stimulating factor 2, CSF-2, also known as granulocyte-macrophage colony-stimulating factor (GM-CSF) was predicted by IPA as the top upstream cytokine regulator in IPF AMs. GM-CSF is a critical cytokine regulating AM development and differentiation from fetal monocytes[172,347]. Multiple sources of GM-CSF have been reported including lung epithelial cells[172,378]. Disruption in GM-CSF signalling causes pulmonary alveolar proteinosis (PAP) characterised by AM dysfunction and alveolar surfactant accumulation[379,380]. Following depletion of lung macrophages in mice, tissue-resident AMs were replenished by local proliferation driven by M-CSF and GM-CSF[346]. The activation of cell cycle genes in IPF AMs in our dataset was predicted to be driven by CSF-2/ GM-CSF. The source of CSF-2 is not predicted by IPA but it is possible, given their close spatial proximity, that epithelial cells in the fibrotic airway niche produce CSF-2 facilitating profibrotic AM proliferation in IPF. Single-cell RNA-seq and spatial transcriptomics could be used to test this hypothesis, as has been applied to investigate the role of macrophage colony-stimulating factor (M-CSF) in a murine model of asbestos-induced lung fibrosis[381].

A significant strength of the work described in this Chapter, was careful consideration of sample purities and emphasis on ensuring high-quality RNA samples were used for sequencing. The input material of an RNA-seq experiment is critically important in

obtaining robust transcriptomic data. AMs were enriched on the basis of CD206 expression using FACS and where not feasible due to sample collection time, MACS. This approach improved confidence that the transcriptomic profile obtained was of AM derivation rather than other immune cell impurities in the sample. Reyfman *et al.* FACS- sorted alveolar macrophages from explant lung tissue using a similar gating strategy, selecting cells that were $CD206^{+}CD169^{+}HLA\text{-}DR^{+}autofluorescence^{hi}$, and discovered a similar pattern of significantly upregulated genes and enriched biological processes[163]. It is important to note that Reyfman *et al*. isolated AMs from tissue obtained from patients with end-stage lung disease (requiring supplementary oxygen therapy) undergoing transplant and included a heterogenous group of fibrotic ILDs including IPF (*n* = 2 subjects) plus other ILDs (systemic sclerosis associated-ILD, myositis-associated ILD and hypersensitivity pneumonitis). In contrast, I was able to identify IPF AMs displaying a fibrotic transcriptional signature using cells isolated from BAL in earlier disease (no supplementary oxygen used by any of the recruited subjects). Hence, the pathways defined in my study perhaps contribute to disease progression at an earlier stage.

The data discussed in this Chapter demonstrate that the mannose receptor CD206, can be used to enrich for airway macrophages in IPF which herald a distinct transcriptomic profile reflecting matrix interactions and remodelling, cell movement and proliferation. CD206+ is expressed by alveolar macrophages[166,333], and monocyte-derived cells in the human lung[333]. Whilst a single-cell RNA-seq approach would delineate macrophage heterogeneity in the samples, digital cytometry using our bulk RNA-seq data and CIBERSORTx showed the sample comprised a spectrum of macrophage polarisation states including uncommitted (M0), alternatively (M2) activated and monocyte-like macrophages. Transcriptomics has shown that the traditional M1 versus M2-polarisation model is an oversimplification, and that a spectrum of macrophage activation states exist[382]. Early immunophenotyping studies of BAL examining the differentiation state of alveolar macrophages in IPF demonstrated a greater proportion of immature, monocyte- like macrophages in pulmonary fibrosis[383]. Genetic lineage-tracing in a bleomycin and adenoviral TGF-β-induced model of pulmonary fibrosis showed monocyte-derived macrophages were crucial to the development of fibrosis[162].

Microenvironmental cues are critically important in shaping the chromatin landscape of macrophages[165], therefore further investigation is required to determine the extent to which the fibrotic niche may reprogram the transcriptome of AMs in IPF. Cadherin-11 mediated interactions between macrophages and myofibroblasts have been discovered in fibrotic lung tissue[186]. Important crosstalk also exists between macrophages and epithelial cells in the airway to modulate immune responses[181] and epithelial repair[384], but little is understood about interactions that occur in IPF. In the next Chapter, single-cell RNA-seq of airway brushings was used to further explore epithelial - AM interactions in IPF.

CHAPTER 5

Single-cell profiling of the IPF airway epithelial niche

5 Single-cell profiling of the IPF airway epithelial niche

5.1 Background and rationale

Bulk transcriptomic analysis of airway brushings in Chapter 4 revealed a relatively small number of differentially expressed genes between IPF and healthy control RNA-Seq samples. It was clear from flow cytometric analysis and immunostained cytospins that airway brushings comprised a mixed population of both epithelial and immune cell subsets concordant with gene expression in the same samples (Fig. 5.1 A and B). As bulk RNA-seq provides a composite gene expression readout of all the cell types present, we reasoned that the degree of cellular heterogeneity present in airway brushings may be masking important cellular and molecular changes occurring in individual cell subsets. Furthermore, immune populations closely associated with the epithelium may provide important information regarding the pathogenesis of IPF. Given the inherent difficulties of flow cytometric sorting of low numbers of human airway epithelial cells from brushes, a single-cell RNA-seq (scRNA-seq) approach was used to directly profile the cellular populations present (Fig. 5.1 C).

IPF has been the focus of a growing number of scRNA-seq studies which have used distal lung tissue from donors with advanced disease[53,145,146,163,164,189,199]. The diverse cellular landscape of the IPF distal lung has been unveiled through high-resolution scRNA-seq data and state of the art computational approaches. An online portal called the IPF Cell Atlas has been created to disseminate the invaluable single-cell data from four key studies[385]. There is universal agreement from single-cell studies that a shift in cellular composition occurs in the IPF distal lung, from alveolar epithelial cells to conducting airway epithelial cell types, a process referred to as 'bronchiolisation' [53,145,164,189]. Two independent single-cell studies[53,145] discovered a novel *KRT5-KRT17+* epithelial cell - also termed 'aberrant basaloid cell'. This cell expresses some but not all basal cell marker genes, mesenchymal genes such as *COL1A1*[386], developmental transcription factor *SOX9*[387], senescence associated genes including *CDKN2A*[388] and *MMP7* encoding matrilysin[126], which have been implicated in IPF pathogenesis. Concordant with the concept of 'bronchiolisation' in the IPF lung, *COL15A1*- expressing vascular endothelial cells - normally found around large proximal bronchi - were discovered in the distal fibrotic lung[53]. A profibrotic alveolar

macrophage population expressing *SPP1* has been consistently reported in single-cell studies[53,163,164]. Transcriptionally distinct fibroblast populations occupying unique anatomical niches in the healthy and fibrotic lung have been identified[145,146], as well as a pathological *CTHRC1* expressing fibroblast unique to the fibrotic lung[146].

The lung tissue used for existing single-cell studies was derived from patients with end-stage pulmonary fibrosis undergoing lung transplant. It is not always possible to deduce from these studies whether the cell types present and/ or molecular pathways upregulated are a unique feature of IPF pathogenesis or simply reflect advanced lung disease. Adams et al.[53] included explant tissue from COPD patients to control for the effects of end-stage lung disease, and interestingly discovered a small number of aberrant basaloid cells in these lung tissues suggesting they may be a common pathological feature of both COPD and IPF. Early- stage disease has not been studied, and merits investigation as it may reveal important transcriptomic changes in cell states, interactions and molecular pathways that contribute to disease progression.

Single-cell transcriptomics has been used to study the airway epithelium and associated immune cells in healthy subjects[179], smokers[198,389], asthmatic patients[168] and COVID-19 patients[180]. Through use of this technology in these studies, different cell states associated with disease have been exposed, differentiation trajectories for rare epithelial cell types have been delineated and cell-cell interactions defined. In view of the shift from alveolar to airway epithelial cell types in the IPF distal lung, I used scRNA-seq to investigate the proximal airway epithelial niche in IPF subjects with early disease. Closer examination of this previously unstudied region of the IPF lung will further our understanding of pathways driving disease initiation and progression.

5.2 Hypothesis

At single- cell resolution there are transcriptomic changes in the epithelial and immune cell populations of the IPF airway niche, and cell-cell interactions that contribute to disease pathogenesis.

5.3 Aims

1. To profile the single-cell RNA landscape of the IPF and healthy airway epithelial niche using bronchial brushes obtained via bronchoscopy.
2. To determine the transcriptomic differences in each airway cell subtype in IPF.
3. To interrogate basal cell progenitor populations in IPF and healthy airways.
4. To assess epithelial differentiation dynamics using a computational approach.
5. To investigate cell-cell interactions taking place in the airway that may contribute to disease.

5.4 Results

5.4.1 Clinical characterisation of subjects recruited

Ten subjects with a diagnosis of IPF following multidisciplinary discussion, and seven healthy control subjects were prospectively recruited and underwent bronchoscopy at the Royal Brompton Hospital between September 2019 and February 2020. The clinical characteristics of our IPF cohort reflect the typical demographic profile associated with the disease[14] (Table 5.1). IPF subjects were almost all male (90%) with a mean age of 69 years ± 10. There was a more equal gender balance in the healthy control group (43% male, 57% female), and a mean age of 46 years ± 9. Baseline pulmonary function tests (PFTs) demonstrated relatively preserved lung volumes (mean FVC, 81% predicted ± 19), but impaired gas transfer (mean haemoglobin- corrected diffusion capacity for carbon monoxide, DL_{co} 48% predicted ± 11) in the IPF cohort. There was no evidence of airflow obstruction (FEV_1/ FVC ratio <0.7) in any of the subjects recruited. The composite physiological index (CPI), which correlates closely with the extent of fibrosis on CT and is a more accurate predictor of mortality than individual PFT values[19], was between 37 – 59. None of the subjects were taking steroids at the time of sampling. IPF subject 5756 was found to be colonised with *Mycobacterium intracellulare* (MAI) and *Klebsiella pneumoniae* (scanty growth) on culture testing of bronchoalveolar lavage (BAL) fluid. In all other subjects there was no clinical evidence of respiratory tract infection. The minor allele (T) of the *MUC5B* promotor polymorphism SNP rs35705950[90] was present in 14% of healthy

controls and 100% of IPF patients. Two to three bronchial brushes were obtained from all subjects and processed for scRNA-seq.

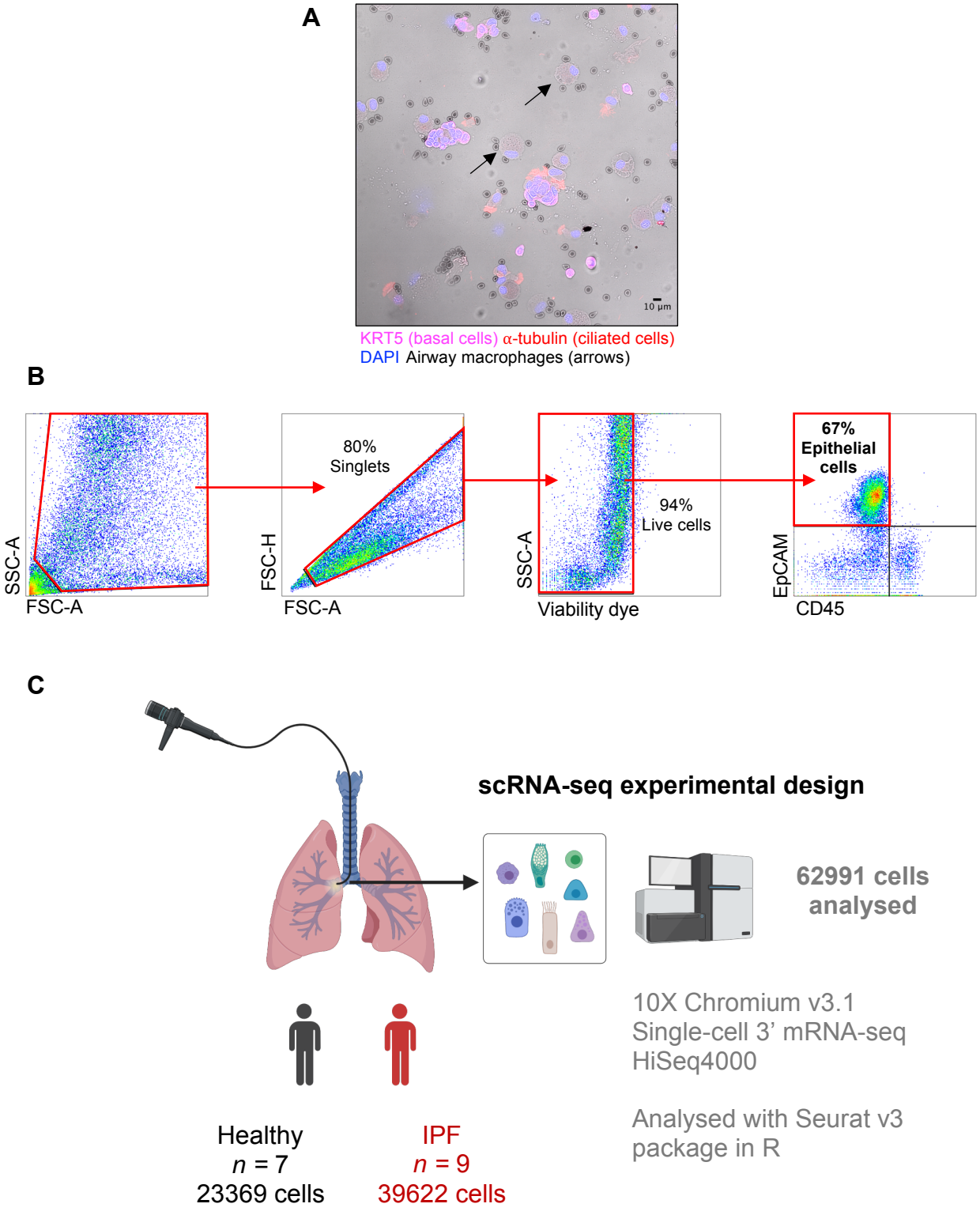


Figure 5.1 Single-cell RNA sequencing to study heterogeneous cell populations in airway brushings. Traditional approaches to identify cell populations including (**A**) cytospin immunostaining, and (**B**) flow cytometric analysis, revealed that airway brushings were comprised of epithelial and immune populations. (**C**) Experimental approach using single-cell RNA seq to explore cellular populations present in the airway niche. Figure created with BioRender.com.

Table 5.1 Clinical characteristics of subjects recruited for scRNA-seq

Subject ID	Diagnosis	Age	Sex	Smoking status	FEV_1 (%pred)	FVC (%pred)	FEV_1/FVC ratio	DL_{CO} (%pred)	CPI	*MUC5B* (rs35705950) genotype
1055	Healthy	47	M	Ex	3.98 (134.2)	4.98 (136.6)	0.80	n/a	n/a	G/G
1056	Healthy	26	F	Ex	2.79 (80.1)	3.21 (78.7)	0.87	n/a	n/a	G/G
1057	Healthy	51	M	Ex	2.86 (92.7)	3.44 (89.5)	0.83	n/a	n/a	G/G
1058	Healthy	47	F	Never	3.39 (108.5)	3.96 (101.4)	0.86	n/a	n/a	G/G
1059	Healthy	50	F	Never	2.4 (89.7)	3.08 (92.4)	0.78	n/a	n/a	G/G
1060	Healthy	55	F	Never	2.74 (92.2)	3.56 (94.4)	0.77	n/a	n/a	G/G
1061	Healthy	44	M	Ex	4.14 (99.8)	4.91 (94.5)	0.84	n/a	n/a	G/T
5756*	IPF	73	M	Never	2.98 (113)	3.42 (98.7)	0.87	4.81 (61.1)	37	G/T
5758	IPF	74	M	Never	2.21 (88)	2.48 (75)	0.89	3.04 (40)	55	G/T
5759	IPF	79	M	Never	2.14 (108.9)	2.47 (93.1)	0.87	2.44 (39.5)	53	G/T
5760	IPF	71	F	Ex	2.35 (102.3)	3.30 (120.3)	0.71	2.55 (33.6)	40	G/T
5761	IPF	84	M	Never	2.27 (99)	2.73 (86)	0.83	3.98 (56)	43	G/T
5764	IPF	61	M	Ex	2.42 (85.2)	2.68 (74.9)	0.90	5.51 (66.5)	37	G/T
5765	IPF	50	M	Never	2.10 (56.9)	2.32 (50.6)	0.91	3.92 (37.8)	59	G/T
5767	IPF	71	M	Ex	2.71 (80.7)	3.49 (79)	0.78	5.38 (55.3)	41	G/T
5772	IPF	66	M	Never	2.48 (88)	2.78 (76.9)	0.89	3.30 (39.8)	54	G/T
5778	IPF	65	M	Ex	2.57 (63.4)	3.25 (60.2)	0.79	5.20 (46.7)	50	G/T

FEV1 – forced expiratory volume in 1 second, FVC – forced vital capacity, DL_{co} – diffusing capacity of the lung for carbon monoxide (haemoglobin-adjusted), CPI - Composite physiological index = 91 – (0.65 x % predicted TLCO) – (0.53 x % predicted FVC) + (0.34 x % predicted FEV1)

*Positive BAL culture - *Mycobacterium intracellulare* (MAI) and *Klebsiella pneumoniae* (scanty growth)

5.4.2 Sample quality control

The mean cell viability (as assessed by an EVE cell counter) was 58% ± 15%, with 414 cells/μl ± 265 obtained per sample (Table 5.2). Of note, IPF sample 5760 was excluded from further analysis at this stage on account of very low cell viability (27%) and the low number of cells sequenced (203 cells). After filtering out empty droplets using the EmptyDrops algorithm[390] in Cell Ranger, an average of 3937 cells per subject were achieved. In total there were a median of 2682 genes per cell and 50722 reads per cell. An R package, DoubletFinder[253] was used to identify the number of cell doublets per sample, showing a mean doublet rate of 7.5%. Our doublet rate for an average of 9836 cells loaded per subject was higher than the stochastic doublet rate quoted in the Chromium Single Cell 3' Reagent Kits v3 User Guide (10X Genomics). As will be discussed later in the Chapter, some doublets may be due to closely interacting cells *in vivo.* In total, 62991 single cells were taken forward for analysis.

Table 5.2 Sample characteristics for scRNA-seq

Subject ID	Diagnosis	Cell viability (%)	Cells loaded	Prefiltered cell no.	Filtered cell no.*	Mean reads per cell	Median genes per cell	Predicted doublets no.**
1055	Healthy	55	9075	4126	3736	39091	2204	280
1056	Healthy	68	8360	4219	3689	46503	2582	277
1057	Healthy	74	3850	3203	2767	54794	2320	208
1058	Healthy	75	10560	5293	4773	47092	3015	358
1059	Healthy	40	12352	4609	4257	55363	2938	319
1060	Healthy	41	10105	2380	1804	46702	1923	135
1061	Healthy	78	13124	3302	2343	110874	2684	176
5756	IPF	49	3234	1127	985	81587	2787	74
5758	IPF	67	14916	6276	5369	21591	1831	403
5759	IPF	66	12540	8117	7399	29752	2639	555
5760 [†]	*IPF*	*28*	*4667*	*203*	*n/a*	*17751*	*2123*	*n/a*
5761	IPF	75	13200	5118	4641	38975	2624	348
5764	IPF	43	8360	2773	2590	47662	3727	194
5765	IPF	64	12408	8631	7974	36092	3204	598
5767	IPF	55	8800	2462	2180	73384	3292	164
5772	IPF	68	11270	6128	5664	43323	3067	425
5778	IPF	44	10395	3135	2820	71736	2631	212

*Filtered cells passing CellRanger's *EmptyDrops* algorithm were taken forward for further analysis.
***DoubletFinder* R package was used to identify (but not remove) doublets.
[†] Sample 5760 excluded from further analysis.

5.4.3 Data quality control

The data was pre-processed with the Seurat v3.0 pipeline[252] in R v3.6.0. After excluding cells with fewer than 500 unique genes, 1000 mapped reads and/ or a high percentage of reads mapping to the mitochondrial genome (> 3 standard deviations from the median) a total of 60462 cells (healthy control *n* = 22559 from 7 subjects; IPF *n* = 37903 from 9 subjects) were taken forward for the integrated analysis. Taking into account the elbow- plot of PCA eigenvalues (Fig 5.2 A), and the p-value distribution of PCs calculated by the jackstraw method[391] (Fig. 5.2 B), the first 50 PCs were used to construct Uniform Manifold Approximation and Projection (UMAP) plots for the integrated data. Using a resolution of 0.2 within the FindClusters Seurat function, revealed seventeen distinct clusters (cluster 0 – 16). A greater percentage of mitochondrial reads were noted in clusters 8, 9 and 11 (Fig. 5.3 A). Higher mitochondrial read counts often suggest low-quality or dying cells, however the average number of unique genes (Fig. 5.3 B) and read counts per cell (Fig. 5.3 C) in clusters 8,9 and 11 does not support this. In support of this higher mitochondrial gene expression has been described by Plasschaert et al. in the context of an epithelial cell cluster expressing *SLC16A7*[201]. Predicted doublets were highest in cluster 7 (49.5%), cluster 13 (32.9%) and cluster 15 (22.7%) (Fig. 5.4, Table 5.3).

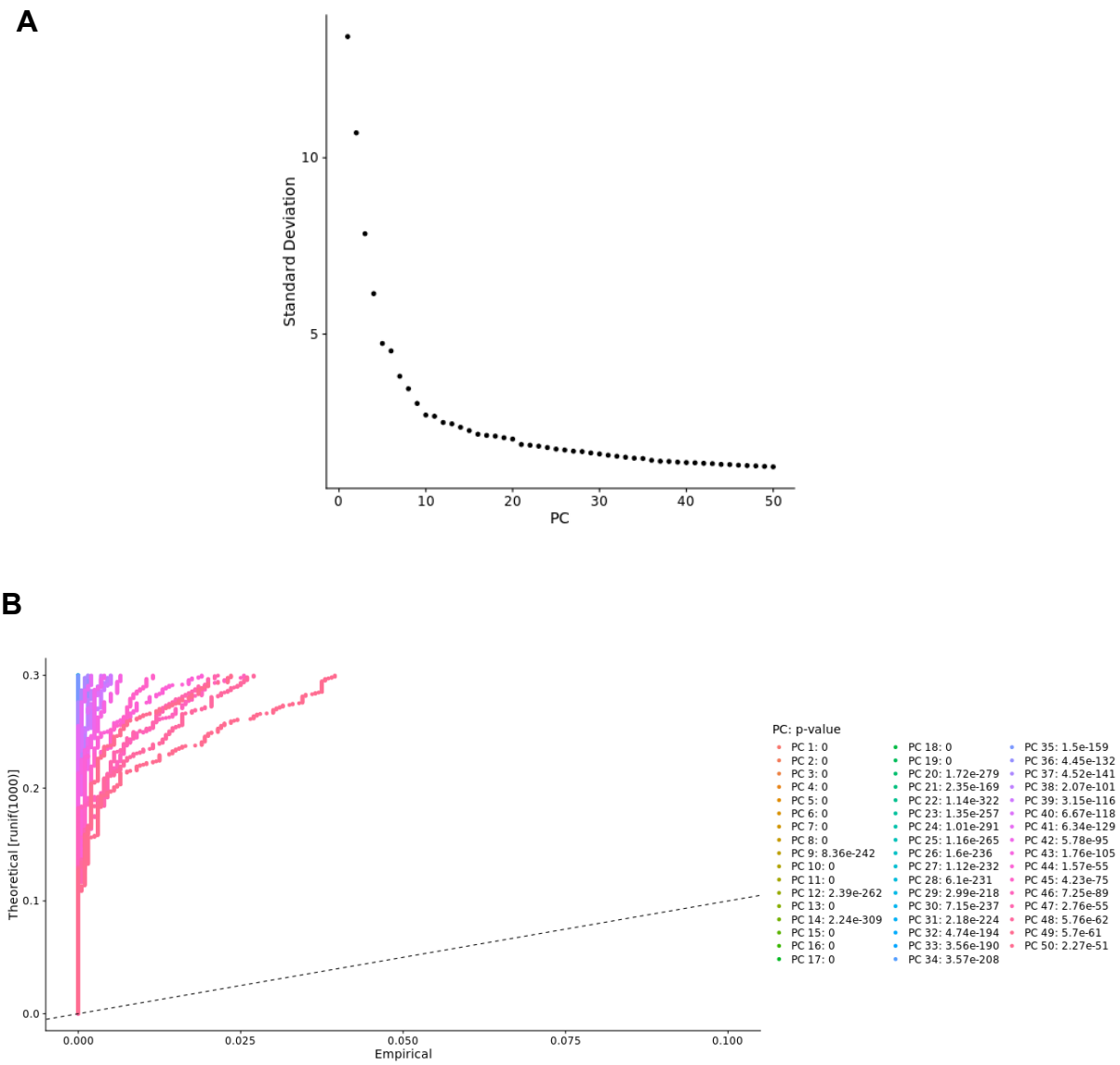


Figure 5.2 Estimation of data dimensionality. (**A**) PCA elbow plot showing percentage of variance explained by the top 50 PCs. (**B**) JackStraw plot comparing the distribution of p-values for each PC with a uniform distribution indicated by the dashed line. Significant PCs are those with a large enrichment of genes with significant p values <0.05. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

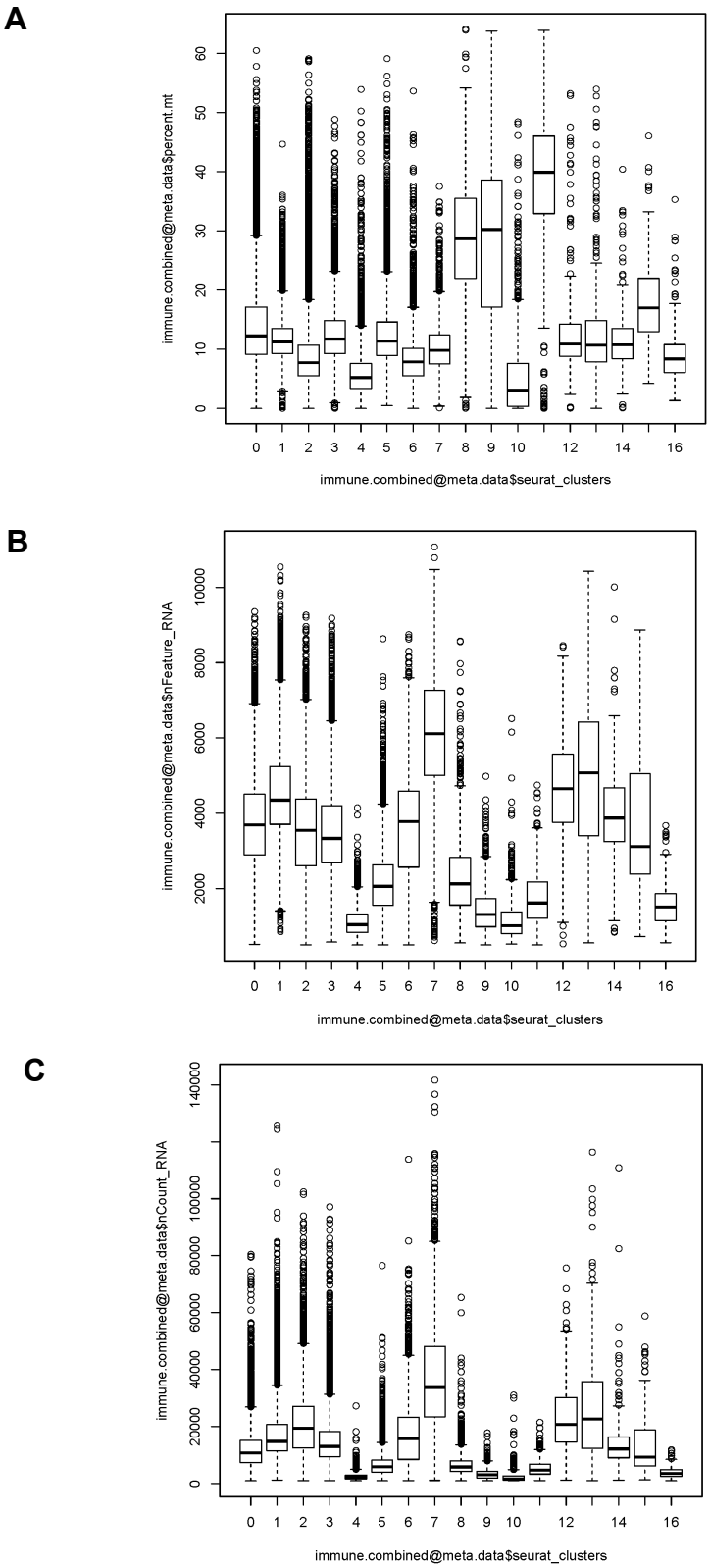


Figure 5.3 Quality control metrics for scRNA-seq of airway brushings by cluster. Box and whisker plots showing (**A**) *percent.mt* – percentage of reads mapping to mitochondrial genome, (**B**) *nFeature_RNA* – number of unique genes detected in each cell, (**C**) *nCount_RNA* - total read counts per cell for each cluster 0 – 16. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

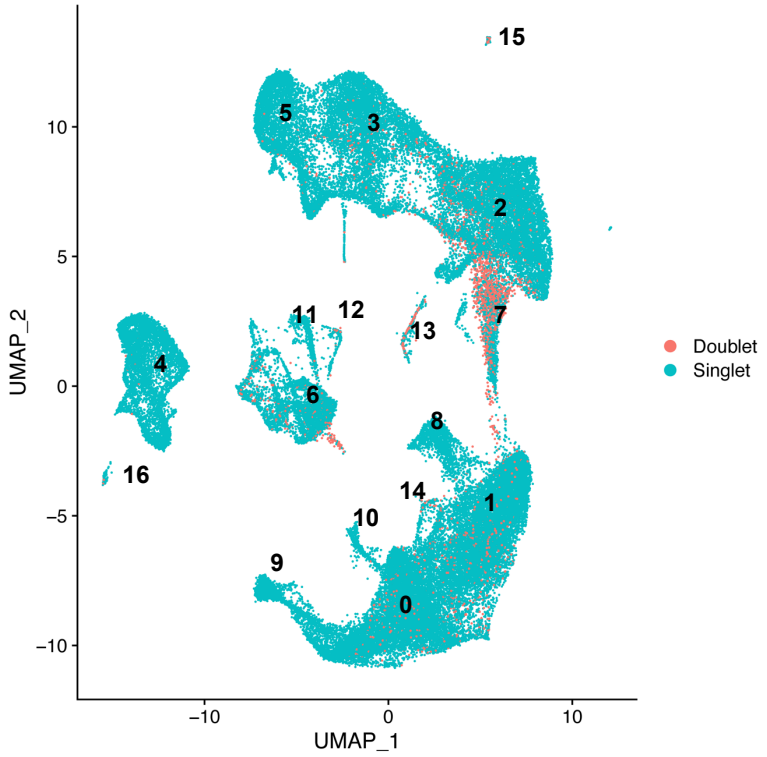


Figure 5.4 Integrated UMAP of single-cell data from all healthy and IPF subjects showing predicted doublets. Each dot represents a single-cell (turquoise) or doublets (pink). Position of seventeen cell clusters indicated (0 – 16). Accompanying table provides related quantitative data. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the DoubletFinder package in R v3.6.0.*

Table 5.3 Predicted singlets and doublets in the integrated dataset of all healthy and IPF subjects

Cluster	Total cells	Singlet	Doublet	% Doublets
0	11710	11508	202	1.7
1	10269	9971	298	2.9
2	9533	9216	317	3.3
3	6125	5993	132	2.2
4	5204	5192	12	0.2
5	4808	4777	31	0.6
6	4055	3763	292	7.2
7	2625	1326	1299	49.5
8	1834	1825	9	0.5
9	1337	1336	1	0.1
10	814	811	3	0.4
11	697	689	8	1.1
12	426	395	31	7.3
13	350	235	115	32.9
14	322	307	15	4.7
15	194	150	44	22.7
16	159	155	4	2.5

5.4.4 Cluster analysis of integrated dataset

The data was first analysed as an integrated object including all samples. Later in the Chapter, the data is analysed by condition (healthy compared to IPF) and by individual subject. The partitioning between clusters, and therefore the number of clusters obtained, was set using a resolution parameter within Seurat. A clustering tree[392] was used to evaluate the impact of a range of clustering resolutions (0.1 – 1.0) on the number of clusters obtained, and an optimal resolution of 0.2 was selected (Fig 5.5).

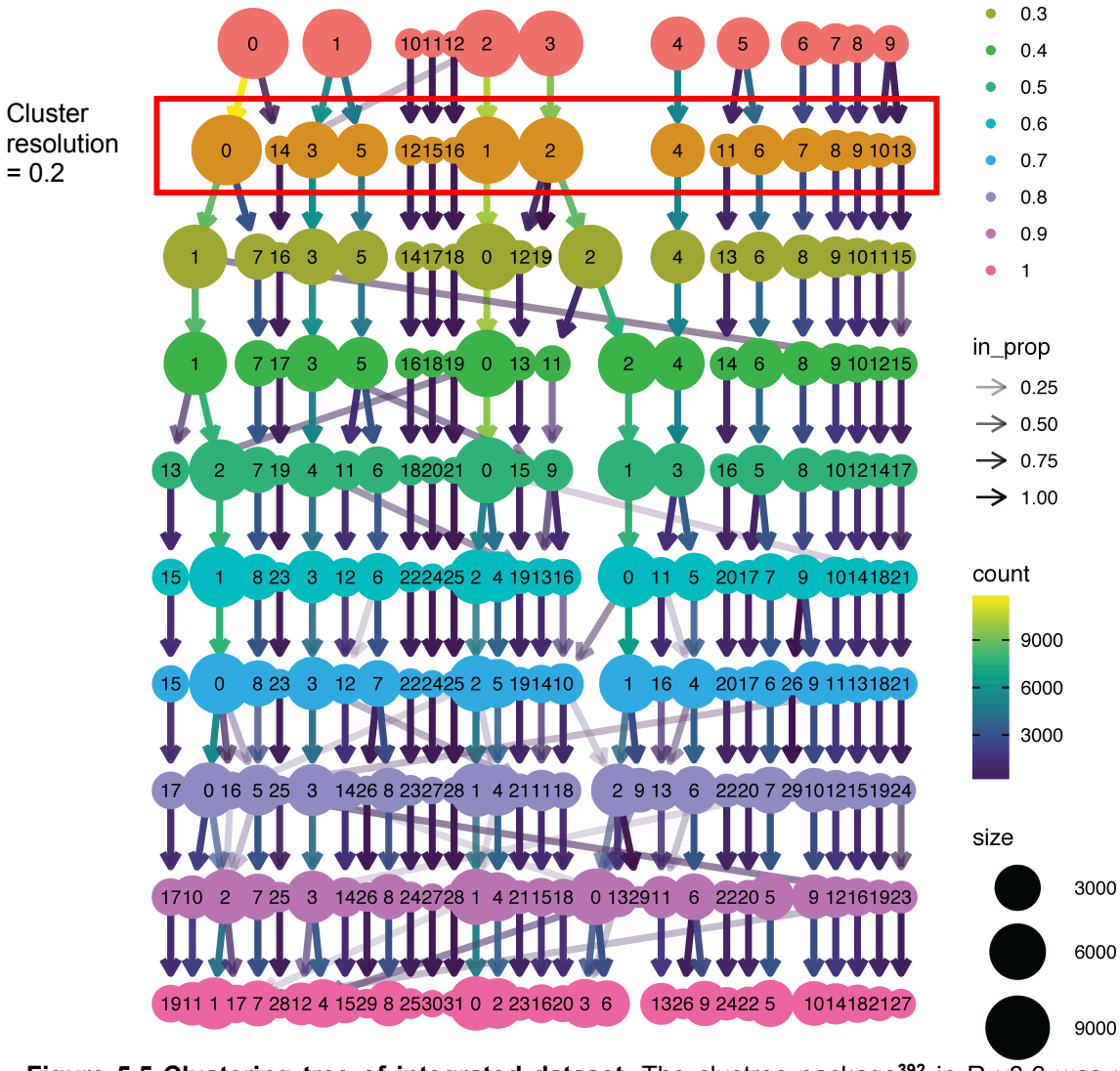


Figure 5.5 Clustering tree of integrated dataset. The clustree package[392] in R v3.6 was used to visualise changes to the clustering as the resolution parameter was increased. A higher resolution results in a greater number of clusters. This method of assessing the granularity of downstream clustering allowed a suitable resolution to be set. Each node represents a distinct cell cluster in the tree. Connecting arrows demonstrate how cells move between clusters as the resolution is increased from 0.1 (top line of tree) to 1.0 (bottom line of tree). The red box indicates cluster resolution (0.2) selected for our analysis, which generates 17 clusters (0 – 16). Plot generated by Probir Chakravarty (The Francis Crick Institute, UK).

Seventeen discrete clusters were revealed which was comparable to previously published scRNA-seq datasets of the human airway[168,179,180] (Fig. 5.6. – 5.7). Cluster labelling (0 to 16) was determined by cell number per cluster (Table 5.3). Individual clusters were annotated based on marker gene expression and published literature[163,168,179] The 12 epithelial clusters identified included all major airway cell types - basal, secretory and ciliated cells. Cluster 0 - ciliated cells, comprised the largest cell population (*n=11710* cells) in the airway brushings. Suprabasal cells were characterised by a lower expression of *KRT5* and higher expression of *NOTCH3, KRT4* and *KRT19* consistent with previous reports[179,204]. Secretory cells (clusters 2 and 7) were defined by expression of canonical secretory genes such as *SCGB1A1, MUC5AC* and *MUC5B*[198]. Interestingly these clusters were enriched in genes linked to biological processes identified using GO of detoxification and immune cell chemotaxis. Six ciliated cell clusters (clusters 0,1,8,9,10 and 14) were identified by characteristic marker gene expression including *SNTN* and *AGR3*, and enrichment of genes in the GO biological process categories of cilium organisation (GO:0044782) and cilium assembly (GO: 0060271) (FDR p-value <0.05). Cluster 14 was notable for enrichment in genes involved in the cellular response to type 1 interferon (IFN) including *IFIT1, ISG15* and *MX1.* Cluster 13 expressed genes involved in centriole assembly and specific markers *CDC20B*, *DEUP1* and *FOXN4* which are unique to precursors of ciliated cells termed deuterosomal cells[179,204,393]. It was possible to identify a total of 194 ionocytes (cluster 15) - a rare and recently discovered epithelial cell type – based on expression of *FOXI1, ASCL3* and *CFTR*[200].

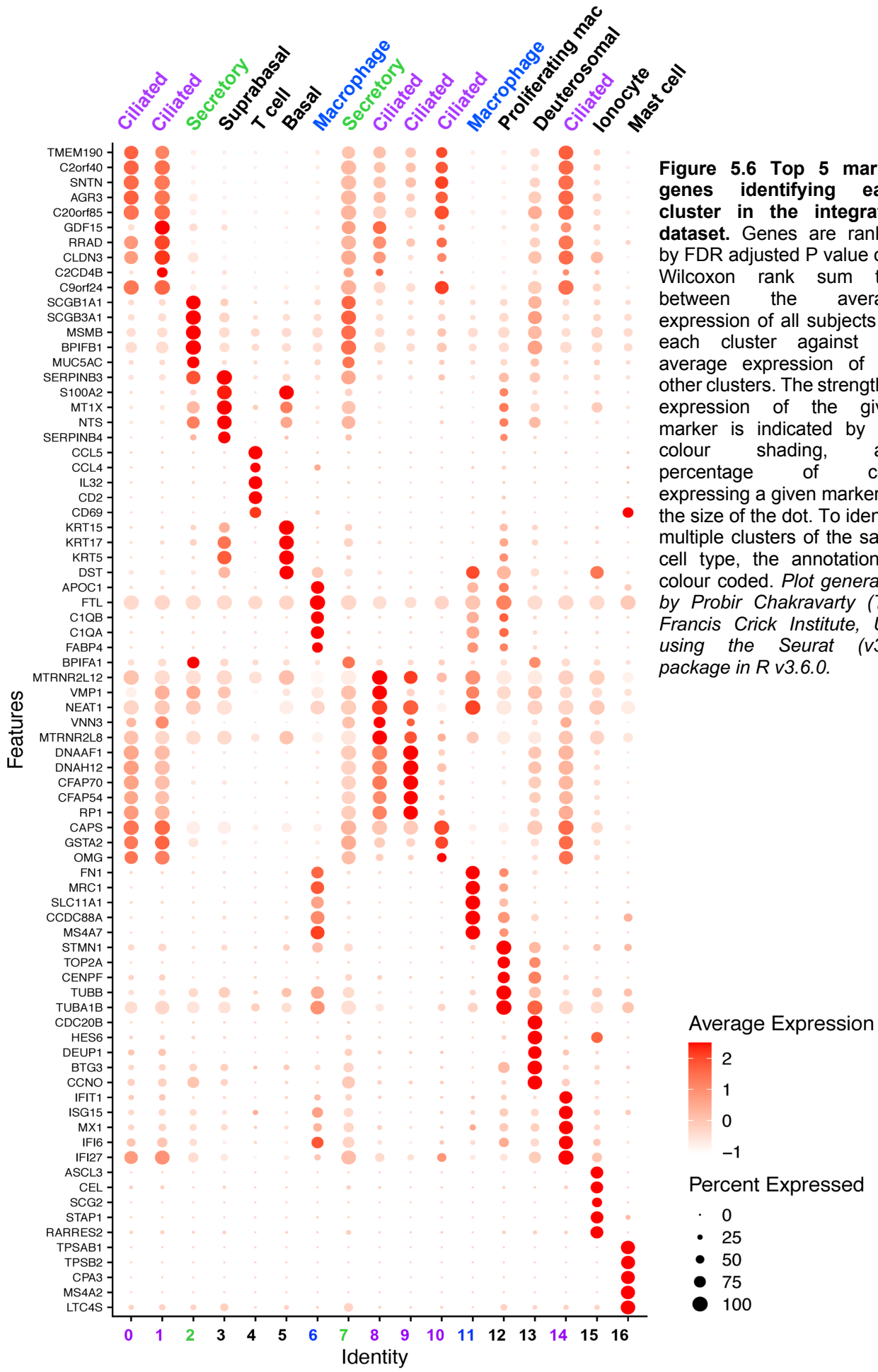
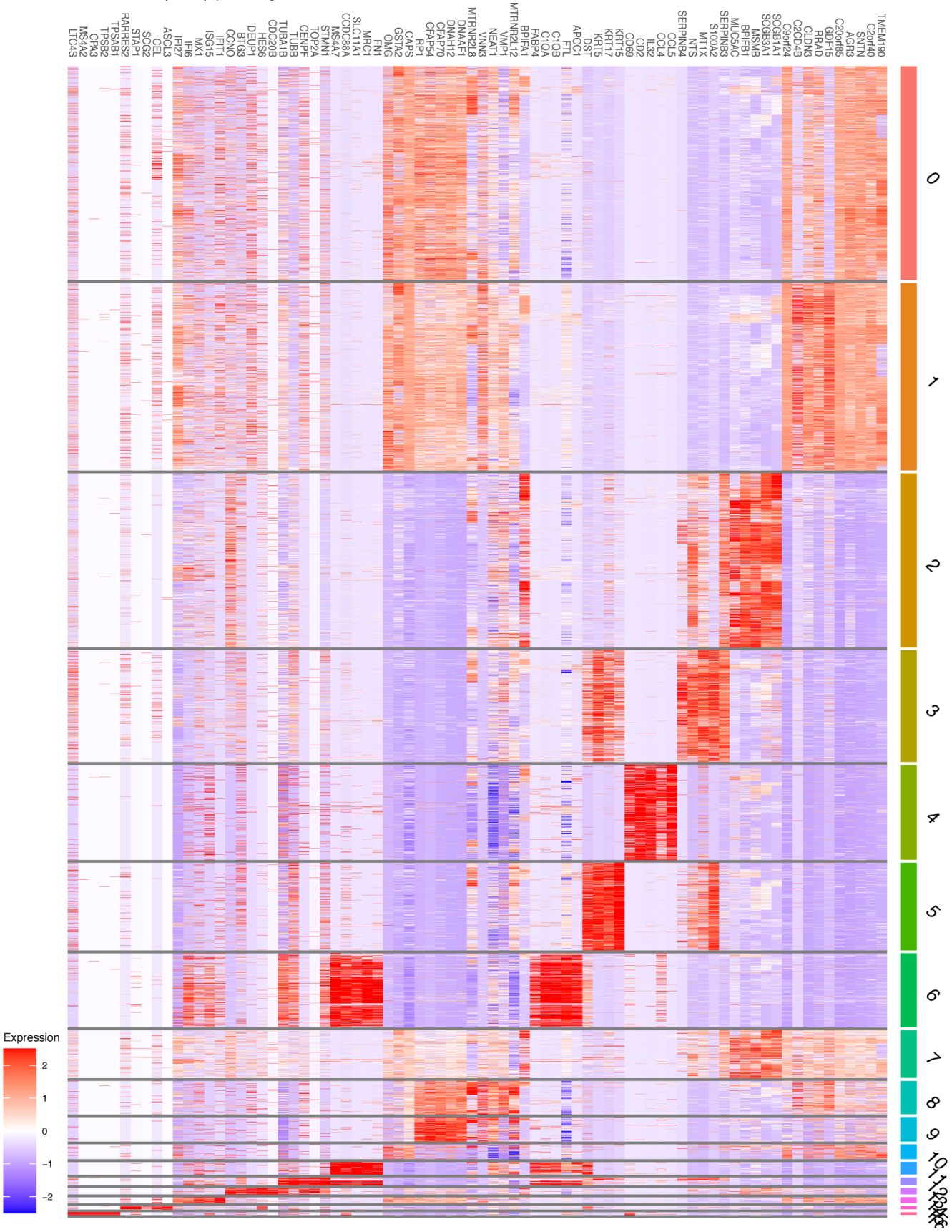


Figure 5.6 Top 5 marker genes identifying each cluster in the integrated dataset. Genes are ranked by FDR adjusted P value of a Wilcoxon rank sum test between the average expression of all subjects for each cluster against the average expression of the other clusters. The strength of expression of the given marker is indicated by red colour shading, and percentage of cells expressing a given marker by the size of the dot. To identify multiple clusters of the same cell type, the annotation is colour coded. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

Figure 5.7 Heatmap of top 5 marker gene expression for each cell in a cluster for the integrated dataset. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*



Manual cluster identification was validated by an automated annotation tool, scCATCH[254] (Table 5.4), and Gene Set Enrichment Analysis (GSEA) using Deprez et al.[179] and Reyfman et al.[163] marker gene lists (Fig. 5.8). scCATCH assigned an annotation of '*SLC16A7*+ cells' to clusters 8 and 9, suggesting that genes associated with this cell type were enriched in these clusters. In agreement with Plasschaert et al.[201], these were also clusters with a higher percentage of mitochondrial reads (Fig. 5.3A). Ciliary motility- related genes including *DNAH12*, *DNAH11* and *CFAP70* were abundantly expressed in these clusters, and thus the higher expression of mitochondrial genes is consistent with the energy requirements for this process[198].

Five immune cell clusters were identified including 3 populations of macrophages, and T cell and mast cell clusters based on unique marker gene expression in conjunction with information from the published literature[163,168,179]. T cells (cluster 4) were enriched in expression of *CCL5, IL32, CD2, TRAC* and genes encoding CD8 and CD3. Tissue-residency markers *CD69* and *ITGA1* were highly expressed in this cluster[394]. Genes associated with cytotoxic CD8+ T cells and NK cells such as *GZMA* and *GZMB* encoding granzymes, *KLRC1* and *KLRB1* encoding transmembrane receptors were also expressed. Mast cells (cluster 16) were the rarest cell population in the airway brushings (*n*=159 cells) and demonstrated unique expression of *TPSAB1, CPA3* and *MS4A2*. Other granulocytes including neutrophils and eosinophils were not identified in the airway brushings and obtaining these cell types using the 10X Genomics single-cell platform is a recognised challenge[395].

Marker genes defining mature macrophages in cluster 6 were associated with lipid transport and metabolism (*APOC1, FABP4)*. Cluster 11 monocyte-derived macrophages were enriched in genes associated with cell migration *(FN1, CCDC88A)* and mature cellular function in the monocyte lineage (*MS4A7)*, and cluster 8 proliferating macrophages expressed cell cycle genes (*STMN1, TOP2A).* The three clusters were not clearly distinguished into myeloid subsets using GSEA as they were enriched for marker genes of both macrophages and monocytes described in the Reyfman et al.[163] and Deprez et al.[179] dataset (Fig. 5.8). Single-cell studies of healthy and IPF distal lung tissue and BAL have previously reported several subsets of macrophages including a profibrotic *SPP1* -expressing macrophage subtype[53,163,164].

The three macrophage clusters identified in our dataset were not segregated by the subset-defining gene markers from the Morse et al.[164] study; *FABP4*, *INHBA* and *LGMN* showed highest expression in all three clusters, perhaps due to the different location or the different stage of disease under study (Fig. 5.9). Macrophages in cluster 6 were highest expressors of *SPP1.*

Table 5.4 Cluster cell type identity

Cluster	Annotation using published marker genes	Annotation using scCATCH	scCATCH celltype_score
0	Ciliated		
1	Ciliated	Ciliated Cell, FOXN4+ Cell	0.5, 0.5
2	Secretory	Club cell	0.8
3	Suprabasal	Basal Cell, Secretory Cell	0.67, 0.67
4	T cell	T Cell	0.77
5	Basal cell	Basal Cell	0.65
6	Macrophage	M2 Macrophage	0.95
7	Secretory	n/a	n/a
8	Ciliated	SLC16A7+ Cell	0.61
9	Ciliated	SLC16A7+ Cell	0.7
10	Ciliated	n/a	n/a
11	Macrophage	Macrophage	0.8
12	Proliferating macrophage	SLC16A7+ Cell	0.63
13	Deutrosomal	FOXN4+ Cell	0.7
14	Ciliated	Ciliated Cell, SLC16A7+ Cell	0.58, 0.58
15	Ionocyte	Ionocyte Cell	0.7
16	Mast cell	SLC16A7+ Cell	0.66

Where more than one cluster exists for a given cell type, these are colour coded.
n/a – automated annotation not possible

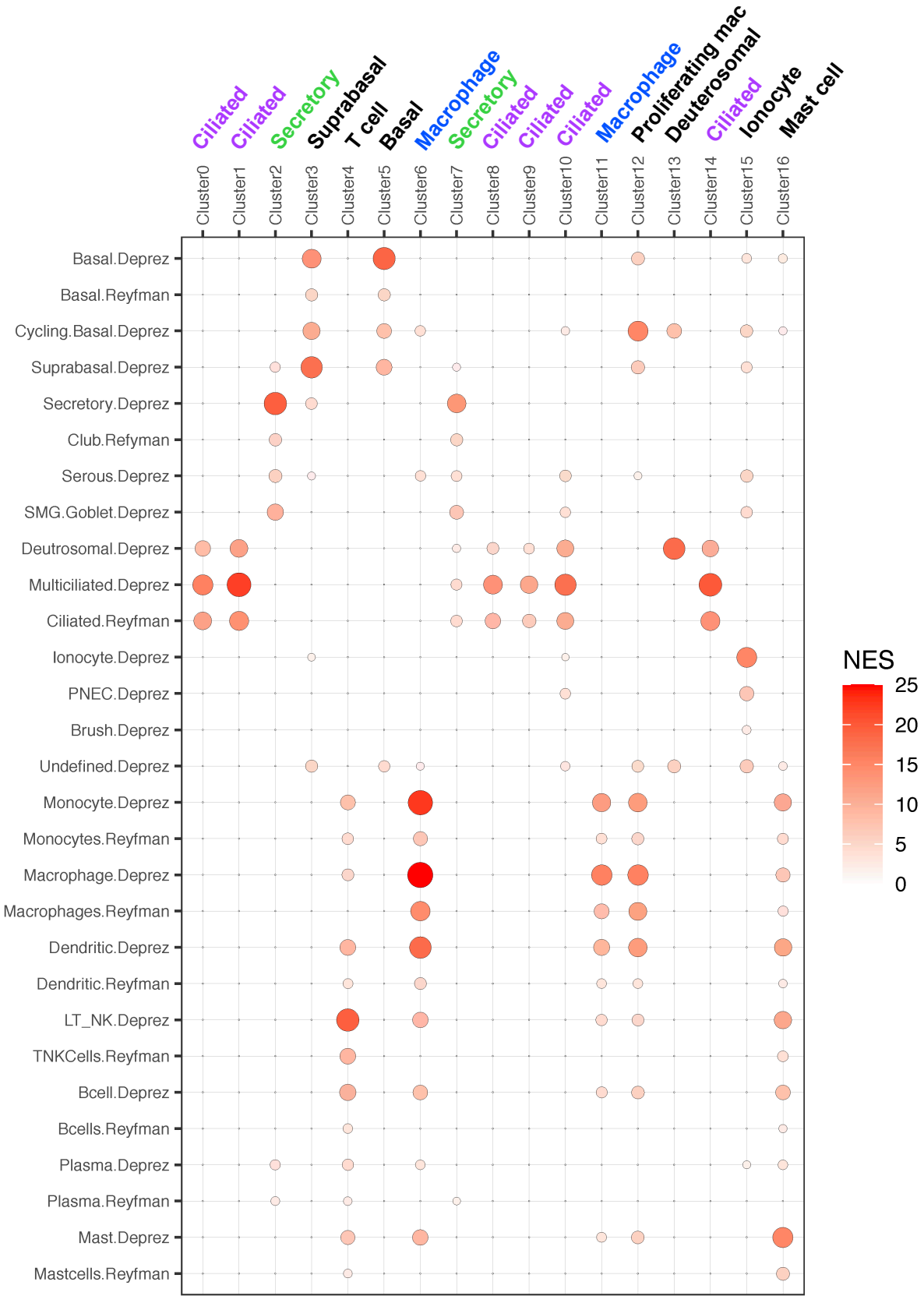


Figure 5.8 Gene set enrichment analysis (GSEA) using cell type specific marker genes from published datasets. GSEA performed using marker gene lists from published single-cell datasets Reyfman et al. 2018 and Deprez et al. 2020. Dataset indicated after cell type on y axis. NES – normalised enrichment score.

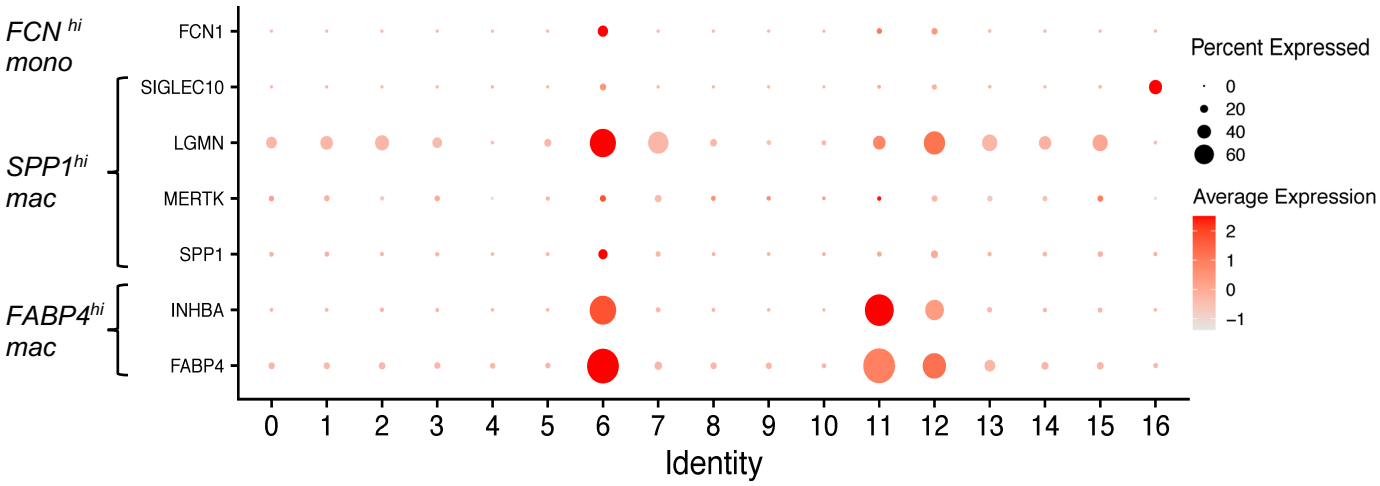


Figure 5.9 Dot plot of macrophage/ monocyte marker genes from Morse et al. dataset.

To visualise and explore the dataset, the Uniform Manifold Approximation and Projection (UMAP)[396] dimensionality reduction technique was used (Fig. 5.10). Compared with other dimensionality reduction methods, UMAP performs better at preserving more of the global structure and relationships from high-dimensional data[396]. A continuity between epithelial cell clusters representing different cellular phenotypes during differentiation (basal cells to secretory to ciliated cells) could be clearly seen in the global structure of our scRNA-seq data (Fig. 5.10).

To validate cluster annotations, GO enrichment analysis using the online PANTHER tool was performed (Fig. 5.11 and 5.12). Ordering significant GO biological processes (FDR <0.05) by fold enrichment (the number of genes observed in the uploaded gene list over the expected number) provided more informative results than ordering simply on the statistical significance of observing genes from a particular category. GO analysis of macrophage clusters (Fig. 5.12) revealed that: antigen processing and presentation via MHC Class II was the top enriched biological process in all 3 clusters; genes linked to wound healing were enriched in cluster 6; regulation of cell adhesion to endothelium enriched in cluster 11; and cell division enriched in cluster 12. Overall, there appeared greater similarity between clusters 6 and 11, as compared to cluster 12.

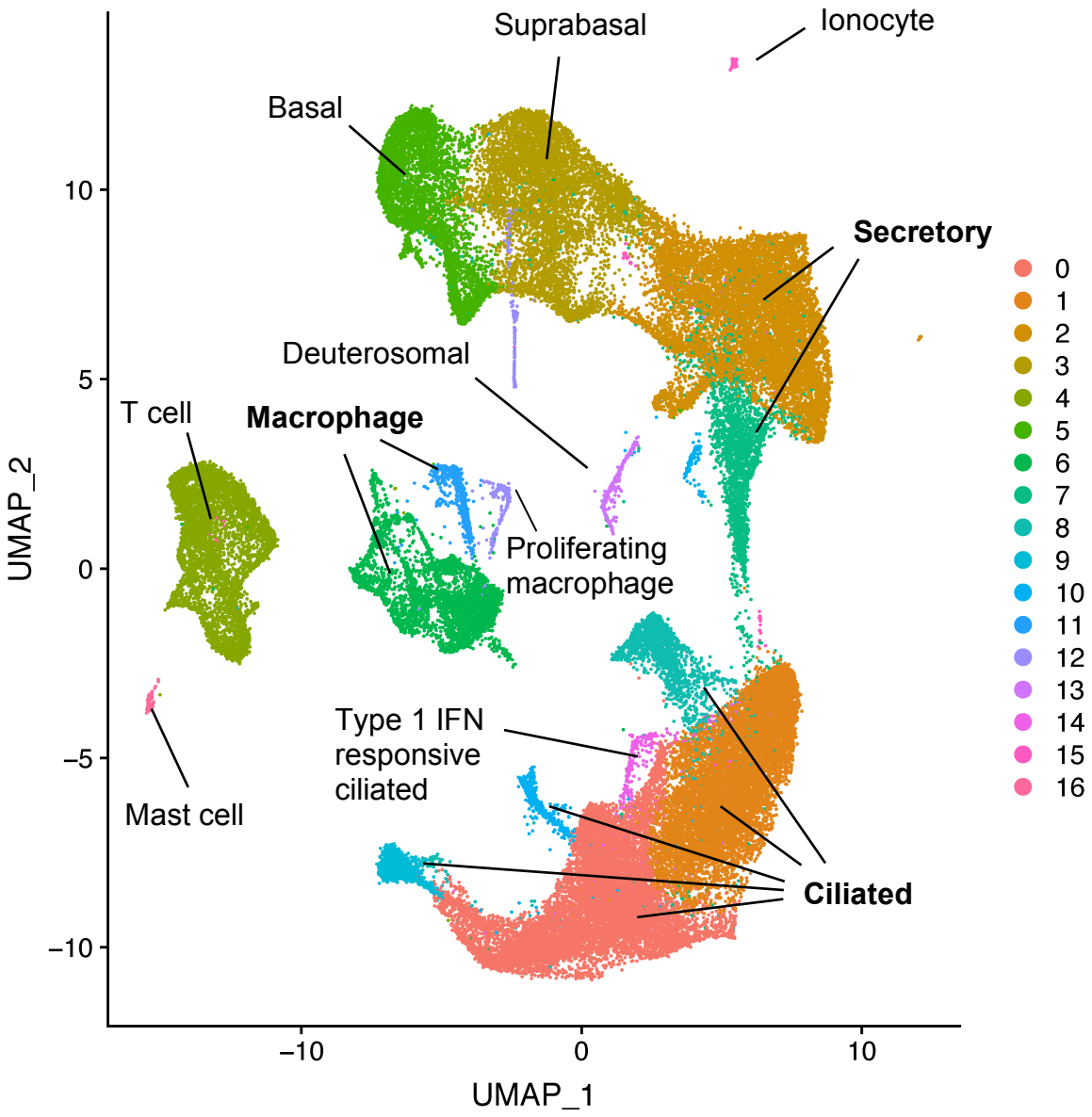


Figure 5.10 Integrated UMAP visualisation of airway cells from all subjects. Single-cell transcriptomes from 60462 cells from 7 healthy and 9 IPF subjects were profiled, integrated and clustered. Seventeen distinct clusters were identified based on unique marker gene expression (genes present in minimum of 10% cells in a cluster, log2 fold change >0.25, adjusted P value <0.05). Each dot represents a single cell, coloured according to cluster cell type. Cell types were assigned based on marker genes from published literature and checked using an automated annotation tool, scCATCH. Where more than one cluster exists for a given cell type, these annotations are in bold. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

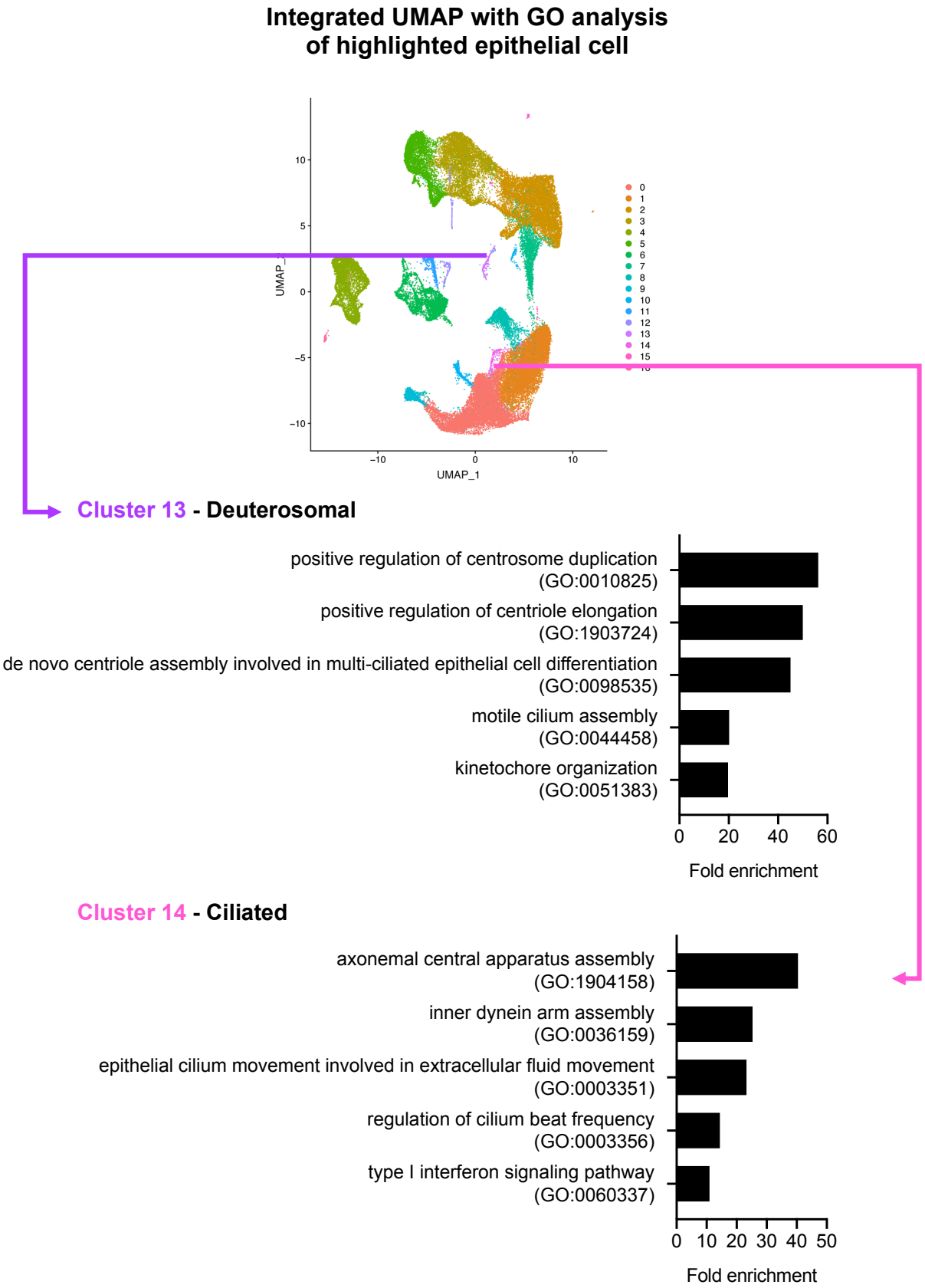


Figure 5.11 Gene Ontology (GO) enrichment analysis for selected epithelial cell clusters. Gene Ontology (GO) enrichment analysis (PANTHER) using gene list generated by *FindMarkers* function in Seurat based on a Wilcoxon rank sum test comparing average expression in cells belonging to one cluster to all cells within the experiment. Only genes showing >0.25 log2 fold change with a Bonferroni adjusted p-value <0.05 were used for GO analysis. Representative enriched GO biological processes shown. Results displayed were analysed with a Fisher's Exact test, FDR P <0.05.

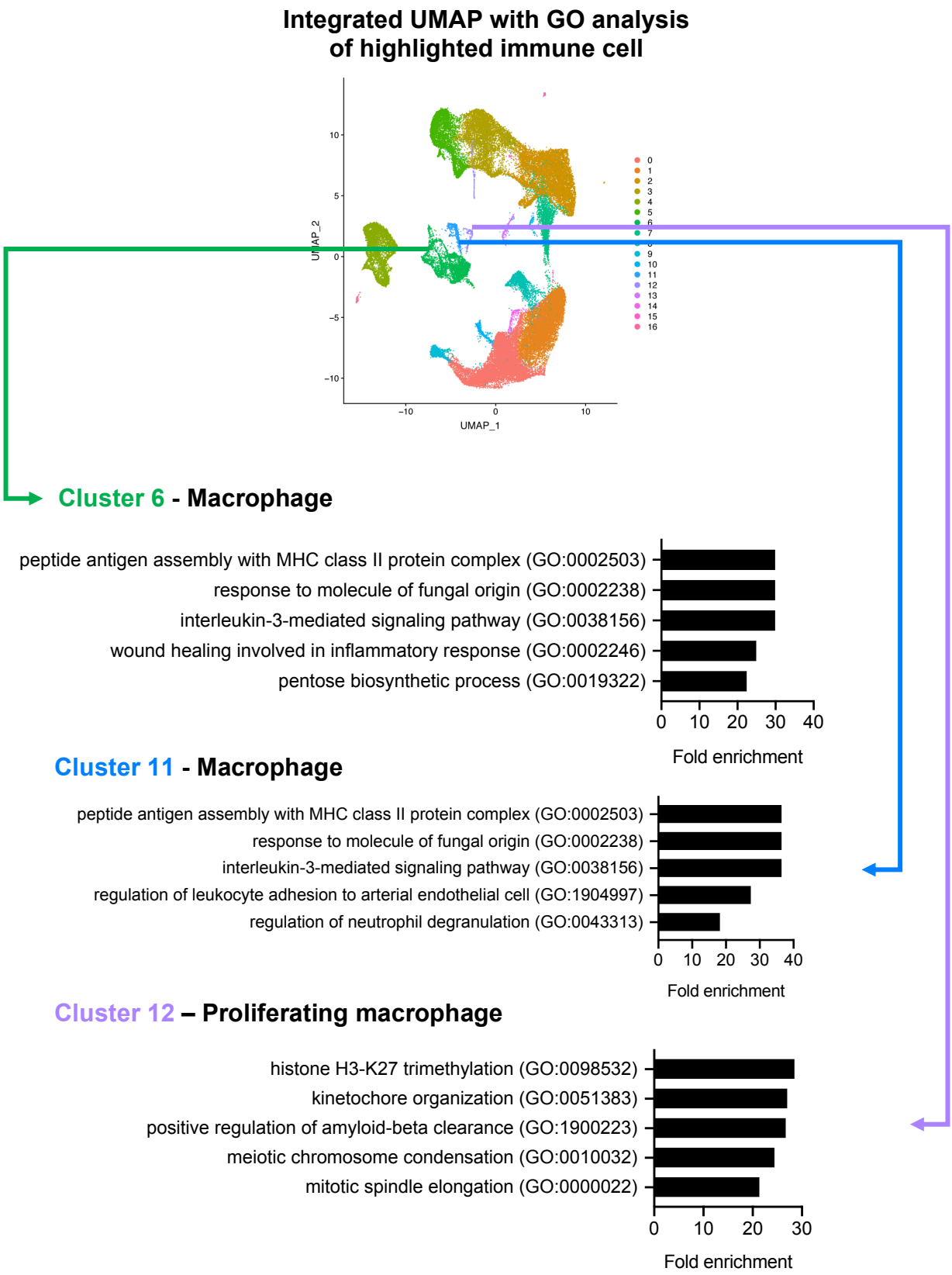


Figure 5.12 Gene Ontology (GO) enrichment analysis for macrophage clusters. Gene Ontology (GO) enrichment analysis (PANTHER) using gene list generated by *FindMarkers* function in Seurat based on a Wilcoxon rank sum test comparing average expression in cells belonging to one cluster to all cells within the experiment. Only genes showing >0.25 log2 fold change with a Bonferroni adjusted p-value <0.05 were used for GO analysis. Cluster 6 *n* = 686 genes, cluster 11 *n* = 561 genes and cluster 12 *n* = 476 genes. Key enriched GO biological processes shown. Results displayed were analysed with a Fisher's Exact test, FDR P <0.05.

5.4.5 Cluster analysis by condition

To investigate differences in the airway cell subsets identified by clustering the integrated dataset, separate UMAP representations were created for IPF and healthy subjects (Fig. 5.13).

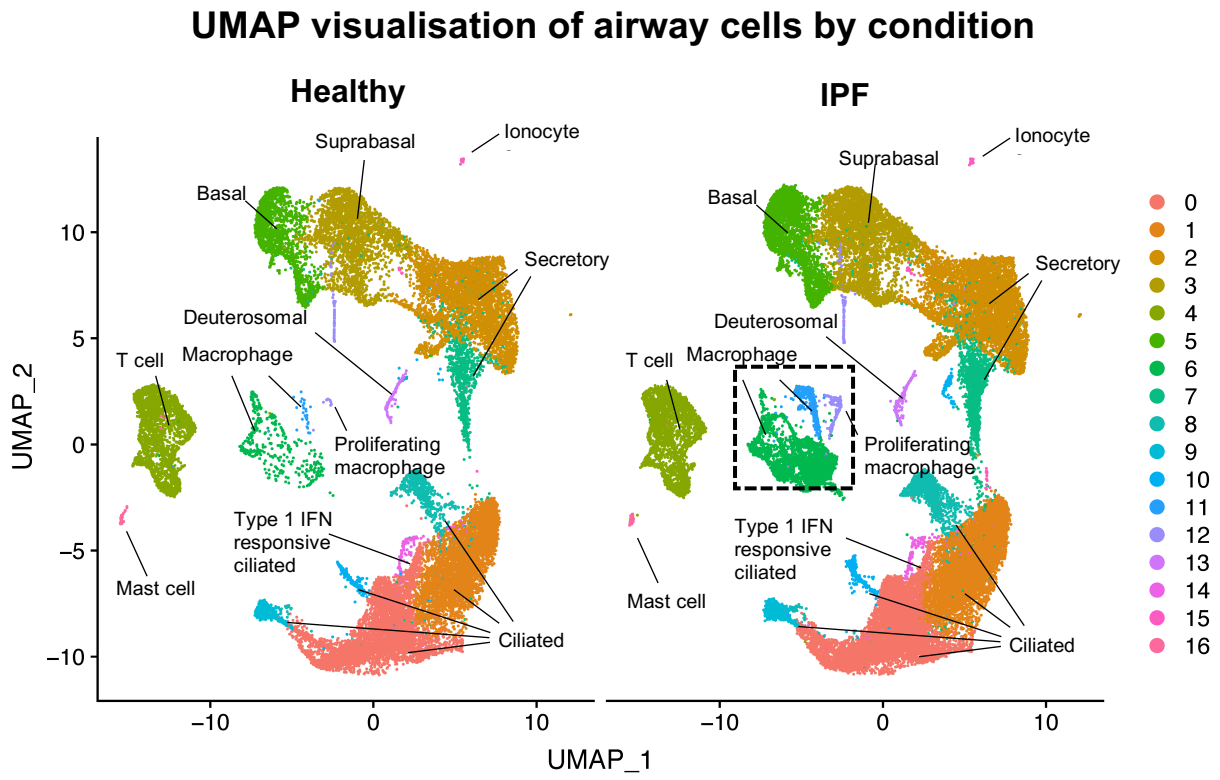


Figure 5.13 UMAP visualisation of airway cells by condition. UMAP representations showing 17 distinct clusters according to condition. Single-cell transcriptomes from 22559 cells from 7 healthy controls *(left),* and 37903 cells from 9 IPF patients (*right*) were used for clustering. Each dot represents a single cell, coloured according to cluster cell type. A dashed box highlights dramatic expansion in macrophage populations in IPF. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

In healthy airway brushings, T cells were the most populous immune cell type comprising 10.8% of all cells, yet in the IPF airway there was marked expansion in macrophage populations and cluster 6 was the most prominent immune population (Fig. 5.14).

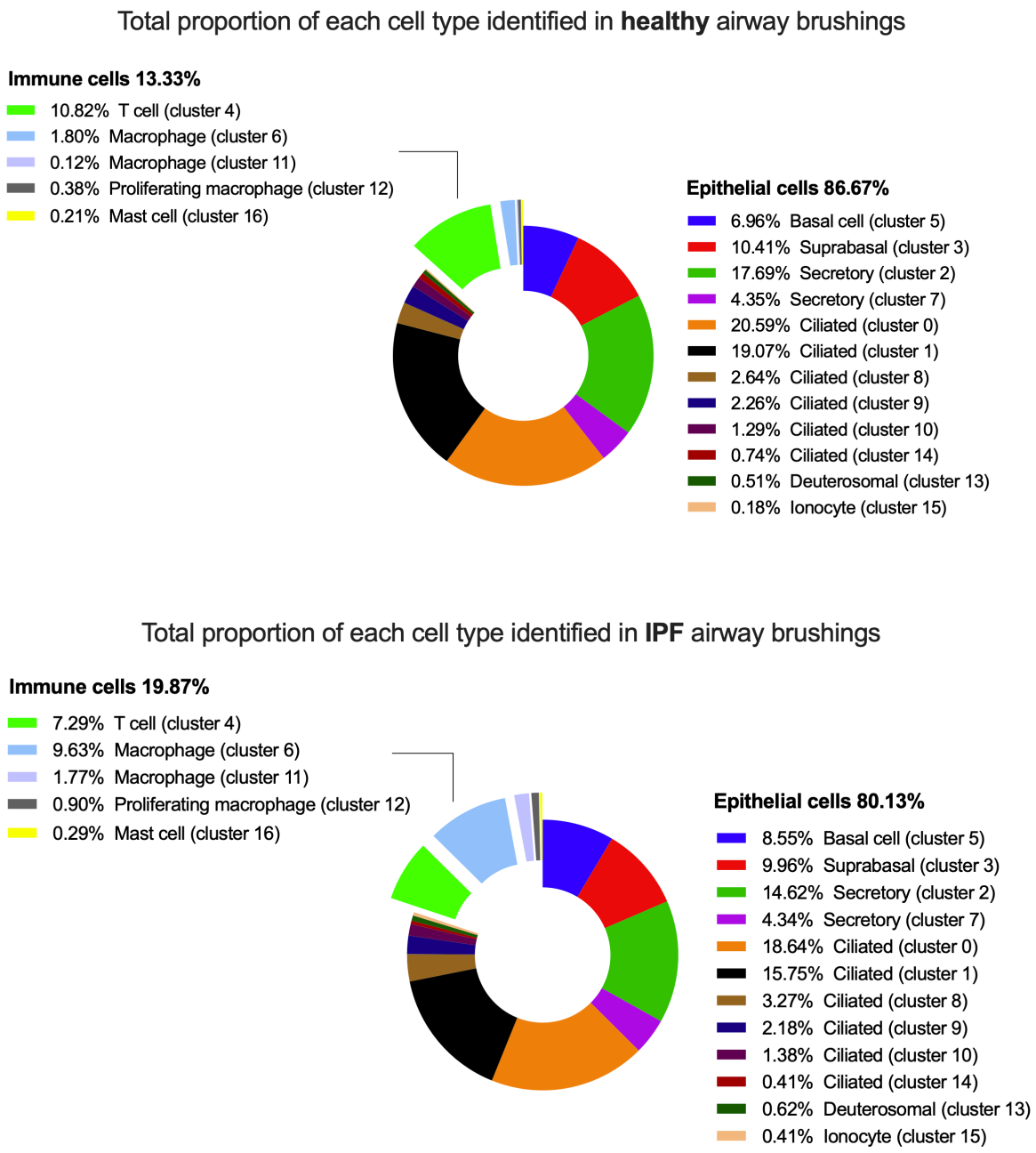


Figure 5.14 Total proportion of each cell type identified in airway brushings by condition. Cell type as a percentage of all cells recovered on airway brushings from (A) healthy controls (n = 7) and (B) IPF patients (n = 9).

To ascertain whether this difference was driven by disease or the contribution of individual donors, UMAP plots per subject were generated (Fig. 5.15). This demonstrated a consistent increase in the number of cells in macrophage clusters 6, 11 and 12 in IPF subjects compared to control.

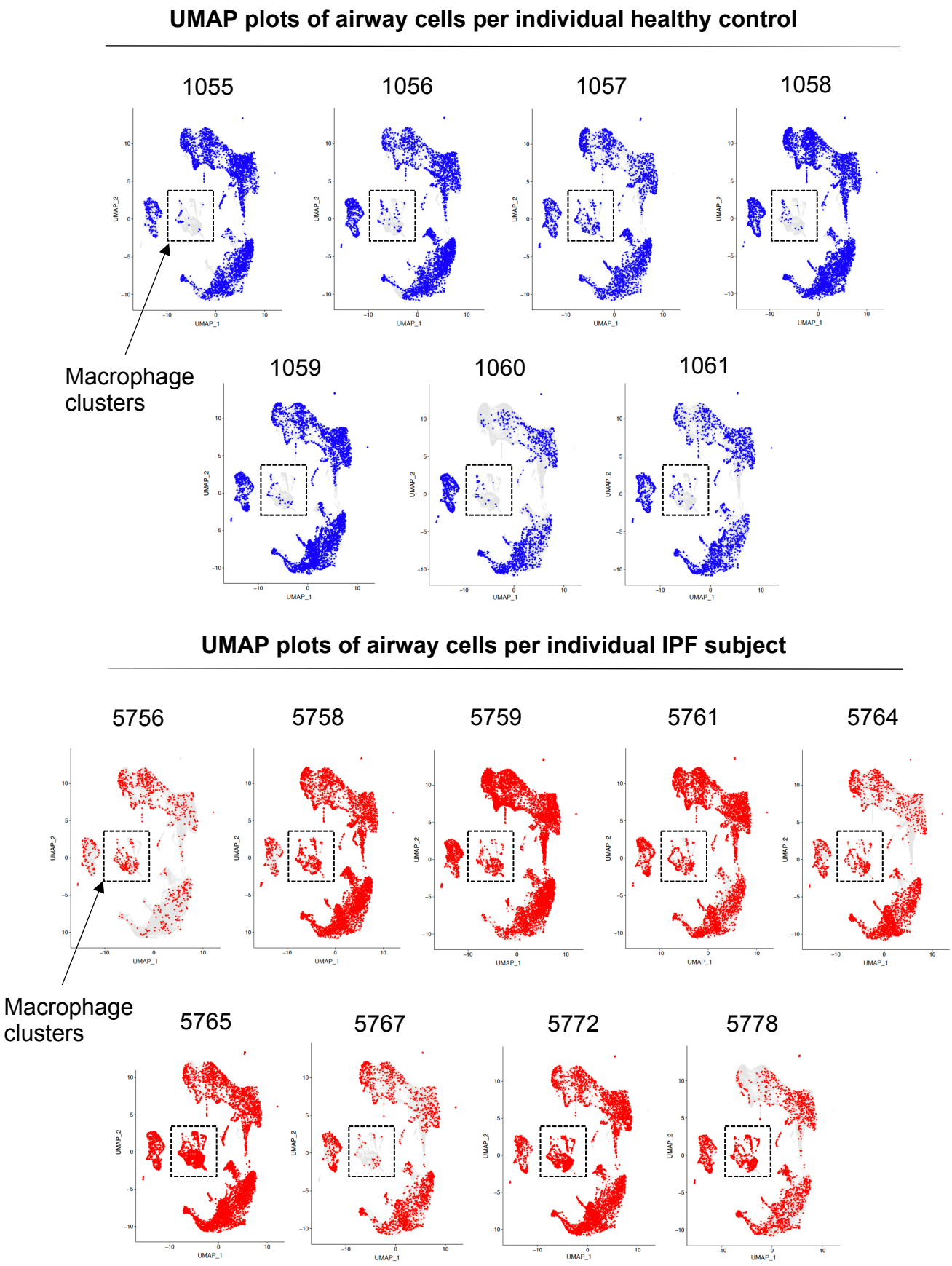


Figure 5.15 UMAP visualisation of airway cells per individual subject. UMAP representations showing contribution to clusters by individual subjects; IPF (red) or healthy control (blue). A dashed box highlights consistent expansion in macrophage populations in IPF subjects. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

To validate these observations quantitatively, the proportion of epithelial and immune cells in each cluster contributed by each donor was plotted (Fig. 5.16 and Fig. 5.17). There were no statistically significant differences in epithelial cluster proportions between IPF and healthy controls, however there was an increase in proportion of basal cells (cluster 5) and ionocytes (cluster 15) in some of the IPF subjects relative to controls. There was a significant decrease in the percentage of T cells (cluster 4) and corresponding increase in macrophages in cluster's 6 and 11. It should be noted that this data was represented as the percentage of all immune cells per donor, therefore rather than a genuine decrease in the number of T cells, it is likely due to an expansion in macrophage clusters. Interestingly there was a significant positive correlation between the percentage makeup of proliferating macrophages (cluster 12) and composite physiological index (CPI), a validated marker of IPF disease extent and prognostic indicator[19] (Fig. 5.18).

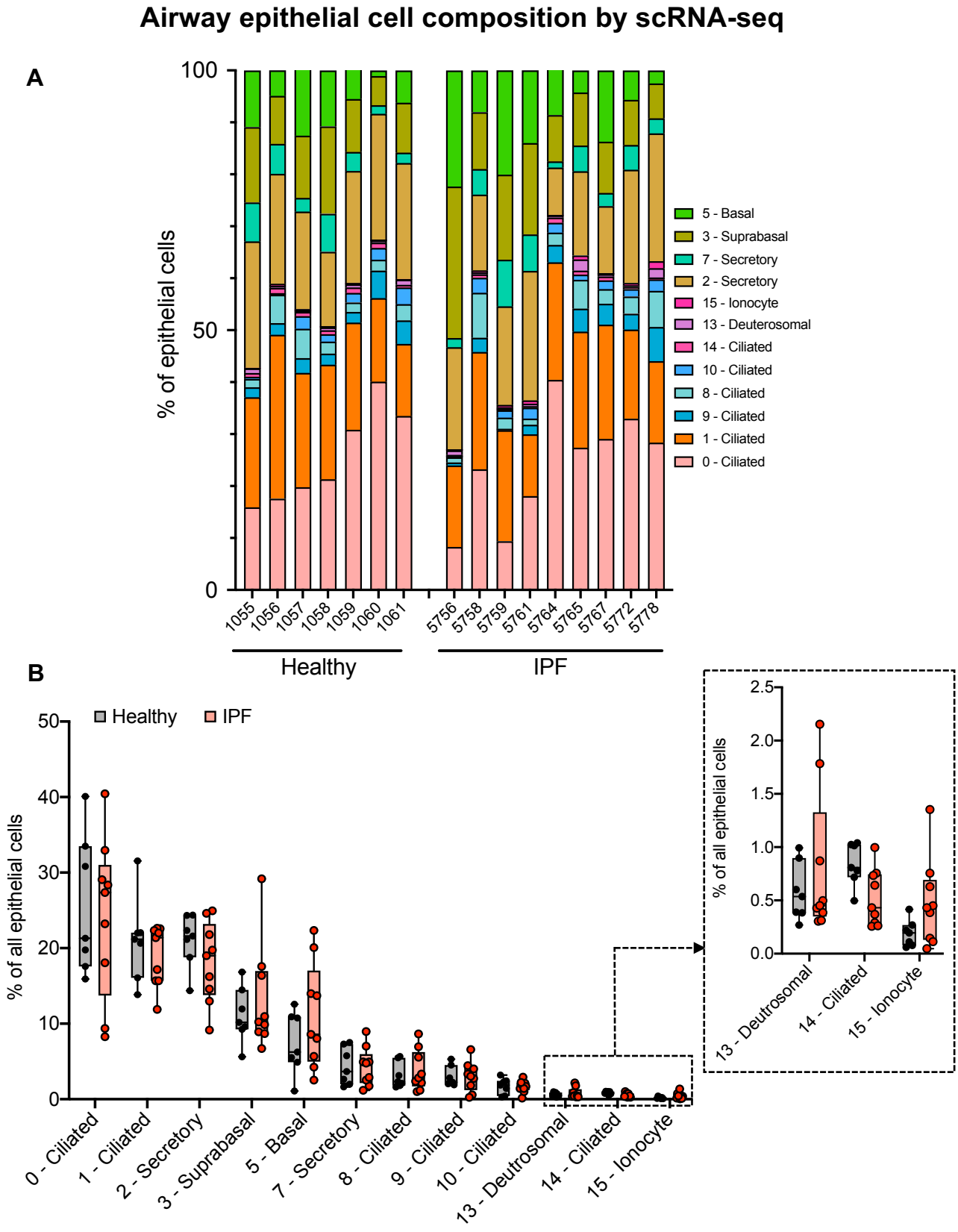


Figure 5.16 Quantification of epithelial cell types in IPF and healthy airways per subject. (**A**) Stacked bar chart quantifying percentage of each epithelial cell cluster per subject. (**B**) Boxplot representing the percentage of each epithelial cell cluster per subject comparing IPF (*n*=9) and healthy controls (*n*=7). Each dot represents a single subject, whiskers showing min. to max values. Clusters highlighted in dashed box have been enlarged for clarity.

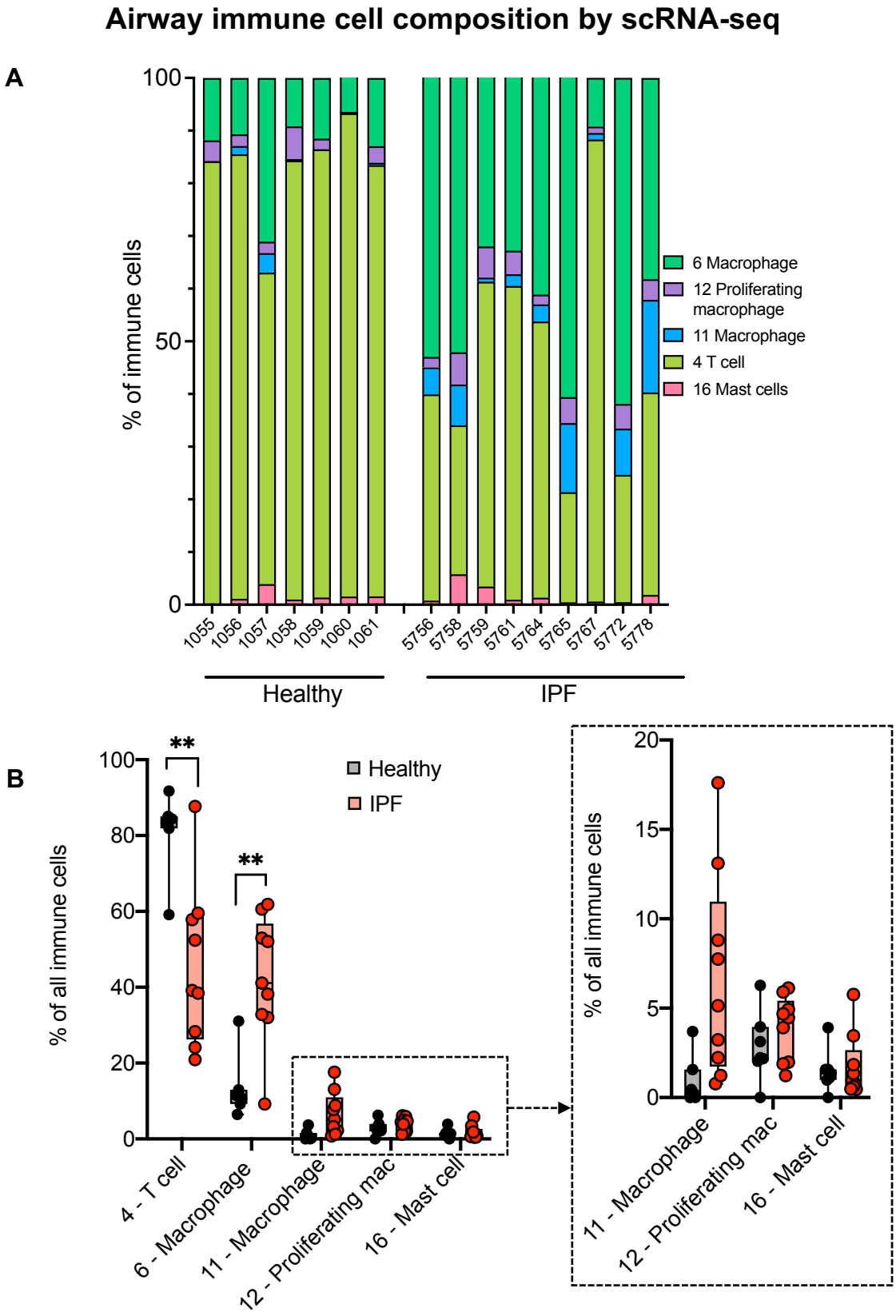


Figure 5.17 Quantification of immune cell types in IPF and healthy airways per subject. (**A**) Stacked bar chart quantifying percentage of each immune cell cluster per subject. (**B**) Boxplot representing the percentage of each immune cell cluster per subject comparing IPF (*n*=9) and healthy controls (*n*=7). Each dot represents a single subject, whiskers showing min. to max values. Clusters highlighted in dashed box have been enlarged for clarity. Unpaired (two-sample) t-test **Holm-Sidak adjusted P value <0.01

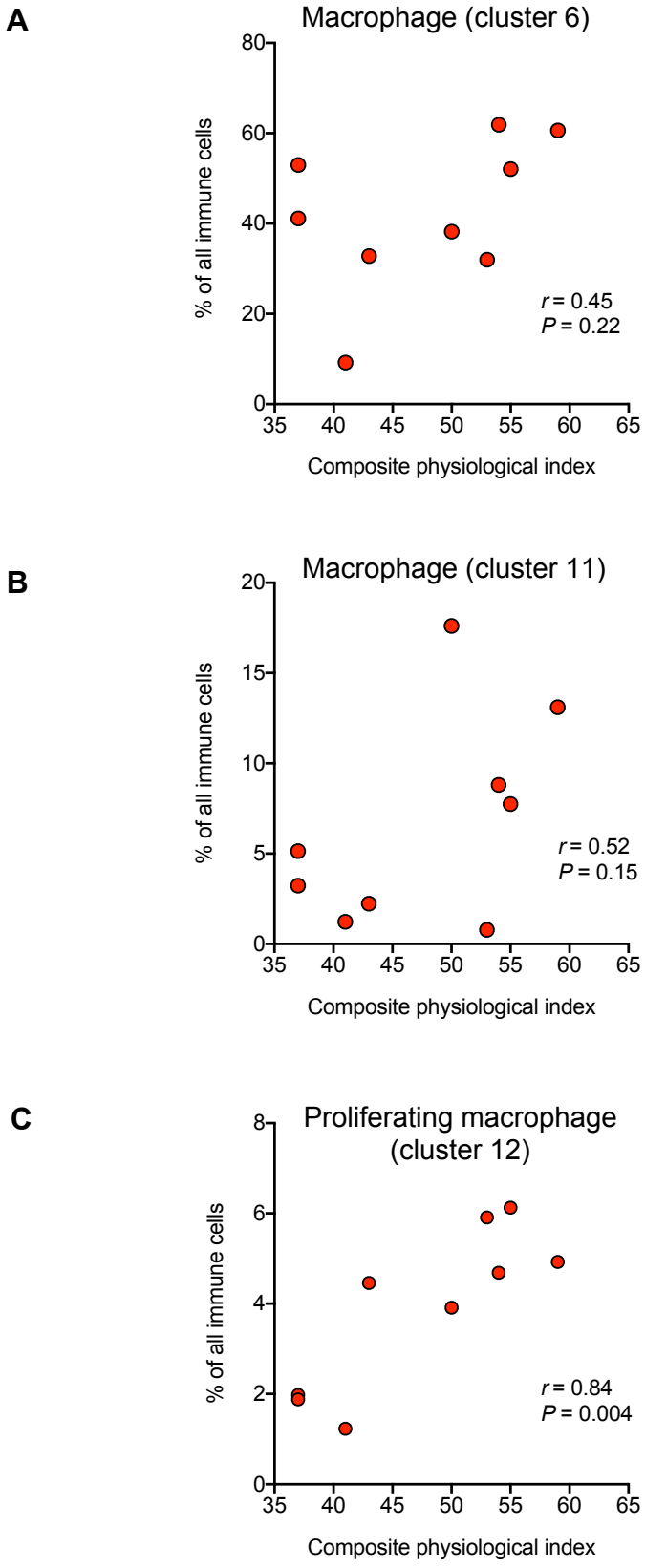


Figure 5.18 Correlation analysis between macrophage cluster percentage and the extent of disease (composite physiological index). Pearson correlation analysis between percentage of all immune cells of (**A**) macrophages in cluster 6, (**B**) macrophages in cluster 11, (**C**) proliferating macrophages in cluster 12. Correlation coefficient, *r*, and two-tailed *P-value* noted on plot.

In Chapter 5, bulk RNA-seq revealed that *KRT13,* a gene marker of recently described 'Hillock' cells[200], was significantly upregulated in IPF airway brushings compared to healthy controls. A rare population of KRT13 expressing cells was also identified on immunostained IPF airway brush cytospins and tissue sections. It was not possible to identify a distinct population of *KRT13* expressing cells in our single-cell dataset using the FeaturePlot function in Seurat which visualises expression on the integrated UMAP (Fig. 5.19). There were small numbers of *KRT13* expressing cells in the suprabasal cell cluster 3 in healthy controls and IPF. Expression of corresponding hillock cell marker *KRT4* was also concentrated in the suprabasal cell (cluster 3).

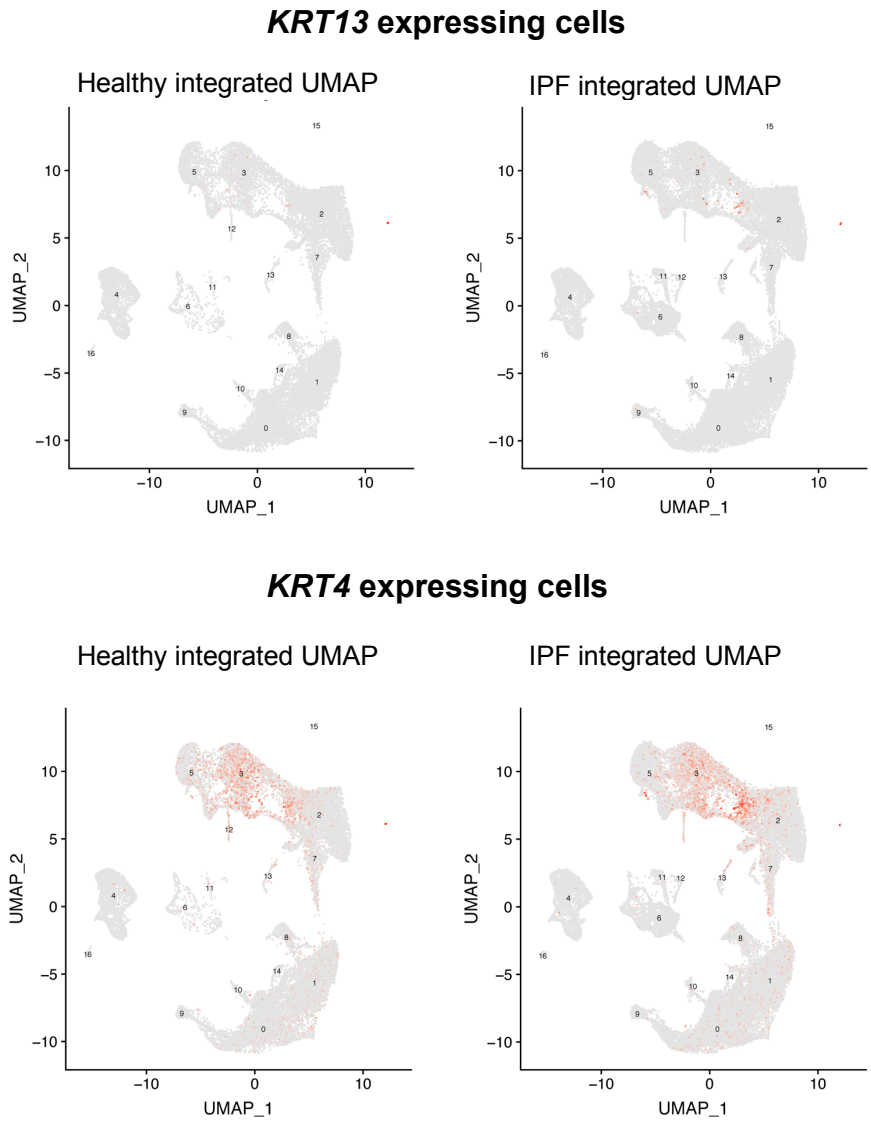


Figure 5.19 *KRT13* and *KRT4* expressing cells visualised on integrated UMAP plots. Distribution of cells expression 'hillock' cell markers (A) *KRT13* and (B) *KRT4* on integrated UMAP representations by condition. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

5.4.6 Differential expression analysis

Differential expression testing of gene expression between cells from clusters in IPF and healthy controls was performed in Seurat. This revealed that the greatest number of differentially expressed genes were in the airway immune compartment, specifically macrophage clusters 6 and 12 (Fig. 5.20 A). By reducing the log2 fold change threshold of average expression between the two groups to 0.1, a larger number of differentially expressed genes in the airway epithelial compartment was exposed (Fig 5.20 B). With this threshold, there were more than 500 differentially expressed genes between IPF and healthy secretory clusters 2 and 7, and ciliated cluster 1. This indicated that whilst the greatest gene expression changes are seen in airway macrophage populations, there are large numbers of significant, low amplitude gene expression changes in airway epithelial cell populations.

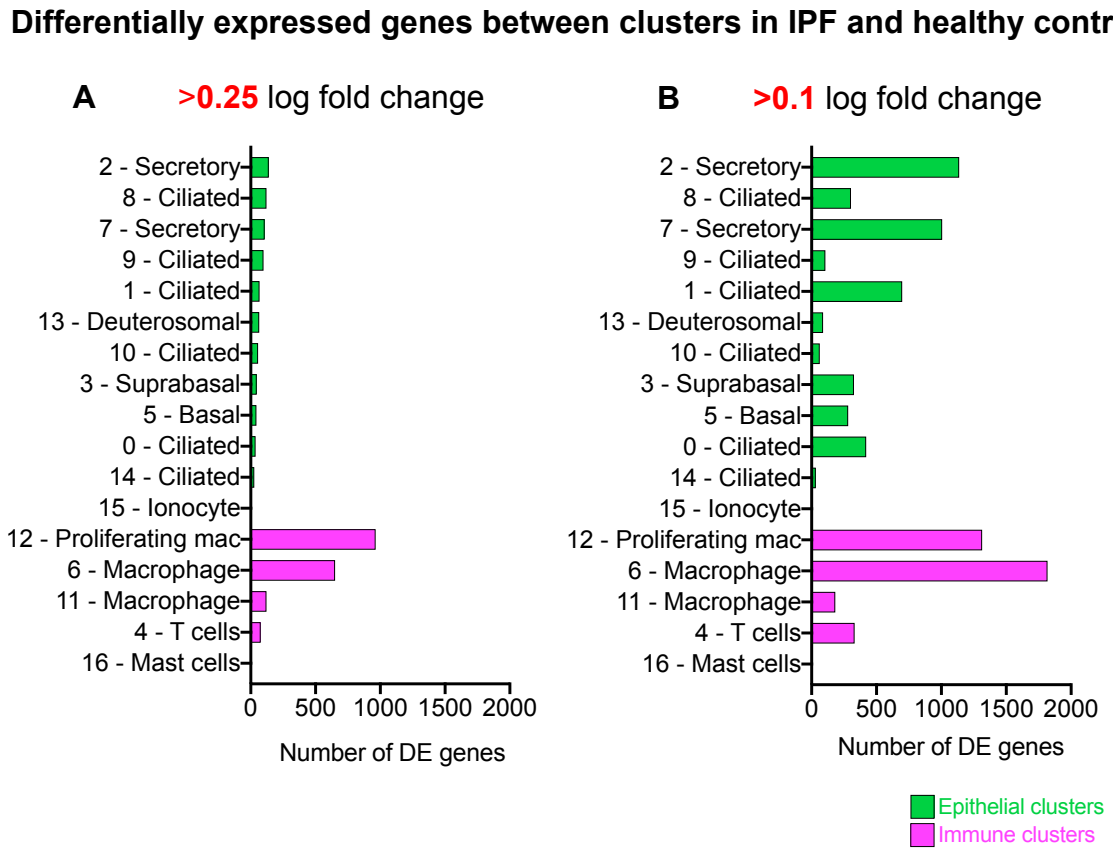
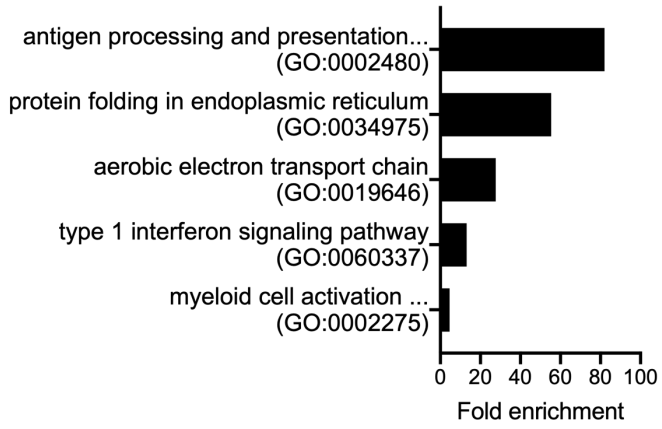


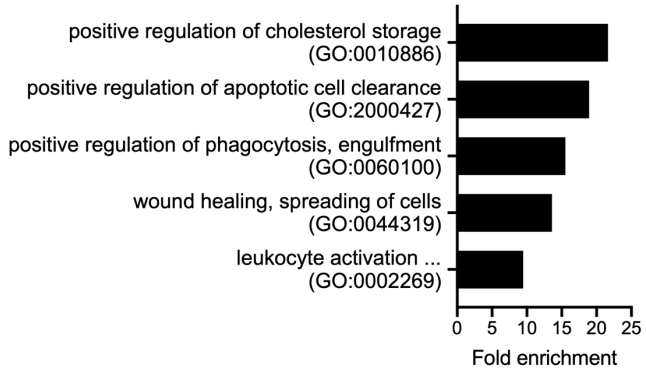
Figure 5.20 Differentially expressed genes between clusters in IPF and healthy control. Differential expression testing performed in Seurat based on Wilcoxon rank sum test. Number of differentially expressed genes (DE) with thresholds of **(A)** >0.25 log2 fold change, Bonferroni adjusted p <0.05, and (**B**) >0.1 log2 fold change, Bonferroni adjusted p <0.05.

To gain further insight into the biological processes enriched in IPF clusters compared to healthy controls, Gene Ontology (GO) enrichment analysis of differentially expressed genes was performed using the PANTHER tool online (Fig 5.21). This highlighted enrichment in processes such as antigen presentation (*HLA-A, HLA-B, HLA-C*), aerobic electron transport (*MT-CO1, MT-CO2, MT-CO3),* type 1 interferon signalling (*IFI27, BST2*) and myeloid cell activation (*ANXA2, S100P, CEACAM6*) in IPF secretory cells (cluster 2) compared to controls. Interestingly, mucosal chemokine *CXCL17*, known to attract immature dendritic cells and monocytes, was also significantly upregulated in this cluster in IPF compared with healthy controls. Genes overexpressed in IPF macrophages (cluster 6), were linked to apoptotic cell clearance (*TREM2, CD300LF, C2)*, phagocytosis (*CD36, RAB31, PPARG),* wound healing (*RHOA, CD44, ADAM17)* and leucocyte activation (*ITGAM, GRN, TYROBP).* IPF proliferating macrophages (cluster 12) demonstrated enrichment in antigen presentation (*HLA-DRB1, HLA-DRA*, *HLA-DMA*), monocyte extravasation (*JAML, CD47, SIRPA)* and IL3 signalling (*SYK, FCER1G, IL3RA)* processes.

A Secretory cells (cluster 2)



B Macrophages (cluster 6)



C Proliferating macrophages (cluster 12)

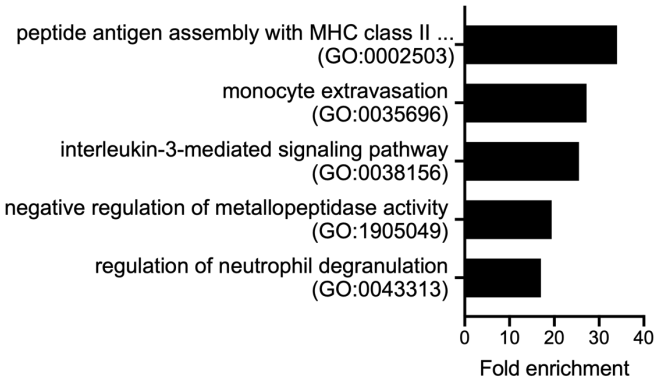


Figure 5.21 Gene Ontology analysis of clusters with greatest number of DE genes. Selected processes from GO Enrichment Analysis using significantly upregulated genes (>0.25 log2 fold change, Bonferroni adjusted p <0.05) for (**A**) secretory cluster 2 (*n=113 genes)* (**B**) macrophage cluster 6 (n=402 genes) and (**C**) proliferating macrophages cluster 12 (n=596 genes). All GO biological processes displayed were analysed with a Fisher's Exact test, FDR P <0.05.

5.4.7 Basal cell subclustering analysis

Basal cell populations have been reported to be expanded in the distal lung of patients with advanced IPF[53,145], but little is known about their potential role in the aetiology of the disease. To shed light on this I focused specifically on the basal cell subsets identified in our scRNA-seq dataset of airway brushings in newly diagnosed cases of IPF.

The proportion of airway basal cells (cluster 5) and suprabasal cells (cluster 3) were increased in some of the IPF donors compared to controls in our study (Fig. 5.16). To examine these populations in more detail, 10666 cells (*n*= 3844 from healthy controls, *n*= 6822 from IPF patients) from clusters 3 and 5 of the complete aggregated dataset were subclustered in Seurat. This revealed 11 basal cell subclusters at a resolution of 0.2 (Fig. 5.22 and 5.23). Cluster annotation was performed using marker genes associated with basal cell subpopulations from published literature[179,198,199], and scCATCH.

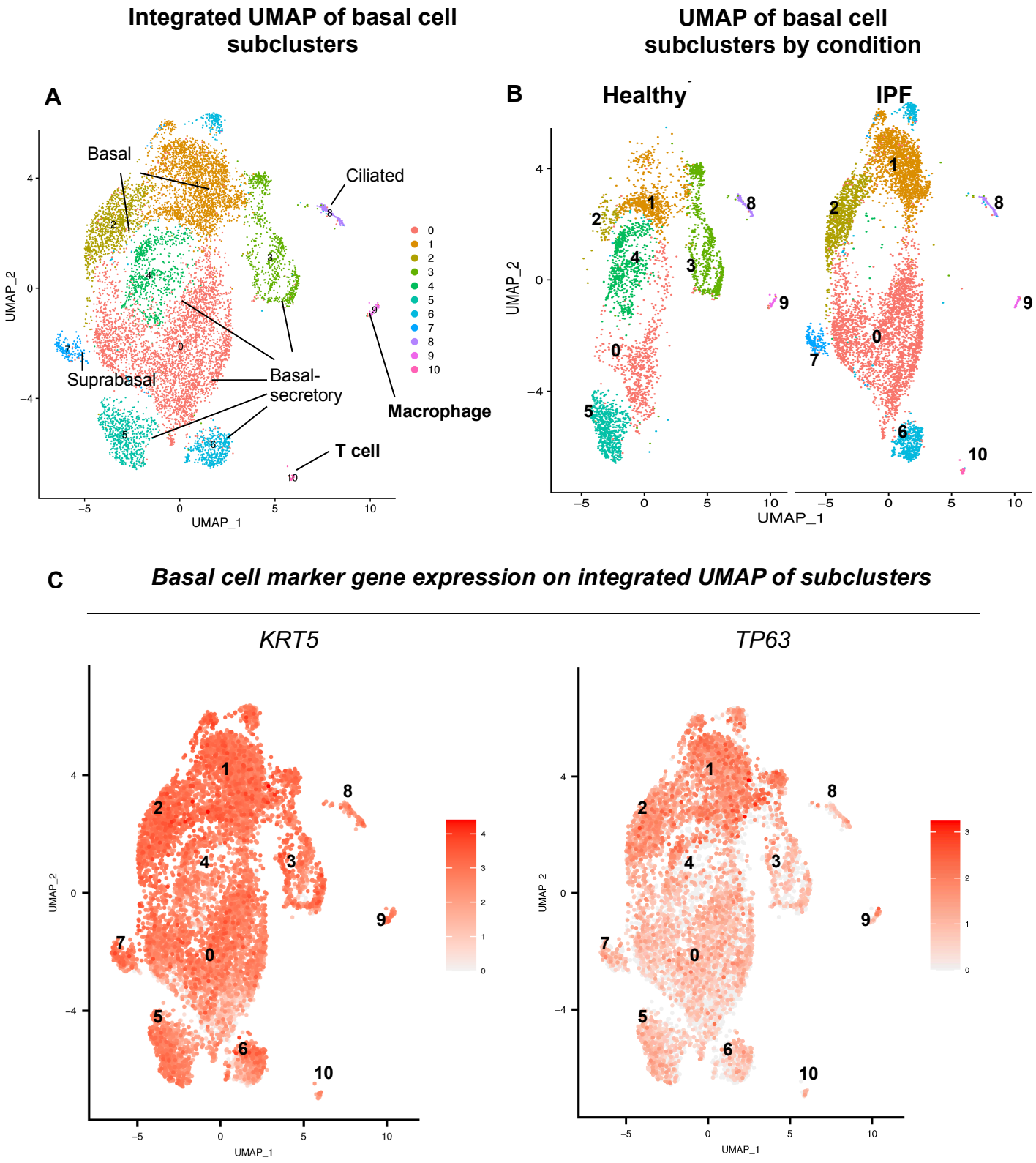


Figure 5.22 Basal cell subclusters. Cells from clusters 3 and 5 of the complete integrated dataset were re-clustered in Seurat. (**A**) UMAP visualisation of basal cell subclusters of integrated dataset (IPF and control) with a cluster resolution of 0.2 (*n*=10666 cells) (**B**) UMAP visualisation of basal cell subclusters by condition (**C**) Visualisation of canonical basal cell marker genes *KRT5* and *TP63* on the integrated UMAP using *FeaturePlot* function in Seurat. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0*

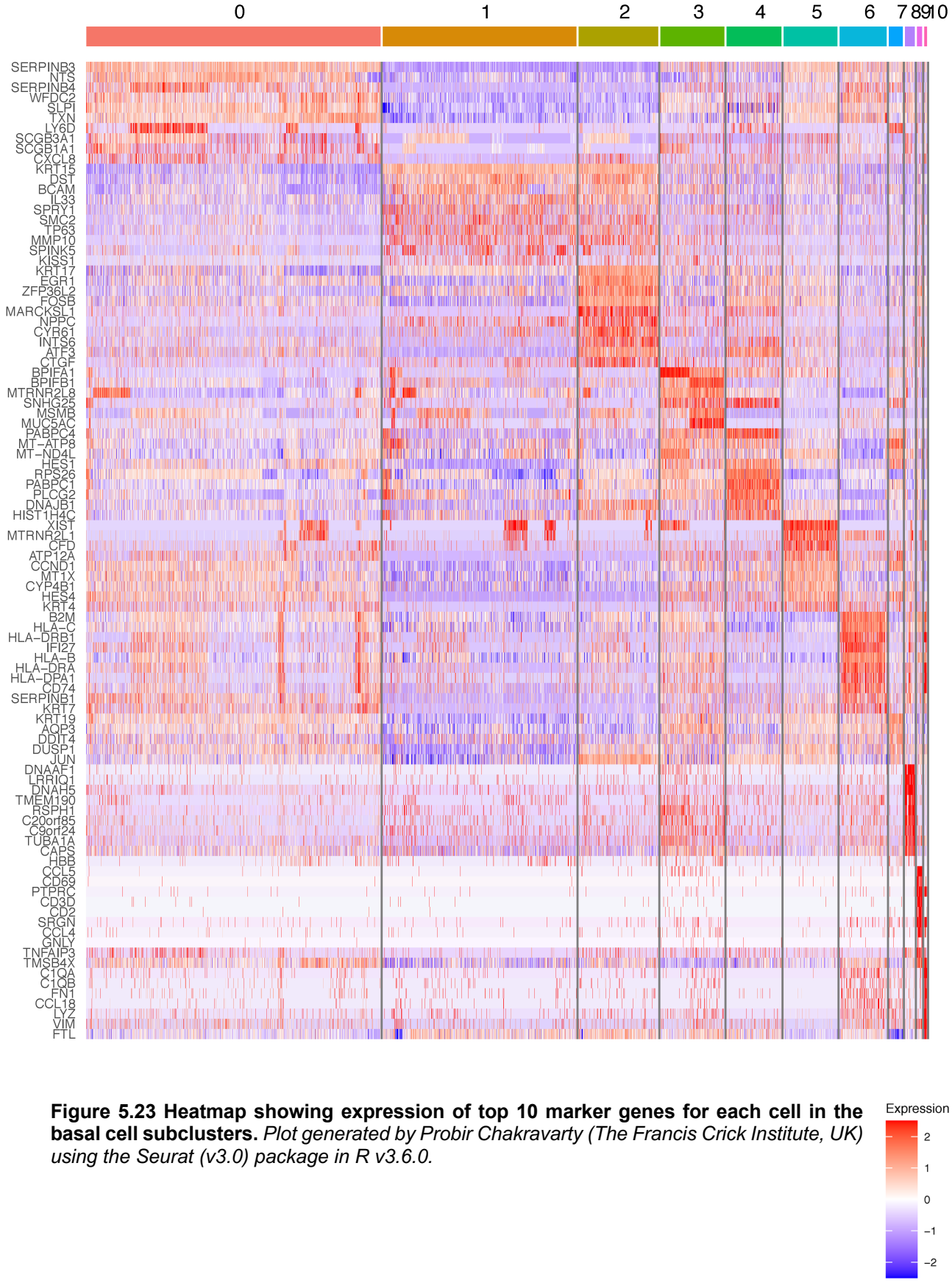


Figure 5.23 Heatmap showing expression of top 10 marker genes for each cell in the basal cell subclusters. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

The highest expression of *KRT5, TP63* and other classical basal cell marker genes (see Table 1.1) were found in subclusters 1 and 2 (Fig. 5.22 C). These subclusters were generally more abundant in IPF subjects than healthy controls. Suprabasal cells in subcluster 7 were characterised by higher expression of *KRT19, NOTCH3* and *HES4*. Interestingly, subclusters expressing secretory marker genes including *SCGB1A1* (subclusters 0 and 4), *SCGB3A1* (subcluster 5), *MUC5AC* (subcluster 3) were also identified. Basal cells are multipotent airway stem cells[397], therefore the presence of secretory gene markers in some basal cell subclusters is likely to represent the early stages of basal cell differentiation into secretory cells. Accordingly, these subclusters were annotated 'basal-secretory'. Subcluster 6 was enriched in genes *HLA-DRA* and *HLA-C,* encoding antigen-presenting Major Histocompatibility Complex (MHC), Class I and II molecules respectively, in addition to *KRT7* a cytokeratin associated with secretory club cells[204]. A small subcluster of basal cells expressing ciliated marker *FOXJ1* (subcluster 8) was also identified.

UMAP visualisation of each individual donor (Fig. 5.24) revealed other clusters driven predominantly by individual donors rather than condition, such as subcluster 4 from subject 1055 (healthy), subcluster 5 from subject 1058 (healthy) and subcluster 7 from subject 5756 (IPF). Surprisingly, there were two small subclusters (9 and 10) expressing marker genes for T cells and macrophages (Fig. 5.25). The T cell subcluster comprised a total of 37 cells and were exclusively present in 6 out of 9 of the IPF subjects. It is likely that these represent doublets between T cells and basal cells (as we did not exclude doublets from the analysis) and may be relevant to the pathogenesis of the disease.

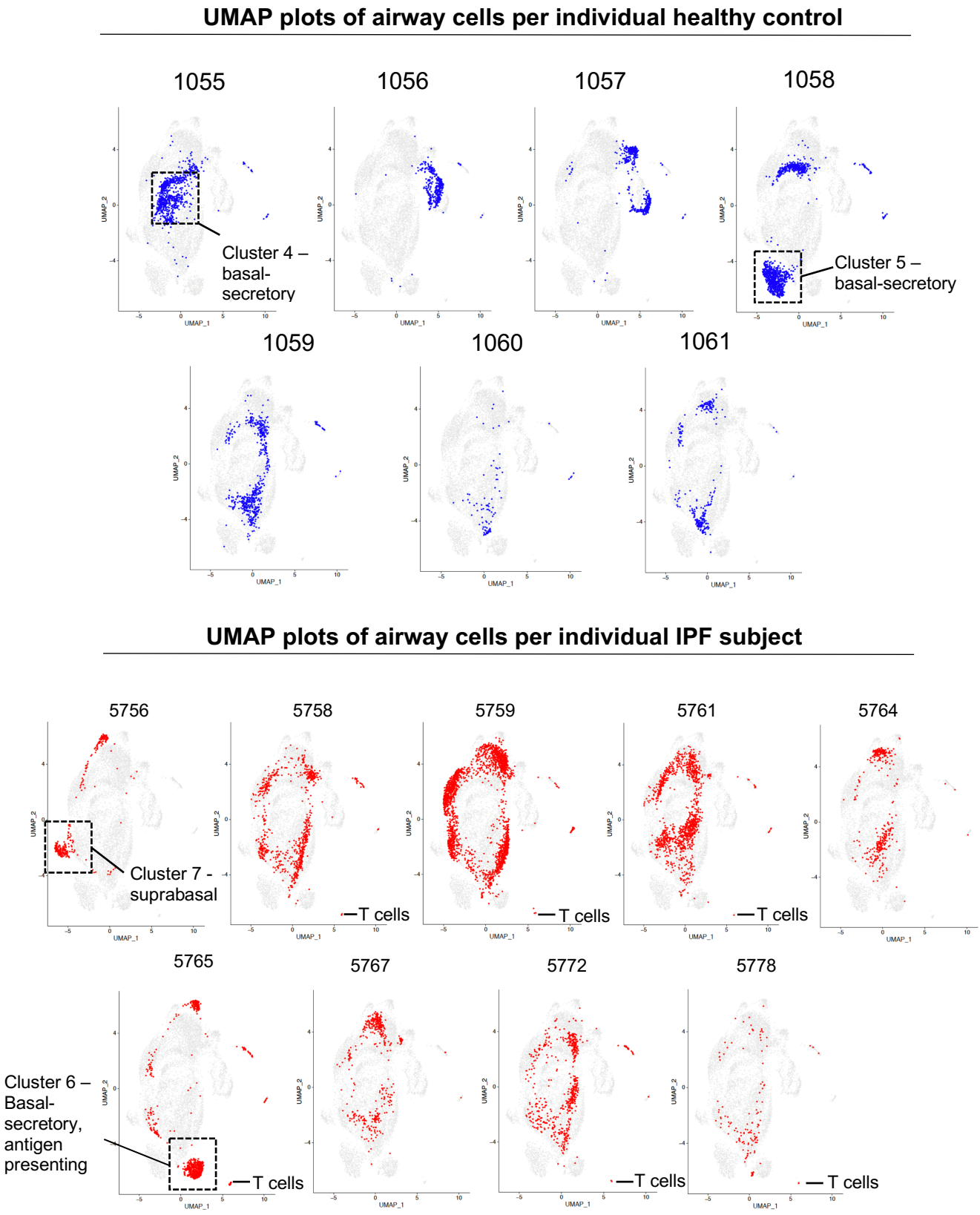


Figure 5.24 UMAP visualisation of basal cell subclusters per individual subject. UMAP representations showing contribution to basal cell sub-clusters by individual subjects; IPF (red) or healthy control (blue). A dashed box highlights clusters driven predominantly by one subject. IPF subjects with a T cell subcluster also indicated. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0*

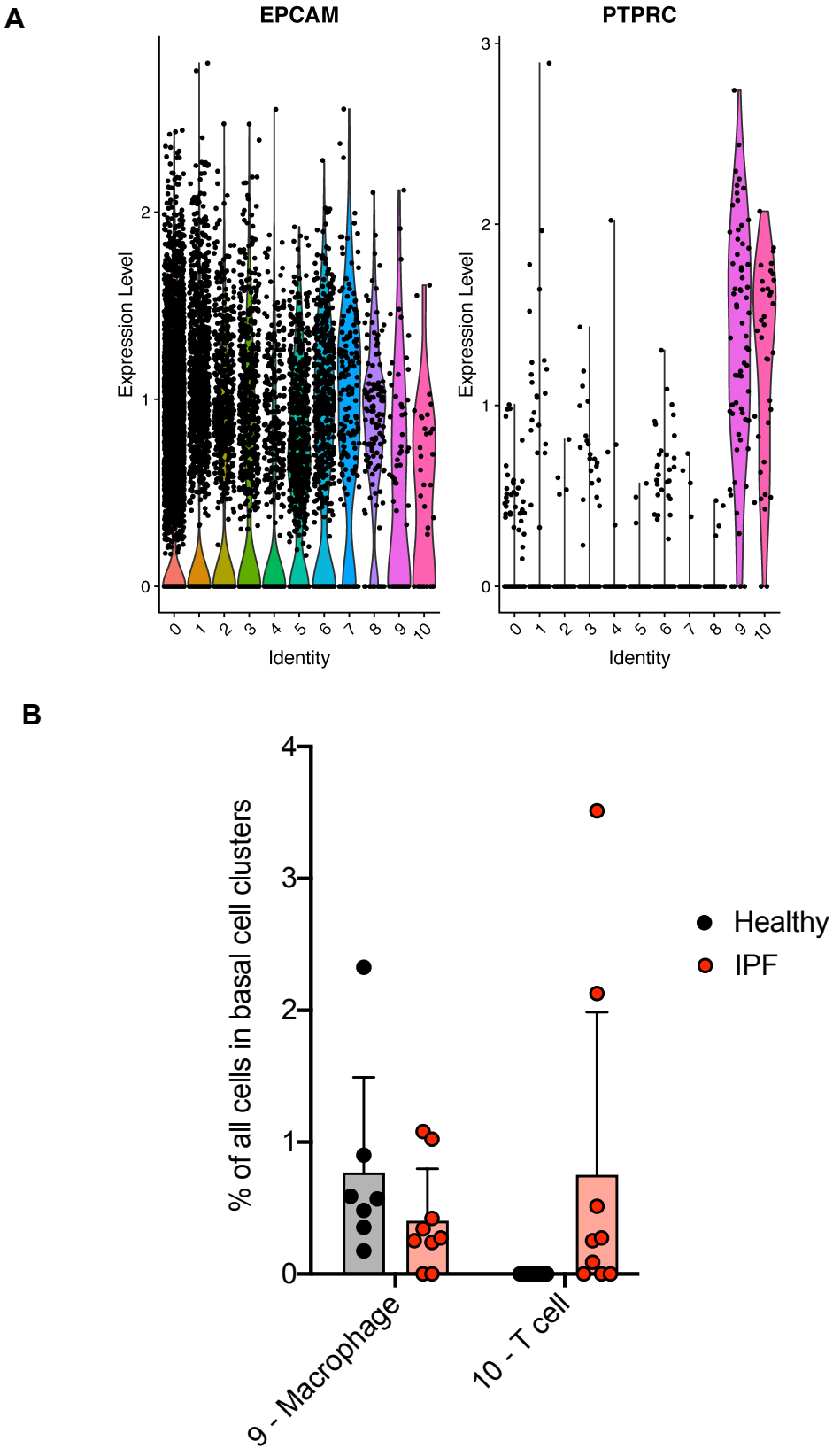


Figure 5.25 Immune cell contamination in basal cell subclustering. (A) Log2 normalised gene expression of epithelial *(EPCAM)* and immune *(PTPRC encoding CD45)* lineage markers in basal cell sub-clusters (0 – 10). **(B)** Percentage of macrophages and T cells in all subclustered cells per donor.

To gain a clearer perspective of basal cell subpopulations in the airway, we removed contaminating immune cell clusters by positively selecting cells expressing epithelial lineage marker, *EPCAM* (normalised expression >0.25) and excluding cells expressing immune cell lineage marker, *PTPRC also known as CD45* (normalised expression <0.1) (Fig 5.25). Subclustering these "filtered" basal cells revealed 8 clusters at a resolution of 0.2 (Fig 5.26 and 5.27). Expression of *KRT5* and *TP63* was confirmed in all subclusters (Fig 5.26 C). To assign annotation to the individual subclusters we performed GSEA using signature marker genes from the Carraro et al.[199] dataset, GO analysis to understand differences in biological processes between subclusters, and scCATCH to further validate.

Basal cells in subcluster 1 were defined by highest expression of *TP63* and other characteristic basal cell markers (see Table 1.1) such as *KRT15* and *BCAM* (Fig 5.27). Subcluster 2 expressed markers associated with the 'activated basal cell' state described by Carraro et al.[199] including *FOSB, JUNB* and *ATF3.* Subcluster 3 was enriched in mitochondrial genes suggestive of a secretory preparatory state[198], secretory protein genes such as *BPIFA1,* and suprabasal marker genes[179,204] including *NOTCH3, HES4* and *CCND1.* Subclusters 0, 4, 5 and 6 were annotated 'basal-secretory cells', as they displayed secretory marker genes and likely represented a differentiating basal cell subsets. For example, early secretory markers, *SCGB1A1* and *WFDC2,* were overexpressed in cluster 0[198]. Subcluster 5 displayed MHC Class I and II genes suggesting antigen presenting capacity and *KRT8* a cytokeratin associated with differentiating luminal cell types[195]. A small subcluster of cells with ciliated genes (cluster 7) was also present.

There were marked differences in the topography of the basal cell subclusters when comparing UMAP representations by condition (Fig. 5.26 B). The cell proportions in each subcluster were quantified in Fig. 5.28. Interestingly, subcluster 2 'basal – activated' were more abundant in IPF compared to healthy controls.

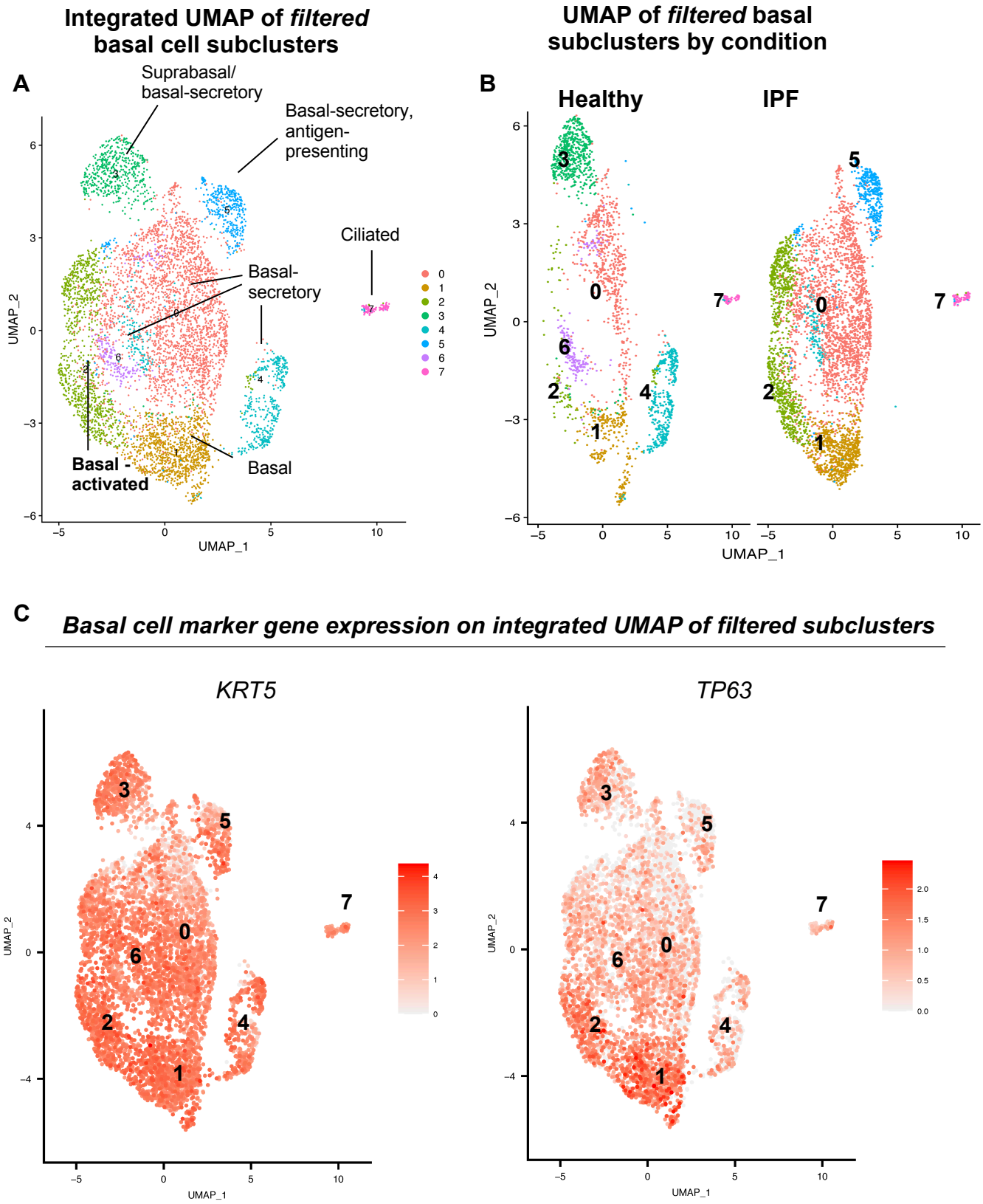


Figure 5.26 Filtered basal cell subclusters. Cells from clusters 3 and 5 of the overall integrated dataset expressing *EPCAM* > 0.25 and PTRPC < 0.1 were re-clustered in Seurat. (**A**) UMAP visualisation of integrated basal cell subclusters (IPF and control) with a cluster resolution of 0.2 (*n=6454* cells) (B) UMAP visualisation of basal cell subclusters by condition (**C**) Visualisation of canonical basal cell marker genes *KRT5* and *TP63* on the integrated UMAP using *FeaturePlot* function in Seurat. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0*

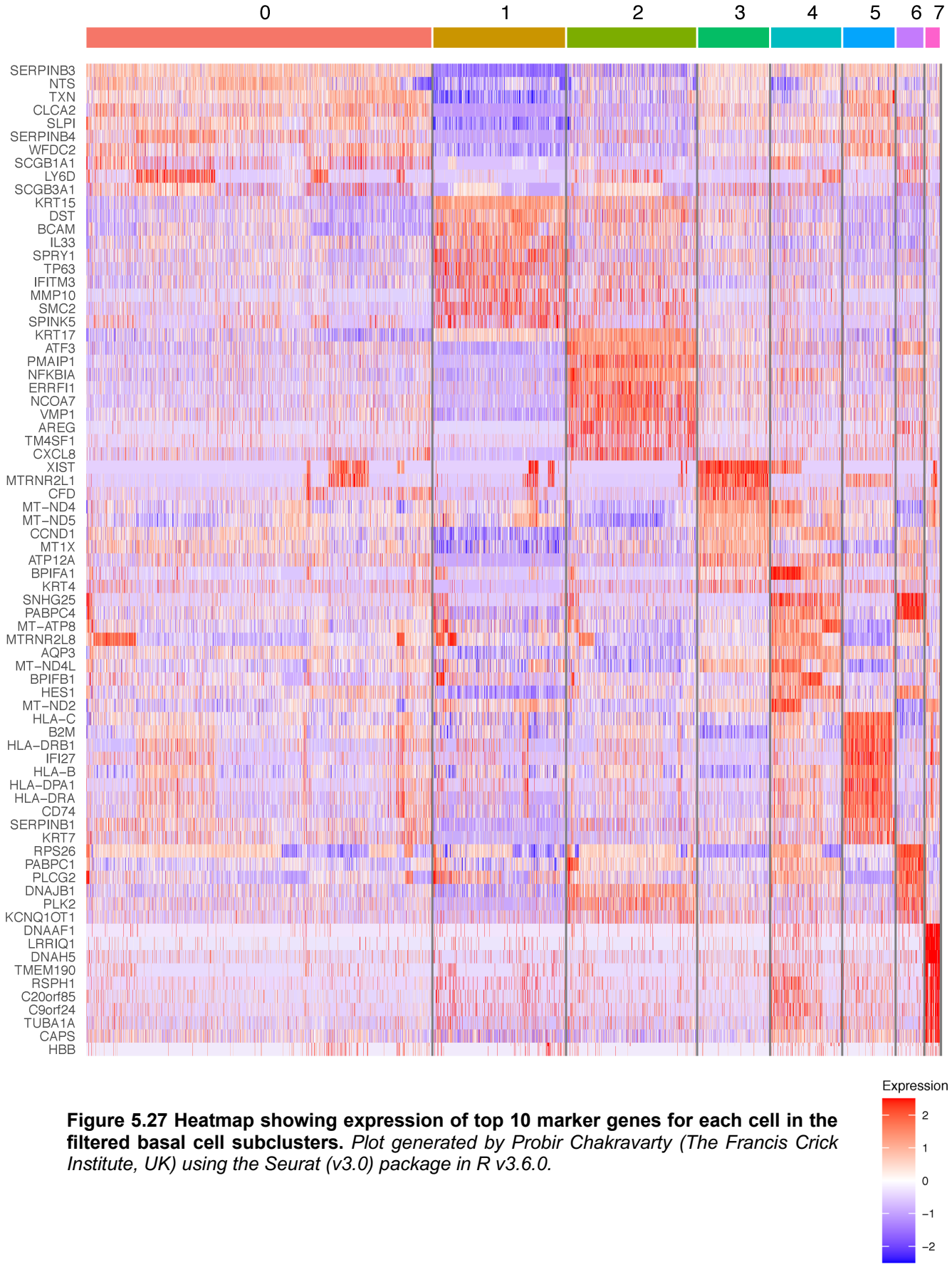


Figure 5.27 Heatmap showing expression of top 10 marker genes for each cell in the filtered basal cell subclusters. *Plot generated by Probir Chakravarty (The Francis Crick Institute, UK) using the Seurat (v3.0) package in R v3.6.0.*

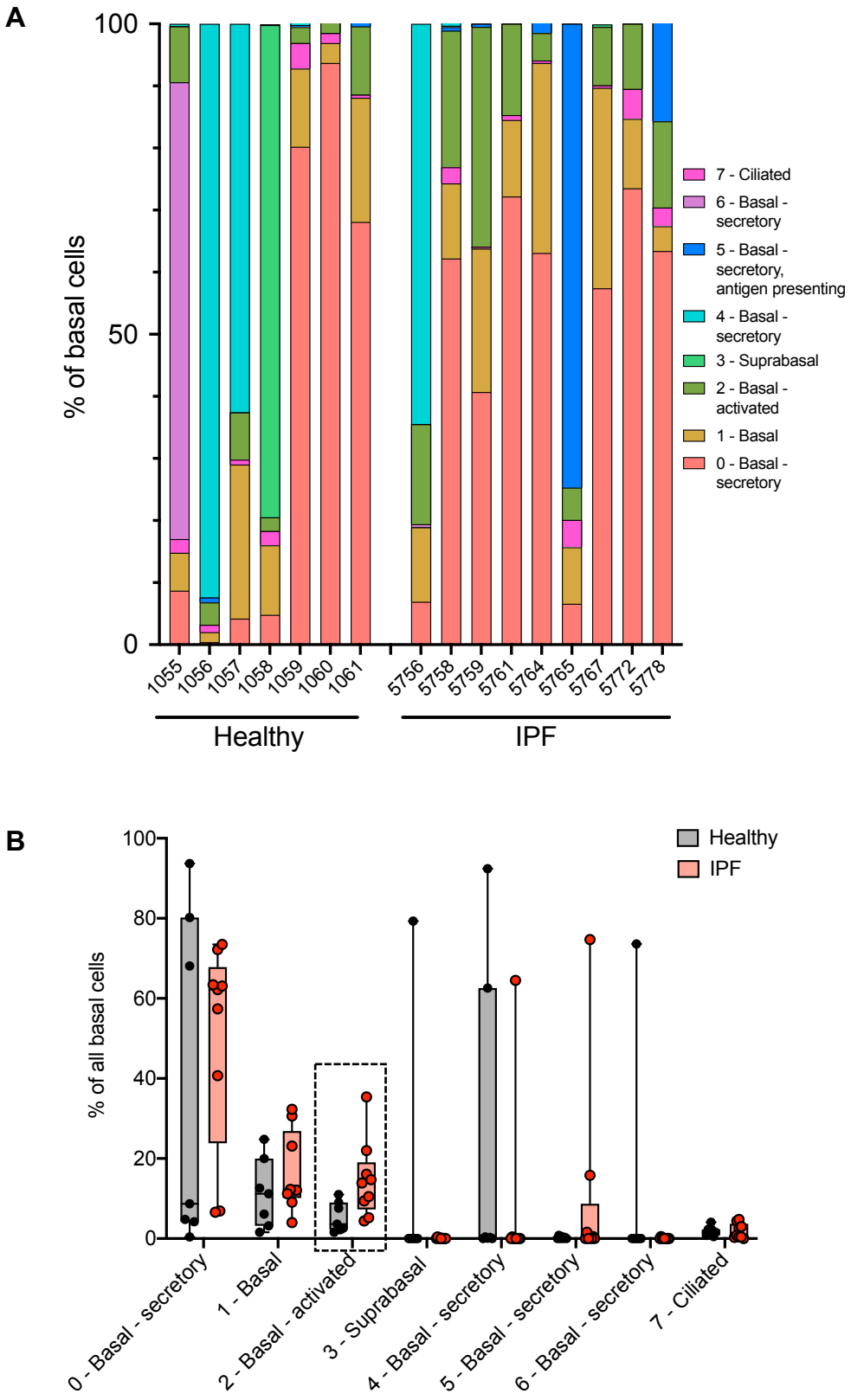


Figure 5.28 Quantification of basal cell subclusters in IPF and healthy airways. (**A**) Stacked bar chart quantifying percentage of each basal cell cluster per subject. (**B**) Boxplot representing the percentage of each basal cell cluster per subject comparing IPF (*n*=9) and healthy controls (*n*=7). Each dot represents a single subject, whiskers showing min. to max values. Dashed box indicates a cluster of interest.

Basal cell subcluster 1 was characterised by GO processes linked to lung development, hemidesmosome assembly and epithelial cell fate (Table 5.5). 'Basal-activated' cells in subcluster 2 were enriched in GO terms associated with 'integrated stress response signalling (GO:0140467)', regulation of fibroblast activity including 'response to fibroblast growth factor (GO:0071774)', and regulation of mesenchymal cell proliferation (GO:0010464). Interestingly, there were also pathways linked to T cell modulation such as 'regulation of gamma-delta T cell differentiation (GO:0045586)' and 'regulation of alpha-beta T cell proliferation (GO:0046640)' which may explain the presence of a small subcluster of possible basal cell - T cell doublets in the initial basal cell subclustering analysis.

The basal-secretory subclusters (0, 3, 4, 5 and 6) were enriched in genes linked to protein targeting to ER, mitochondrial electron transport and detoxification. Interestingly, GO processes such as response to wounding and cell migration were overrepresented in subcluster 4. Basal-secretory cluster 5 was enriched in pathways linked to antigen presentation, protection from immune cell mediated cytotoxicity and protein localisation to the telomere.

Noteworthy was that basal-secretory subcluster 5 with a gene signature of antigen-presentation (Table 5.5), was most highly represented in IPF subject 5756 who was subsequently found to be colonised with *Mycobacterium intracellulare* (MAI) and *Klebsiella pneumoniae* on culture testing of bronchoalveolar lavage (BAL) samples. Changes in clusters 3 and 6 were driven predominantly by individual subjects, rather than condition; subcluster 3 from subject 1058, and subcluster 6 from subject 1055.

Table 5.5 GO enrichment analysis for basal cell subclusters

Cluster *(genes submitted for GO)*	GO biological process	Fold enrichment	FDR
0 – Basal- secretory (*n* = 268 genes)	cellular response to prostaglandin D stimulus (GO:0071799)	56.46	6.11E-03
	aerobic electron transport chain (GO:0019646)	33.87	1.67E-08
	retinoic acid biosynthetic process (GO:0002138)	28.23	2.36E-02
	protein targeting to ER (GO:0045047)	26.65	4.56E-37
	detoxification of inorganic compound (GO:0061687)	20.91	1.11E-03
1 – Basal (*n* = 301 genes)	bud elongation involved in lung branching (GO:0060449)	29.3	1.95E-02
	lung saccule development (GO:0060430)	22.79	3.03E-02
	hemidesmosome assembly (GO:0031581)	22.79	5.80E-03
	epithelial cell fate commitment (GO:0072148)	19.53	8.39E-03
	epithelial cell differentiation involved in prostate gland development (GO:0060742)	18.64	4.30E-02
2 – Basal - activated **(*n* = 484 genes)**	regulation of TRAIL-activated apoptotic signaling pathway (GO:1903121)	32.05	7.68E-03
	hemidesmosome assembly (GO:0031581)	17.8	9.29E-04
	PERK-mediated unfolded protein response (GO:0036499)	16.43	1.19E-03
	mammary gland epithelial cell proliferation (GO:0033598)	14.24	7.76E-03
	fibroblast migration (GO:0010761)	13.15	9.52E-03
3 – Suprabasal / Basal-secretory (*n* = 292 genes)	electron transport coupled proton transport (GO:0015990)	55.05	4.38E-04
	positive regulation of mammary gland epithelial cell proliferation (GO:0033601)	20.64	2.70E-02
	positive regulation of sodium ion transmembrane transporter activity (GO:2000651)	17.2	8.22E-03
	hemidesmosome assembly (GO:0031581)	17.2	3.72E-02
	mammary gland epithelial cell differentiation (GO:0060644)	16.19	9.62E-03
4 – Basal- secretory (*n* = 618 genes)	electron transport coupled proton transport (GO:0015990)	26.69	5.77E-03
	regulation of transepithelial transport (GO:0150111)	25.02	3.73E-02
	positive regulation of metallopeptidase activity (GO:1905050)	14.83	2.25E-02
	regulation of vascular endothelial growth factor production (GO:0010574)	7.08	9.33E-03
	positive regulation of response to wounding (GO:1903036)	4.42	2.19E-02
5 – Basal- secretory (*n* = 1068 genes)	peptide antigen assembly with MHC class II protein complex (GO:0002503)	18.97	1.07E-02
	MHC class I protein complex assembly (GO:0002397)	18.97	4.74E-02
	protection from natural killer cell mediated cytotoxicity (GO:0042270)	12.65	2.50E-02
	retinoic acid biosynthetic process (GO:0002138)	11.86	8.40E-03
	positive regulation of establishment of protein localization to telomere (GO:1904851)	11.38	2.83E-03

Table 5.6 GO enrichment analysis for basal cell subclusters (continued)

Cluster	GO biological process	Fold enrichment	FDR
6 – Basal-secretory (*n* = 320 genes)	positive regulation of intrinsic apoptotic signaling pathway by p53 class mediator (GO:1902255)	38.73	1.49E-02
	cotranslational protein targeting to membrane (GO:0006613)	28.76	1.68E-42
	protein targeting to ER (GO:0045047)	25.14	8.21E-40
	translational initiation (GO:0006413)	22.26	7.53E-43
	mitochondrial electron transport, cytochrome c to oxygen (GO:0006123)	16.14	3.13E-03
7 – Ciliated (*n* = 1331 genes)	axonemal central apparatus assembly (GO:1904158)	15.67	1.93E-02
	intraciliary retrograde transport (GO:0035721)	14.46	1.45E-06
	inner dynein arm assembly (GO:0036159)	13.71	1.34E-07
	protein localization to ciliary transition zone (GO:1904491)	13.71	1.91E-03
	epithelial cilium movement involved in determination of left/right asymmetry (GO:0060287)	13.05	4.35E-05

Gene Ontology (GO) enrichment analysis (PANTHER) using gene list generated by *FindMarkers* function in Seurat based on a Wilcoxon rank sum test comparing average expression in cells belonging to one cluster to all cells within the experiment. Only genes showing >0.1 log2 fold change with a Bonferroni adjusted p-value <0.05 were used for GO analysis. Key enriched GO biological processes shown. Results displayed were analysed with a Fisher's Exact test, FDR P <0.05.

Notch signalling is critically important in basal cell fate decisions[195,204], therefore we evaluated the expression of genes in this pathway in our basal cell subclusters (Fig. 5.28). In comparison to basal-secretory subclusters (0, 3, 4, 5, 6) where Notch pathway genes were upregulated, basal cell subclusters (1 and 2) generally showed lower expression of receptor genes *NOTCH2* and *NOTCH3,* and pathway target genes *HES4* and *HEY1*[204]*.* As expected however, basal cell subclusters 1 and 2 expressed higher levels of Notch ligands *JAG2* and *DLL1*[204]*.* Overall, these findings support the conclusion that there are basal cell subclusters undergoing active differentiation.

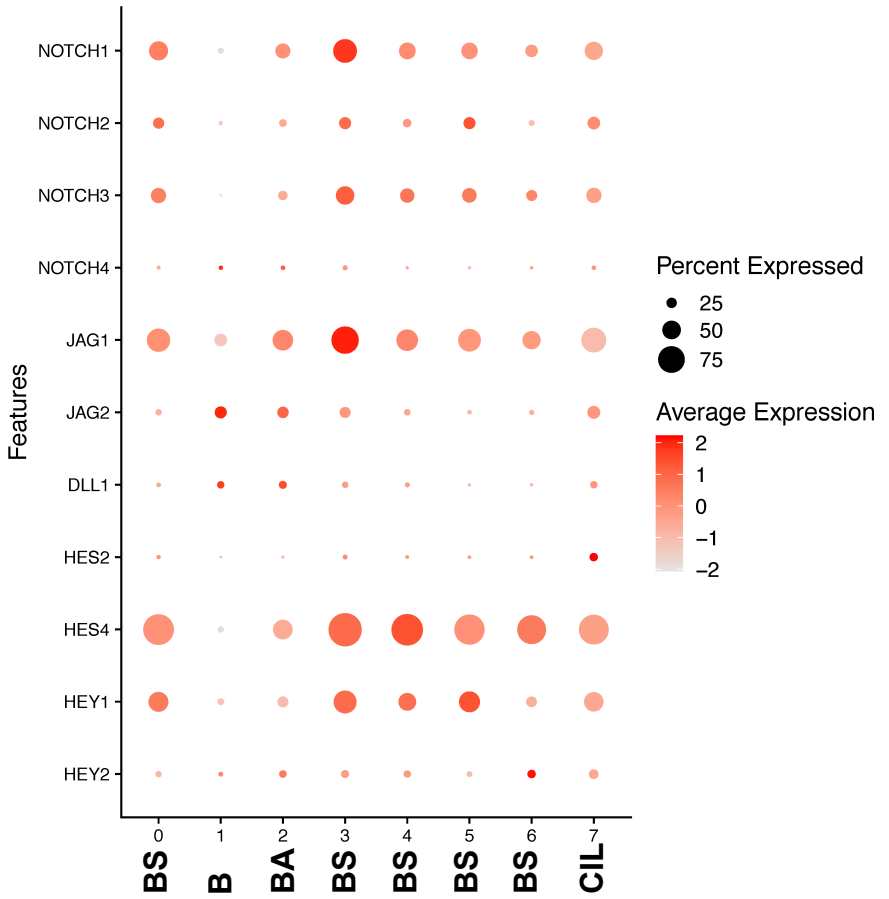


Figure 5.29 Dot plot of the expression of genes related to the Notch pathway in each of the basal cell subclusters.

B = basal
BA = basal activated
BS = basal-secretory
CIL = ciliated

Overall, this subclustering analysis provided valuable insight into the nuanced processes occurring in basal cell subsets in newly diagnosed cases of IPF. Some of the subsets were more abundant in the IPF airway compared to healthy controls therefore may be relevant in the evolution and progression of the disease.

5.4.8 Trajectory analysis of integrated dataset

Trajectory inference methods also known as pseudotime analysis, use single-cell transcriptomic data to computationally model cell differentiation by ordering cells along a developmental trajectory according to similarities in gene expression patterns[203]. We first used an unsupervised algorithm, 'Monocle version 3'[256,257], to infer epithelial cell differentiation in the healthy and IPF airway samples. Immune cell clusters were

excluded from the analysis by taking forward only those cells which expressed epithelial marker, *EPCAM* (log2 normalised expression >0.4) and not immune cell marker, *PTPRC/ CD45* (log2 normalised expression <0.1) (n=49140). The global structure of the epithelial dataset as visualised by UMAP demonstrated a natural ordering of clusters from basal cells, through secretory cell clusters to ciliated clusters (Fig. 5.30 A). Cluster 13 (deuterosomal cells) and cluster 15 (ionocytes) - also within the epithelial lineage - were detached from the main epithelial cell clusters. Basal cells represent the earliest developmental stage in epithelial differentiation, so the root-node of the trajectory was selected by taking the average of the co-ordinates of each cell in cluster 5. Default parameters for Monocle 3 were used in the analysis, except for the cluster cells function in which we used k = 3, res = 10^{-5}. Using the UMAP plot co-ordinates, Monocle 3 was unable to call a single trajectory that incorporated clusters 13 and 15 (Fig. 5.30 A). This was likely due to the absence of intermediate cells transitioning to cluster 13 and 15. The algorithm drew an average trajectory path through the three main clusters; basal, secretory and ciliated (Fig. 5.30 B). A number of branches were evident in the ciliated clusters. Cells were then ordered in pseudotime which in this case quantitates the progress of a cell through the process of epithelial differentiation based on gene expression of varying genes (Fig. 5.30 C). Ciliated cluster 9 was last in pseudotime so represented a terminal differentiated state. Other branches could be seen with multiple clusters appearing at the same time point for example clusters 1, 8 and 14. There were no significant differences between IPF and healthy controls when comparing epithelial cell cluster ordering on mode of pseudotime values (Fig. 5.30 C). However, there was a population of secretory cells in cluster 7 which appeared marginally earlier in healthy controls as compared to IPF which may warrant further investigation. Marker genes defining epithelial clusters were plotted over pseudotime to confirm known and predicted trends showing basal cell marker *KRT15* highly expressed early in pseudotime, and cilitated cell marker gene *TMEM190* expressed late in pseudotime.

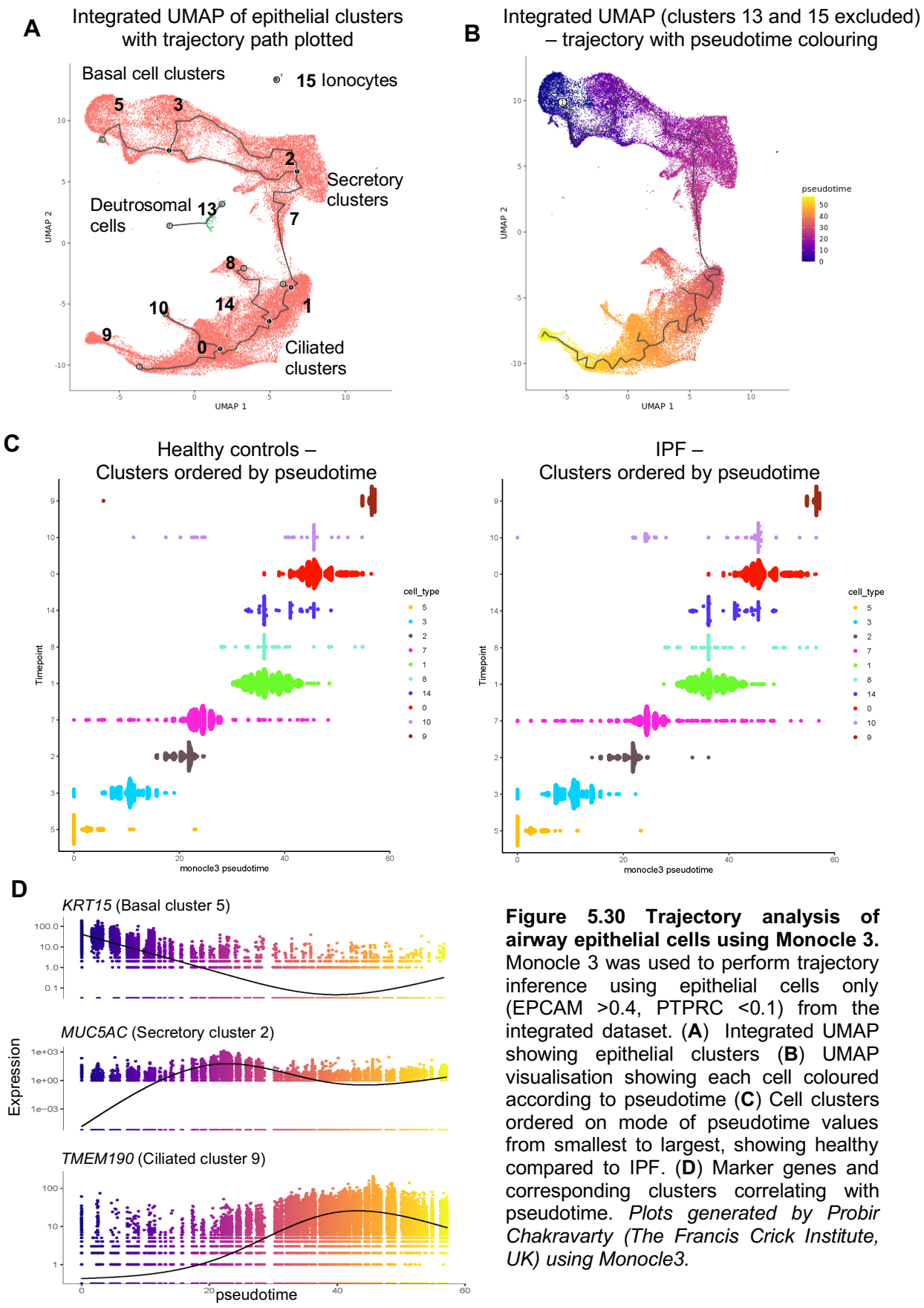


Figure 5.30 Trajectory analysis of airway epithelial cells using Monocle 3. Monocle 3 was used to perform trajectory inference using epithelial cells only (EPCAM >0.4, PTPRC <0.1) from the integrated dataset. (**A**) Integrated UMAP showing epithelial clusters (**B**) UMAP visualisation showing each cell coloured according to pseudotime (**C**) Cell clusters ordered on mode of pseudotime values from smallest to largest, showing healthy compared to IPF. (**D**) Marker genes and corresponding clusters correlating with pseudotime. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using Monocle3.*

Monocle 3 is only able to use UMAP co-ordinates for pseudotime analysis. For the reasons described above clusters 13 and 15 could not be incorporated into the pseudotime analysis using this computational approach. As a result, we opted to use 'Slingshot', which allows for pseudotime analysis using either UMAP or PCA co-ordinates.

'Slingshot'[258] is an alternative method for lineage and pseudotime inference which uses PCA data to build the trajectory. Immune cells were again removed prior to trajectory analysis by taking forward only those cells with *EPCAM* expression greater than 0.4 and *PTPRC/ CD45* expression less than 0.1. Using the top two prinicipal components, cluster 13 and 15 were situated closer to the other epithelial clusters – and more importantly there appeared to be intermediate cells between the epithelial clusters and the terminal cluster of cluster 15 (Fig. 5.31 A). In a PCA plot the similarity between cells is proportional to the distance between cells. Basal cell cluster 5 was set as the start of the lineage and ciliated clusters 8, 9, 10, 14 and ionocyte cluster 15, set as the terminal cell states. Slingshot-inferred pseudotime was calculated for cells in the PCA plot (Fig. 5.31 B) and seven trajectories were predicted (Fig. 5.31 C). Allowing for the multiple ciliated branches which arise because we chose not to combine these ciliated clusters, the trajectory analysis was largely consistent with prior biological knowledge of lineage relationships in the human airway epithelium[259]. The exception was deutrosomal cells in cluster 13, which are described as a transitionary cell type between secretory and ciliated cells in the literature[204], but appeared to be a terminally differentiated state in my dataset. Comparing cell clusters between IPF and Healthy using the pseudotime values per cluster or the order of cell differentiation showed that there was no significant differences between IPF and healthy controls (Fig 5.31 D).

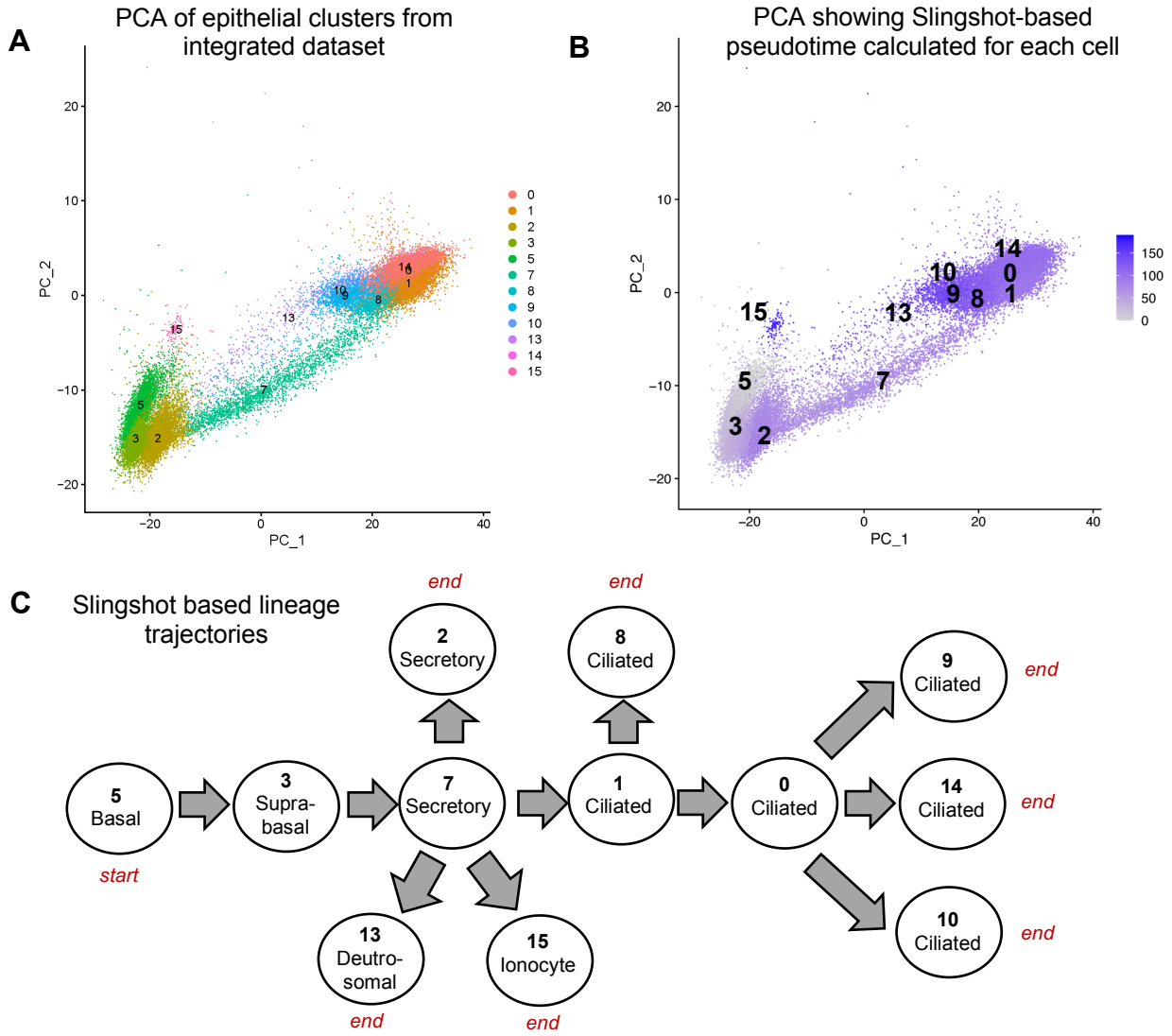


Figure 5.31 Trajectory analysis of airway epithelial cells using Slingshot. (A) PCA of first two principal components from integrated dataset including only epithelial cells (EPCAM >0.4, PTPRC <0.1) **(B)** PCA coloured by Slingshot- based pseudotime; a grey- to- purple gradient indicates progression of a cell through pseudotime **(C)** Schematic showing the ordering of clusters and branching nature of lineages **(D)** Horizontal violin plot demonstrating the distribution if Slingshot- calculated pseudotime per cluster. Clusters are ordered from bottom to top using the median pseudotime per cluster. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using Slingshot.*

5.4.9 Cell- cell interactions in the IPF airway

Having defined the epithelial and immune populations that exist in the airway niche, CellPhoneDB[261] was used to perform an unbiased analysis of the cell-cell interactions taking place. CellPhoneDB offers a publicly available repository of ligands, receptors and their interactions which is integrated with a computational approach that predicts cellular interactions between two cell types from scRNA-seq data. It has been previously used to dissect the cellular interactome in airway brushings from asthmatic[168] and COVID-19 patients[180]. As per Vieira Braga et al.[168], only ligands or receptors present in at least 30% of the cells in a cluster were considered in this analysis. In the healthy airway, predicted interactions were predominantly between secretory cells (clusters 2 and 7), suprabasal cells (cluster 3) and proliferating macrophages (cluster 12) (Fig. 5.33 and 5.34). In healthy controls and IPF subjects, secretory cluster 7 had the greatest total number of predicted significant interactions with all cell clusters (indicated by node size in Fig. 5.34), but in particular with proliferating macrophage cluster 12. In IPF, there was a marked increase in predicted epithelial-immune interactions between secretory cluster 7 and macrophage cluster 6, and also interactions between cells within the same lineage, with macrophage cluster 6 and proliferating macrophage cluster 12 (Fig. 5.33 and 5.34).

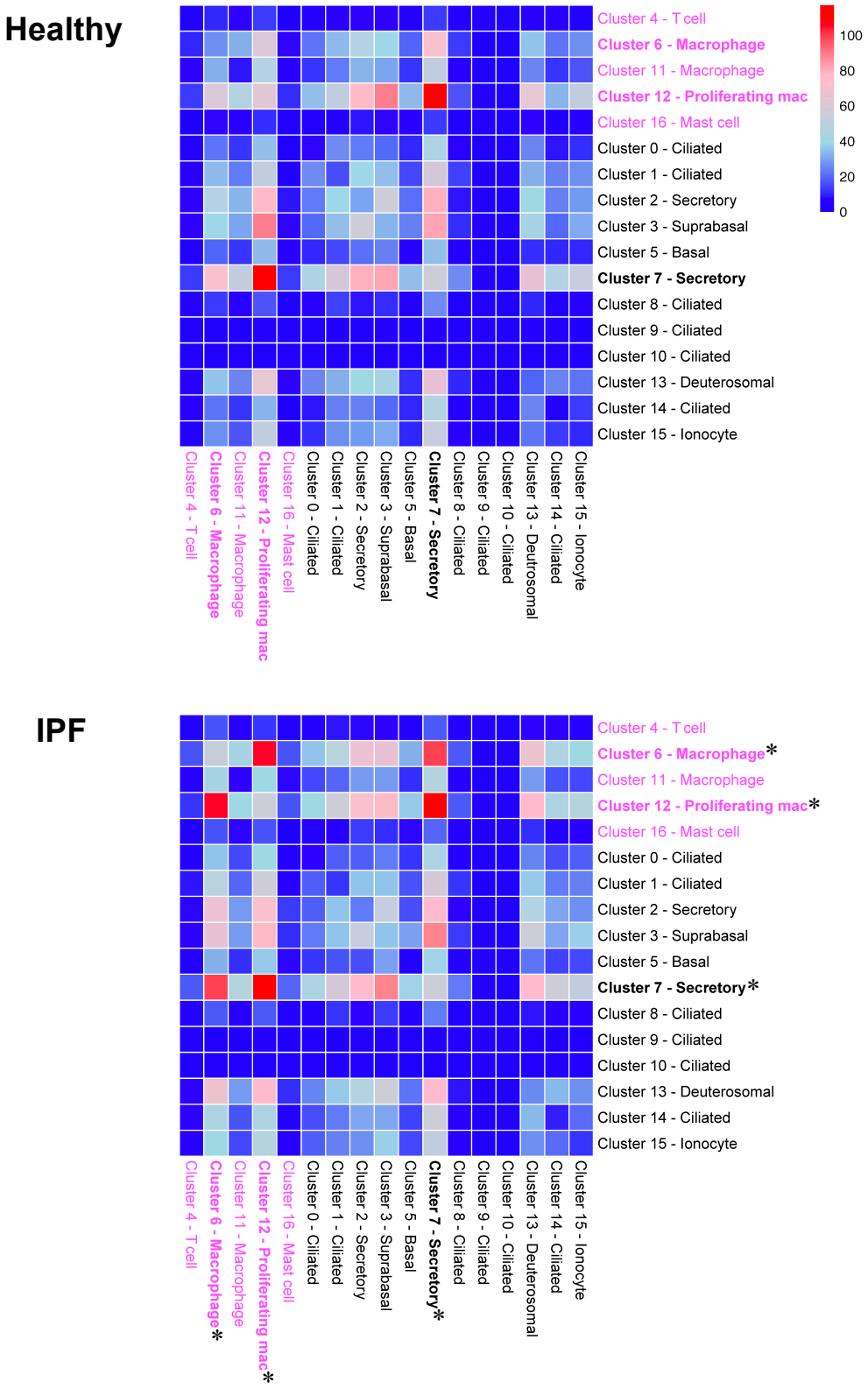


Figure 5.32 Heatmaps displaying the number of interactions between all clusters in healthy and IPF samples. Data plotted are raw counts derived from significant means table output of CellPhone DB. Immune cluster labels coloured magenta. *Clusters of interest. *Plots generated by Probir Chakravarty (The Francis Crick Institute, UK) using CellPhoneDB.*

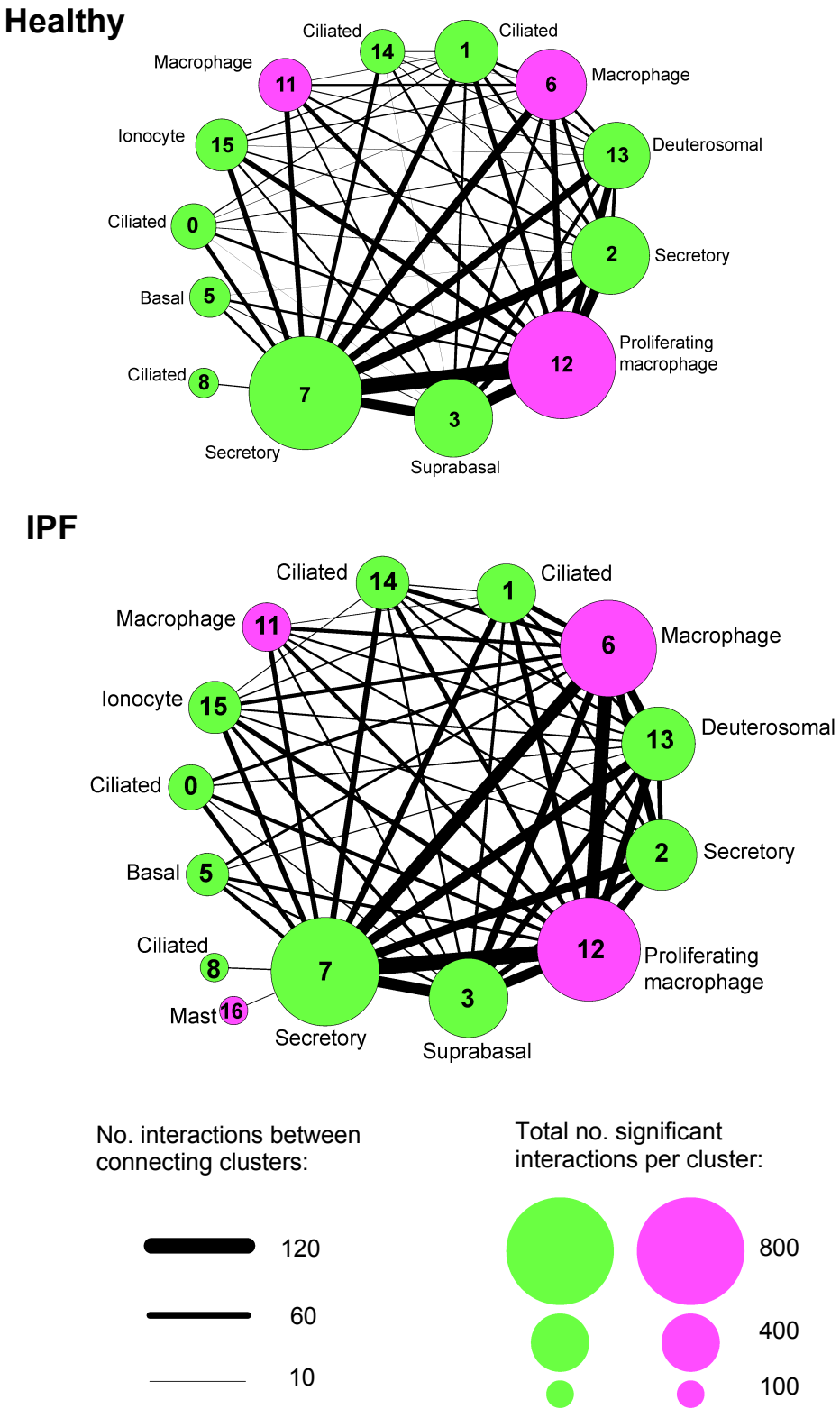


Figure 5.33 Network displaying interacting cell clusters in healthy and IPF samples. Node size is proportional to total number of interactions per cluster. Line thickness is proportional to no. of interactions between connecting clusters. Indicative scales noted below network plots. Epithelial cell nodes in green, immune cell nodes in magenta. Any clusters with fewer than 20 interactions were not included in this network analysis. *Data generated by Probir Chakravarty (The Francis Crick Institute, UK) using CellPhoneDB* and plotted in Cytoscape.

Specific ligand – receptor interactions were then explored between the clusters with the greatest number of interactions. From a list of the top 10% of all interacting ligand-receptor pairs, we focused on interactions showing the greatest change between healthy controls and IPF (Fig. 5.35 – 5.37). CellPhone DB predicted enhanced interactions between *ANXA1* expressed by IPF secretory cells (cluster 7) and formyl peptide receptors, *FPR1* and *FPR3* on proliferating macrophages (cluster 12) in IPF but not healthy controls (Fig. 5.35). Other genes upregulated exclusively by IPF secretory cells with their corresponding ligand or receptor in proliferating macrophages (cluster 12), were receptors implicated in cell adhesion (*CD38, CD99, CD55, CD47)* and secreted proteins involved in mediating the immune response (*SCGB3A1, MIF*).

CellPhoneDB revealed several unique interactions between secretory cells in cluster 7 and macrophages in cluster 6 (Fig. 5.36). Cytokine *TNFSF10* was expressed by IPF secretory cells and predicted to interact with *RIPK* - a regulator of apoptosis and inflammation - expressed by macrophages (cluster 6). In IPF, cluster 7 secretory cells also express a pleiotropic growth factor *MDK* which interacts with *SORL1* expressed by cluster 6 macrophages. Strikingly in IPF glycoprotein fibronectin, *FN1,* expressed by macrophages in cluster 6 and cluster 12 was predicted to interact with integrin receptors $\alpha V\beta 5$*,* $\alpha 3\beta 1$ and $\alpha 2\beta 1$ expressed by secretory cells in cluster 7 (Fig. 5.35 and 5.36).

Within the myeloid lineage, *CCL18* – *CCR1* and *CD74* mediated interactions between macrophages in cluster 6 and proliferating macrophages in 12 were enriched in IPF compared to healthy controls (Fig. 5.37).

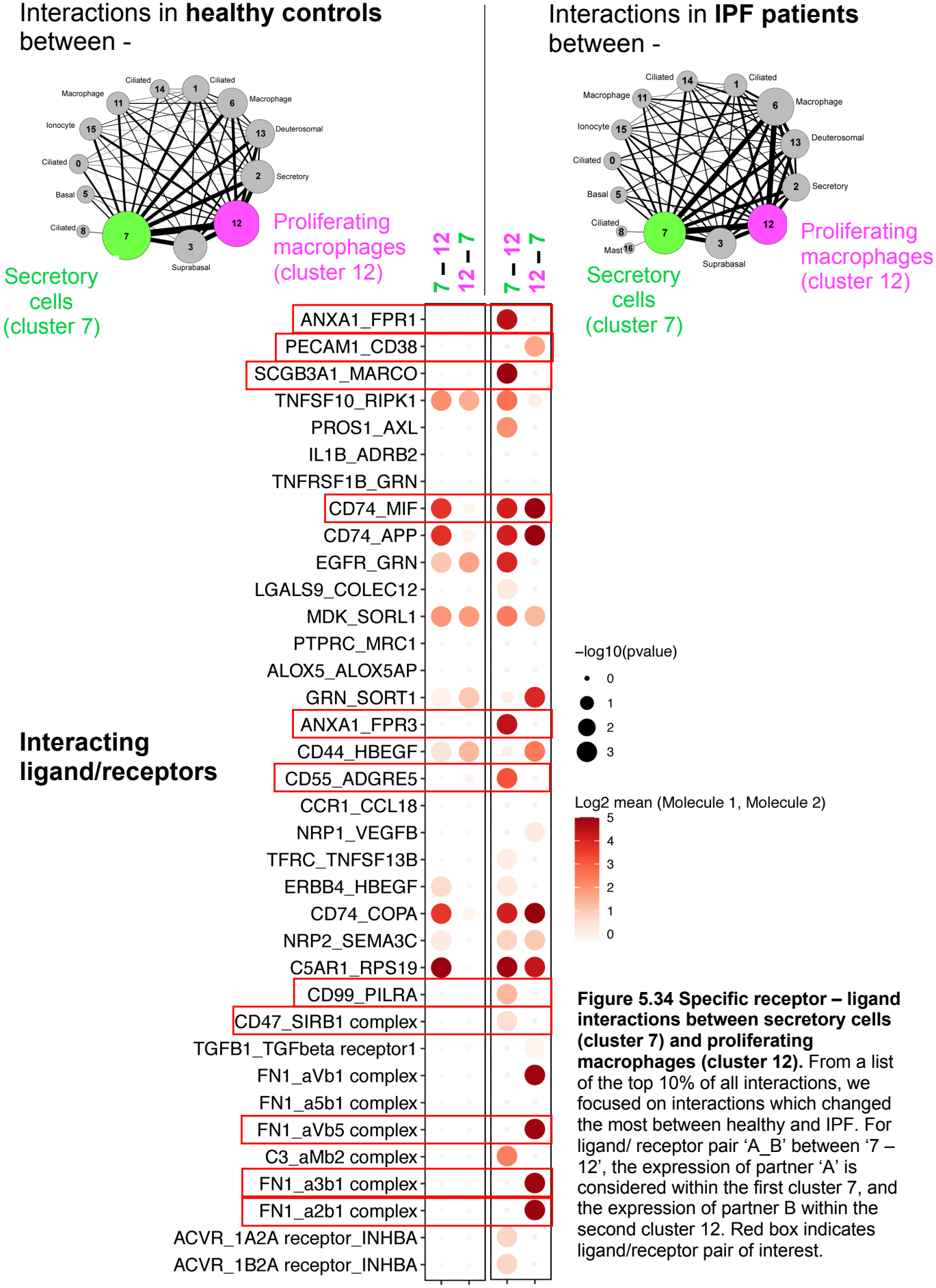


Figure 5.34 Specific receptor – ligand interactions between secretory cells (cluster 7) and proliferating macrophages (cluster 12). From a list of the top 10% of all interactions, we focused on interactions which changed the most between healthy and IPF. For ligand/ receptor pair 'A_B' between '7 – 12', the expression of partner 'A' is considered within the first cluster 7, and the expression of partner B within the second cluster 12. Red box indicates ligand/receptor pair of interest.

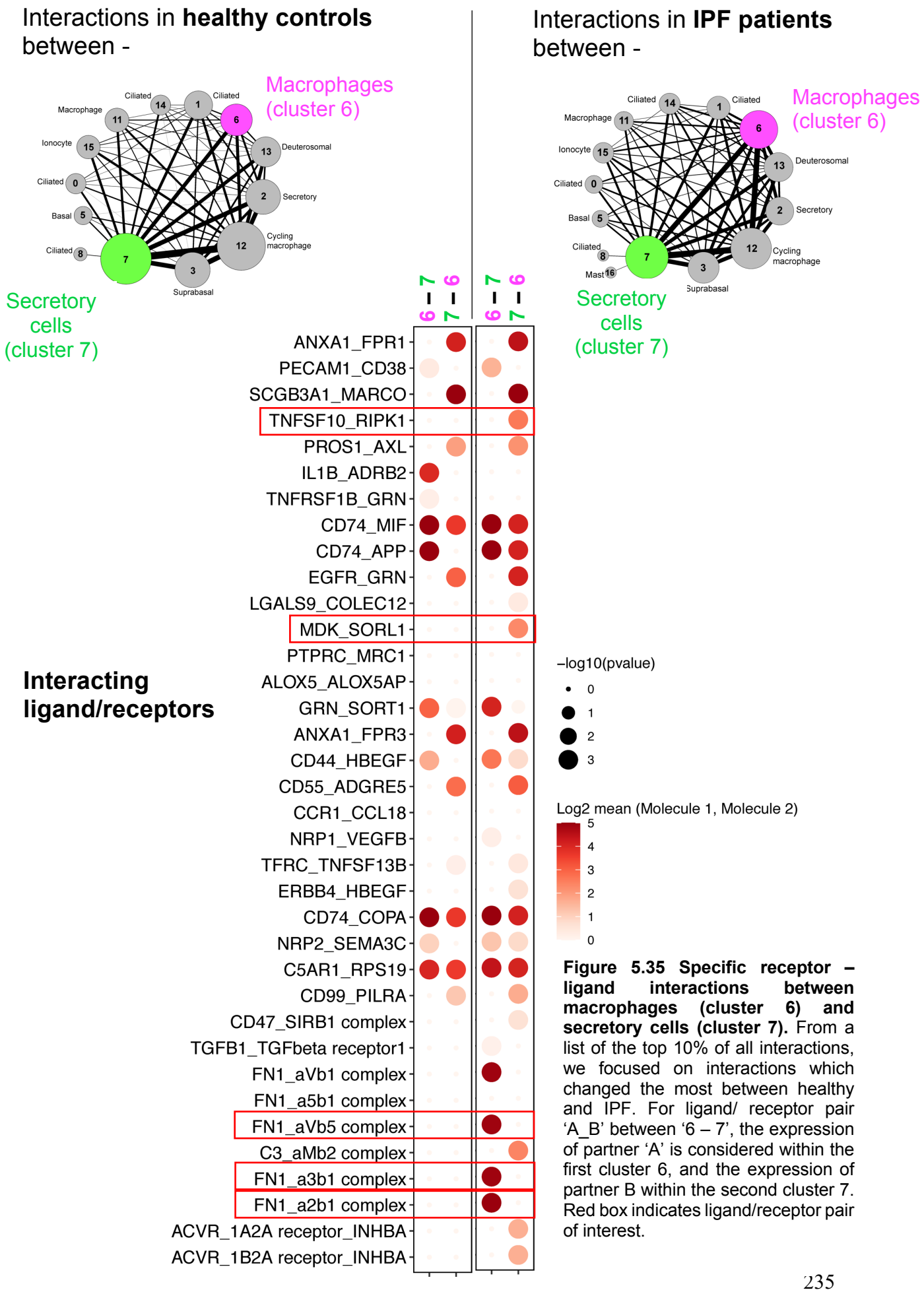
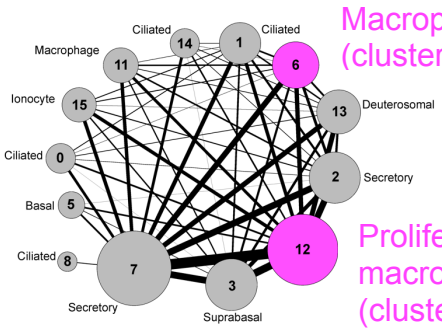


Figure 5.35 Specific receptor – ligand interactions between macrophages (cluster 6) and secretory cells (cluster 7). From a list of the top 10% of all interactions, we focused on interactions which changed the most between healthy and IPF. For ligand/ receptor pair 'A_B' between '6 – 7', the expression of partner 'A' is considered within the first cluster 6, and the expression of partner B within the second cluster 7. Red box indicates ligand/receptor pair of interest.

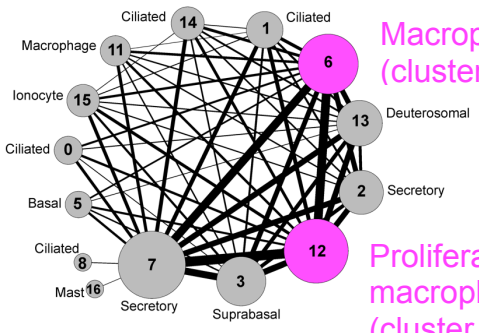
Interactions in **healthy controls** between -



Macrophages (cluster 6)

Proliferating macrophages (cluster 12)

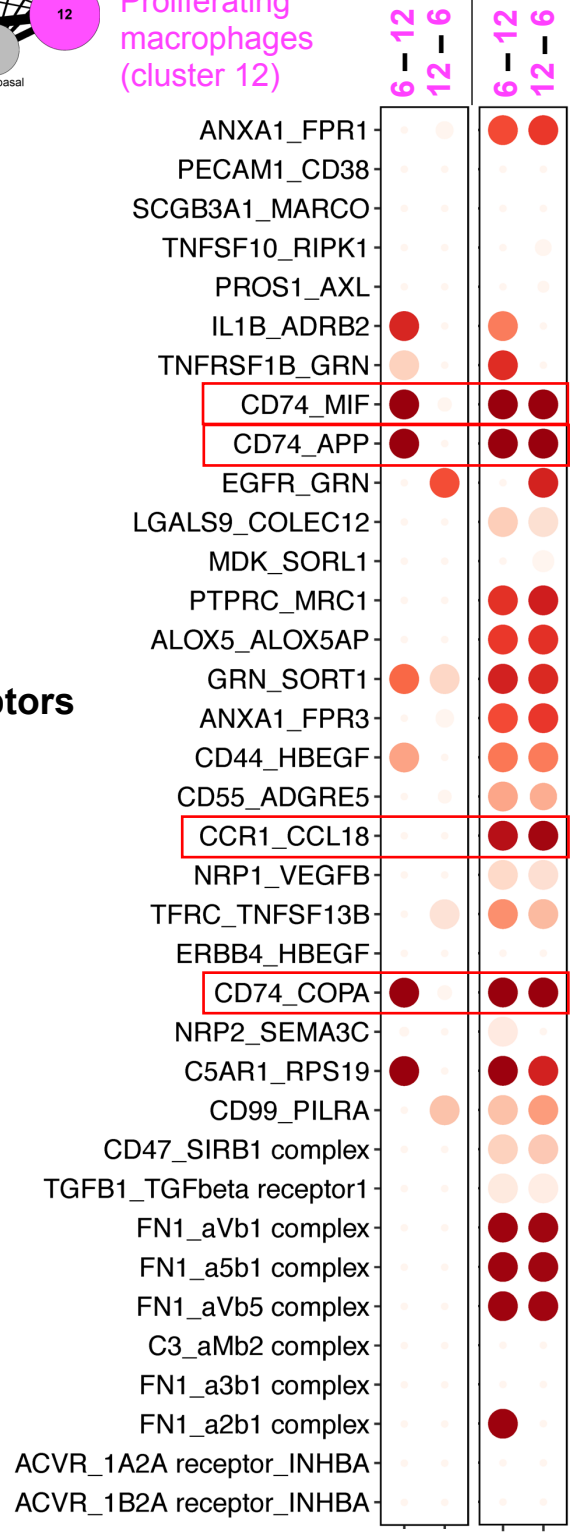
Interactions in **IPF patients** between -



Macrophages (cluster 6)

Proliferating macrophages (cluster 12)

Interacting ligand/ receptors



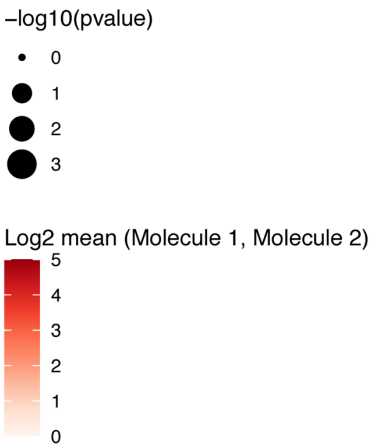


Figure 5.36 Specific receptor – ligand interactions between macrophages (cluster 6) and proliferating macrophages (cluster 12). From a list of the top 10% of all interactions, we focused on interactions which changed the most between healthy and IPF. For ligand/ receptor pair 'A_B' between '6 – 12', the expression of partner 'A' is considered within the first cluster 6, and the expression of partner B within the second cluster 12. Red box indicates ligand/receptor pair of interest.

5.5 Discussion

The data presented in this Chapter provides the first comprehensive single-cell analysis of the cellular landscape of the proximal airway in newly diagnosed cases of IPF. The airway brushing samples that were sequenced were acquired during diagnostic work-up. The data therefore captures an earlier stage in the disease process than previously published scRNA-seq studies that evaluated distal lung tissue samples from advanced cases of IPF[53,145,163]. Single-cell RNA-seq exposed the cellular diversity of the airway niche and revealed transcriptomic changes in individual populations that were not appreciated using a bulk transcriptomic approach in Chapter 4. Notably, there was a significant expansion in macrophage populations in the IPF airway, and an increase in secretory epithelial cell-macrophage crosstalk. On this basis, the Chapter 5 hypothesis was accepted.

Using the 10X Genomics Chromium droplet-based scRNA-seq technology[398], epithelial populations previously reported in single-cell studies of the human airway[168,179] were identified in airway brushings from IPF and healthy donors. This included ionocytes, a rare cell type comprising 0.3% of all epithelial cells in our dataset. The absence of other rare epithelial cell types including pulmonary neuroendocrine cells (PNECs) and tuft cells may be due to the anatomical location of sampling as PNECs are focused at airway branch points[399] and/or the clustering resolution or depth of sequencing[400]. Three transcriptionally distinct macrophage populations, T cells and mast cells were identified in the airway samples. In accordance with findings from Deprez et al.[179], we did not identify neutrophils or eosinophils in the airway brushings. Similarly, in a study analysing BAL fluid from COPD patients, granulocytes (neutrophils, eosinophils) were lost with the 10X Genomics droplet-based approach but not with a well-based sequencing method[395]. A recognised limitation of 10X Genomics' Chromium technology, is the difficulty in sequencing neutrophils and other granulocytes on account of low levels of RNA and/ or high RNase content. It is therefore not possible to reliably comment on airway neutrophils or eosinophils in ours or other scRNA-Seq studies.

Surprisingly T cells were the most abundant immune cells in healthy airway brushings, a finding that was highly reproducible between donors. Deprez et al. comprehensively mapped cell types present in 10 healthy subjects according to sampling location from nasal to distal airways[179]. Macrophages were more abundant than T cells in this study, however methodological differences exist between our studies which may explain the discrepancy in findings. Deprez et al. sequenced biopsies and distal brushings and did not sample the airway location used for our study; the right main bronchus. Secondly, Deprez et al. used a cold active protease from *Bacillus licheniformis* to dissociate airway brushings[401]; we used collagenase D which is active at 37^{o}C.

Overall, the cellular composition of the airway epithelium was comparable between healthy controls and IPF patients. However, the proportion of basal cells was increased in some IPF subjects compared to controls which mirrors the increase in basal cells found in single-cell studies of IPF distal lung tissue[53,145]. This may indicate a process contributing to disease progression is operating in the airways in the early stages of IPF. Closer examination of basal cell heterogeneity in our dataset was achieved by subclustering all epithelial-lineage cells from clusters 3 and 5. This revealed a continuum of differentiation states from multipotent basal cells (subclusters 1 and 2) to 'basal-secretory cells' (subclusters 0, 3, 4, 5 and 6) enriched in both basal and secretory genes, reminiscent of basal cell subsets previously reported[198,199,204]. An intermediate cell population co-expressing club cell marker *SCGB1A1* and basal cell marker *KRT5* - as demonstrated in subcluster 0 of our dataset - was identified in a single-cell analysis of human small airway club cells from healthy volunteers[402]. Ciliated genes were strongly expressed in subcluster 7. There are reports of direct differentiation of basal cells into ciliated cells following sulphur dioxide[403] or polidocanol-induced injury[201] in mice. However, murine epithelial lineage tracing[197] and trajectory analyses using human single-cell data[168,404] – including the work presented in this Chapter - suggest that basal cells first differentiate into secretory cells and subsequently into ciliated cells. Therefore, rather than reflecting an intermediate cell type, subcluster 7 may represent basal – ciliated cell doublets which could conceivably occur if cell dissociation was insufficient during sample processing for sequencing. Interestingly, 'basal-activated cells' (subcluster 2) annotated according to the gene signature described by Carraro et al.[199], were more abundant

in the IPF airway compared to healthy controls, and were enriched in GO biological processes associated with the integrated stress response, apoptosis and proliferation, but also the regulation of fibroblast and T cell activity. Remarkably, a subcluster enriched for T cell marker genes was identified in the IPF unfiltered basal cell analysis. Rather than a highly atypical mixed cell lineage subcluster, it is likely that this represented T cell – basal cell doublets. Stochastic doublets have been shown to occur using 10X Genomics single-cell technology, at a rate of between 0.4 – 7.6% depending on the number of cells loaded (Chromium Single Cell 3' Reagent Kits v3 User Guide). Non-random, cell-cell complexes of biological significance have also been reported, and strategies to distinguish these from singlets - such as imaging flow cytometry - have been evaluated[405]. Therefore, this mixed cell lineage subcluster raises the question of whether physical interactions between basal cells and T cells in the airway wall may be a unique pathological feature of IPF. This could be further investigated by PIC-Seq, a novel approach for cell sorting and sequencing physically interacting cells (PICs)[406]. T cell subsets have been described in IPF tissue explants[407], and implicated in disease pathogenesis, albeit through use of an imperfect mouse model[408].

Pseudotime analysis of epithelial cell lineage clusters from the integrated dataset generally conformed with the currently accepted view of epithelial differentiation dynamics in the human airway[259]. There were no stark differences observed between IPF and healthy controls, which corresponds with data presented in Chapter 3 showing the differentiation of primary airway basal cells harvested from IPF patients was comparable to healthy controls when cultured in an air-liquid interface model. Given the differences noted in the basal cell subclustering analysis, a future trajectory analysis focusing on these alone may unveil more nuanced changes to progenitor cell differentiation in IPF compared to healthy controls.

There were striking compositional changes in the immune cell repertoire in the IPF airway brushings, with a significant expansion in macrophage populations (clusters 6, 11 and 12). Gene markers of macrophage maturity including *APOC1* and *APOE* involved in lipoprotein metabolism*, FABP5* encoding a fatty acid binding protein, and innate immune receptors *MRC1* and *MARCO,* were highest in cluster 6. This cluster

comprised resident airway macrophages and was the population most dramatically increased in IPF compared to healthy controls. A recent study in our laboratory demonstrated that following lung transplant, most airway macrophages in the lung were replaced by circulating monocytes from the recipient[174]. Monocyte associated marker genes including *FCN1, VCAN* and *CD14* did not identify with one specific cluster but were expressed to some extent in all three macrophage clusters. Pseudotime analysis has been used to study gene expression profiles during the maturation of monocytes and monocyte-derived macrophages in airway sputum samples from cystic fibrosis patients[167]. In the future, this approach could be used to better define differentiation trajectory in our macrophage clusters. GO analysis highlighted enrichment of processes shared by clusters 6 and 11, including "MHC class II protein complex assembly", "interleukin 3-mediated signalling" and "wound healing involved in inflammatory response". Macrophage cluster 12 was distinguished by the presence of cell cycle genes, consistent with a proliferative group of resident airspace macrophages previously described in the murine lung during homeostasis and inflammation[409]. Interestingly there was a significant, positive correlation between the proportion of proliferating macrophages in the airway and a score of disease extent in IPF patients; the composite physiological index (CPI).

Aberrations in macrophage populations have been shown in the alveolar and interstitial regions of the IPF lung[53,163,164]. However, this is the first report of changes in airway macrophages closely associated with epithelial cells on proximal airway brushings in newly diagnosed cases of IPF. Several studies have discovered an IPF-specific population of macrophages expressing *SPP1*[53,163,164], a gene encoding osteopontin, which has been previously implicated in pulmonary fibrosis[410]. A profibrotic SPP1- high population was not demonstrable in our dataset, however Morse et al.[164] reported there to be few *SPP1*- expressing macrophages in the IPF upper lobes which would be anatomically closer to the site that airway brushings were derived from for our study.

Comparison of the transcriptomic profiles of individual epithelial and immune cell clusters between IPF and healthy controls revealed marked numbers of differentially expressed genes in macrophage cluster 6 and proliferating macrophage cluster 12.

The genes upregulated in both clusters generally reflect a state of myeloid cell immune activation. More specific processes in resident airway macrophages (cluster 6) were linked to apoptotic cell clearance, phagocytosis and wound healing; and in proliferating macrophages (cluster 12) were monocyte extravasation, antigen presentation and IL3 signalling. Fibronectin, *FN1,* a multifunctional glycoprotein recognised to play important roles in cell-matrix interactions and tissue remodelling[411], was the top upregulated gene in both macrophage clusters in IPF compared with healthy controls. This is consistent with data published over 30 years ago which showed levels of fibronectin in BAL from IPF patients was significantly higher than that of healthy controls[412]. Moreover, IPF macrophages express more fibronectin mRNA transcripts[413] and production of fibronectin by IPF alveolar macrophages in vitro occurs at a rate 20 times the rate of normal macrophages[414].

Within the epithelial cell clusters in IPF and healthy controls, secretory cells (cluster 2) exhibited the greatest number of differentially expressed genes. However, it should be emphasised that this was a modest number in comparison to the macrophage clusters. Genes linked to processes supporting the production of secretory proteins such as the aerobic electron transport chain (providing energy) and protein folding in the ER were upregulated in IPF. Similar processes have been shown to be enriched during the early phases of secretory cell differentiation in a pseudotime analysis of single-cell data from human tracheal samples[198]. The upregulation of genes linked to antigen presentation machinery is noteworthy; MHC class II molecule HLA-DR has previously been detected on nasal epithelial cells[415], and both class I and II molecules have been shown to be expressed on bronchial epithelial cells[416,417]. Epithelial cells are critically important in orchestrating immune responses[52] and this differential expression analysis suggests changes in the immune function of secretory cells in IPF.

There was higher expression of monocyte chemoattract, *CXCL17,* in airway secretory cells (clusters 2 and 7) in IPF patients compared with controls. In a previous report, CXCL17 was detectable in bronchial epithelial cells by immunohistochemistry, exhibited antimicrobial activity *in vitro*, and was significantly elevated in the BAL of IPF patients[418]. To explore the complex cellular cross- talk occurring in the airway, communicating pairs of cells were predicted by the expression of interacting ligands

and receptors, using CellPhone DB[260,261]. In IPF there were increased interactions between Annexin 1, *ANXA1* expressed by secretory cells (cluster 7) and formyl peptide receptors 1 (*FPR1*) and 3 (*FPR1*) on proliferating macrophages (cluster 12). Annexin 1 has been shown to elicit chemotaxis of human peripheral blood monocytes in vitro[419]. Underscoring the importance of this pathway, interactions between *ANXA1* and *FPR1* and *FPR3* were also increased between macrophages (cluster 6) and proliferating macrophages (cluster 12) in IPF. A study using mice lacking the Club cell secretory protein (CSSP) gene (CCSP-/-), showed secretory cells play an important role in post-translational modification of the ANXA1 protein, which was localised within ciliated cells and airway macrophages from BAL[420]. This work supports the concept of secretory cell crosstalk with other epithelial and immune cell types in the airway as suggested by our study. Similarly, interactions were found between secretoglobulin, *SCGB3A1* and macrophage scavenger receptor with collagenous structure (*MARCO)* expressed by proliferating macrophages. These findings support data showing another member of the secretoglobulin family UGRP1/ SCGB3A2 is also a ligand for MARCO on alveolar macrophages[421].

There were increased interactions between complement decay accelerating factor, *CD55*[402], expressed by secretory cells (cluster 7), and adhesion G protein-coupled receptor E5, *ADGRE5 / CD97* expressed by proliferating macrophages (cluster 12). *ADGRE5 / CD97* has been identified on alveolar macrophages by IHC[422], and plays an important role in leucocyte trafficking[423]. A number of interactions were also predicted to be mediated by expression of CD74 (MHC class-II-associated invariant chain II) on IPF proliferating macrophages (cluster 12), such as that with proinflammatory 'macrophage migration inhibitory factor', *MIF*[424,425] expressed by secretory cells. Remarkably in IPF, there was a shift in the pattern of interactions between secretory cells (cluster 7) and proliferating macrophages (cluster 12). This was in line with that observed between secretory cells (cluster 7) and macrophages (cluster 6) in both the healthy and IPF airways. Increased interactions were predicted between fibronectin, *FN1* expressed by macrophages (clusters 6 and 12) and integrin receptor complexes expressed by secretory cells. The integrin repertoire expressed by airway epithelial cells modulates multiple processes including cell adhesion to extracellular matrix, signalling, proliferation and wound repair[426]. Taken together, it is

clear that interactions exist between airway secretory cells and macrophage populations in patients with newly diagnosed IPF, that are not present in healthy controls. These interactions may contribute to the progression of disease.

Prominent interactions between transcriptionally distinct populations of mature (cluster 6) and proliferating (cluster 12) airway macrophages were predicted in IPF such as between the CC-chemokine ligand 18, *CCL18* and *CCR1.* This interaction is pertinent because high concentrations of chemokine CCL18 in bronchoalveolar lavage fluid, has been reported to be predictive of mortality in IPF[427]. Increased production of CCL18 by IPF alveolar macrophages upregulated collagen I production by *in vitro* cultured normal lung fibroblasts[428].

The CellPhoneDB data provides valuable insight into the complex crosstalk occurring between ligands and their receptors, and cell types in the IPF airway (Fig. 5.38). Importantly, this is a computational prediction based on single-cell gene expression data. To translate these findings into meaningful therapeutic targets, it will be important in future studies to validate the presence of the specific interactions cited by staining airway samples using immunohistochemistry/ immunofluorescence, and additionally by performing functional co-culture experiments.

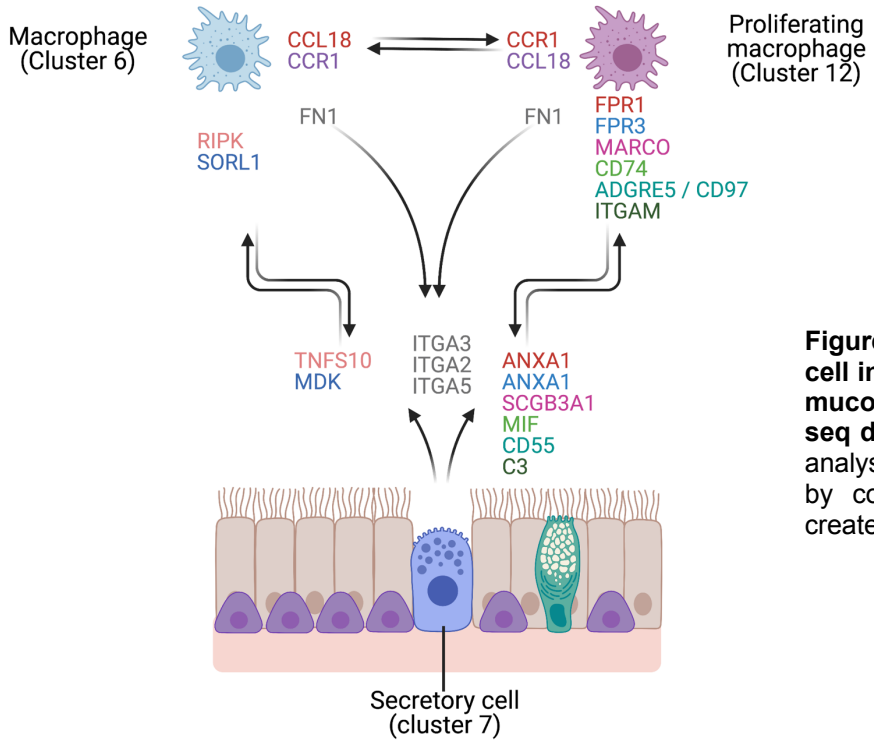


Figure 5.37 Summary of key cell-cell interactions in the IPF airway mucosa predicted from scRNA-seq data. Data from CellPhone DB analysis. Interacting pairs indicated by corresponding colours. Figure created with BioRender.com.

The data in this chapter was strengthened by a consistent approach to airway sampling and use of a validated protocol for processing human bronchial brushings[168]. During the analyses, emphasis was placed on evaluating the contribution of individual donors to the clusters obtained in the integrated IPF and control datasets. This was critically important in determining if changes in a cell cluster were truly driven by the disease process or by an individual subject. There were no cluster differences influenced by a single donor in the overall integrated dataset, however some donor-driven differences were revealed by subclustering basal cells. For the analysis, integration of single-cell datasets from individual subjects was performed in Seurat[252], but other computational tools exist to perform this important step. For example, 'Harmony' is an integration algorithm reported to account for experimental and biological variation when integrating datasets across donors, whilst preserving fine-grained subpopulations in the data[429]. Use of 'Harmony' or similar computational tools may improve data integration and reduce donor-related effects in the sub-clustering analysis.

There were several limitations to the work described. The first was the gender and age imbalance in our recruited cohort of subjects; the IPF group comprised older males after exclusion of a female IPF subject on account of obtaining too few cells. A study investigating sex-determined differences in the human transcriptome across 44 tissue sources as part of the Genotype-Tissue Expression project, found 37% of all genes showed sex-bias in at least 1 tissue[430]. At present we do not possess the bioinformatic tools to correct for the influence of gender in single-cell studies therefore we have engaged in a collaboration with Martijn Nawijn and colleagues (University Medical Centre Groningen, Netherlands) who will be providing scRNA-seq data from six male, age-matched healthy controls to integrate with our dataset and therefore present a totally gender matched datset.

Whilst scRNA-seq technology provides transcriptomic data on thousands of cells from a subject, this comes at the expense of lower sequencing depth which may miss low expression genes relevant to disease pathogenesis[400]. It is possible that this is the reason that bulk RNA-seq which I described earlier in Chapter 4 (affording greater sequencing depth) showed that *KRT13* was significantly upregulated in IPF brushings,

yet very few cells expressing this gene were detectable in our scRNA-seq dataset. In the future, parallel scRNA-seq and bulk RNA-seq from the same individual and brushings sample may help remedy this issue.

There is currently no consensus in the field of single-cell transcriptomics on marker genes that define particular cell types and cell states. In line with Deprez et al.[179] it was not possible to sub-categorise secretory cells into club or goblet cells on the basis of published marker genes, since *SCGB1A1* (club) and *MUC5AC* (goblet) were expressed in both clusters. Similarly, the T cell cluster in our dataset also shared some genes associated with NK cell function. To validate cell types assigned to clusters, a combination of manual annotation from published literature, GSEA using marker genes from two single-cell studies of the human lung[163,179] and an automated tool were used.

Single-cell transcriptomics of IPF airway brushings has provided a unique insight into cellular changes that occur at a relatively early stage during the evolution of the disease. These changes were identified in the proximal airway, an anatomical site distant to the main pathological lesion; the fibrotic foci in the distal lung. Our data suggest that airway macrophage populations closely associated with the epithelium are expanded in IPF, and in part this may be orchestrated through direct interactions with secretory epithelial cells. One striking finding was the predicted interaction between extracellular matrix protein fibronectin, expressed by airway macrophage populations, and secretory epithelial cells expressing integrin receptors. To what extent changes occurring in the cellular microenvironment - specifically the extracellular matrix - influence airway epithelial cell phenotype and function in IPF is unknown and will be explored in the next chapter.

CHAPTER 6

Basal cell- matrix interactions in IPF

6 Basal cell - matrix interactions in IPF

6.1 Background

Basal cells are multipotent airway stem cells, characterised by expression of transformation-related protein 63 (TP63) and cytokeratin 5 (KRT5), that are anchored to the basement membrane[194]. They play a crucial role in homeostatic turnover of the epithelium and regeneration following injury. Basal cells have the capacity to differentiate into club/ secretory cells, and rare epithelial cell subsets (tuft cells, PNECs and ionocytes)[198,200,259]. Diseases of the respiratory tract including chronic obstructive pulmonary disease (COPD)[300] and chronic allergic rhinosinusitis (CRS)[404] have been associated with basal cell dysfunction.

Accompanying the dramatic structural remodelling that occurs in the distal lung in IPF, is the intriguing emergence of a population of cells with phenotypic characteristics of airway basal cells[36,187,205,431]. This has now also been confirmed by single-cell RNA-sequencing analysis of cell populations in end-stage IPF lung tissue specimens[53,145,189]. Colonisation of the damaged alveolar compartment by ΔNp63-positive basal cells in IPF has been termed “bronchiolisation”[188]. The origin of these so-called “metaplastic” airway basal cells and their contribution to the pathobiology of IPF is currently unknown. Research into this important question is hindered to some extent by the tools available to study this process in human lungs. Murine studies using genetic lineage tracing, demonstrate that injury to the alveolar epithelium from influenza virus infection, triggers the proliferation and migration of a Krt5+ /Trp63+ progenitor cell population into damaged areas of lung parenchyma forming “Krt5+ pods”[104,105,206,207]. Notably, these Krt5+ cells do not arise from club cells (SCGB1A1+) or AT2 cells (SFTPC+), but rather from SOX2 lineage-negative progenitors in the distal airways[207]. In mice, permanent cystic structures in the alveolar parenchyma resembling IPF ‘honeycomb cysts’ develop and are lined by KRT5 basal cells[104]. Interestingly in a bleomycin model of pulmonary fibrosis, mesenchymal stromal cells expressing *GLI Family Zinc Finger 1* (*Gli1)* - a transcription factor activated by the sonic hedgehog signalling pathway – were able to drive KRT5+ metaplastic differentiation of migrating SOX2+ airway progenitors[432]. The contribution of distal

airway stem cell populations to alveolar regeneration remains to be elucidated and is an area of active research[99].

The lung extracellular matrix (ECM) provides a physical support scaffold for the adhesion of resident cells but importantly also hosts a complex array of biochemical and mechanical signals which influence cell behaviour[433]. The consequential changes in ECM that occur during fibrosis also catalyse the pathobiology of the disease[434]. Increased ECM stiffness activates lung fibroblasts through YAP and TAZ signalling[435], and promotes myofibroblast invasion through basement membrane matrices[436]. The process of cell migration can also be modulated by other ECM properties including molecular composition, topology, cross-linking and porosity[437]. Basal cells spread, migrate, proliferate and differentiate as part of a coordinated wound repair response in the airways[438]. It is not known if human KRT5 basal cells possess the capacity to migrate as widely as demonstrated in the mouse model by Vaughan et al.[104]. It is also unclear how KRT5+ basal cells interact with and navigate the complex 3D ECM microenvironment of the fibrotic lung.

6.2 **Hypothesis**

Changes in the extracellular matrix (ECM) influence basal cell migration and redistribution in the fibrotic lung.

6.3 **Aims**

1. To establish the colocalization of KRT5+ basal cells with fibrotic ECM in IPF lung tissue
2. To compare regional collagen organisation in the IPF distal lung by second harmonic generation (SHG) microscopy
3. To assess basal cell migratory dynamics on different ECM substrates
4. To identify compositional differences between control and fibrotic lung matrices that may influence basal cell migration

6.4 **Results**

6.4.1 KRT5+ basal cells populate parenchymal lung tissue and co-localise with collagen-rich fibrotic matrix IPF

An expansion of KRT5+ expressing cells have been reported in the distal alveolar regions and honeycomb cysts in IPF[36,205]. Immunofluorescence staining of FFPE lung tissue sections from eight IPF patients with a typical UIP histological pattern, were compared to normal lung tissue obtained peripheral to resected lung cancer from five control patients. Demographic information pertaining to these samples is outlined in Appendix, Table 7.2. In control lung tissue KRT5+ basal cells lined the basement membrane of small airways but were absent from alveolar regions (Fig. 6.1 A and B). In the IPF lung, KRT5+ cells could also be seen associated with the basement membrane of the ciliated epithelium of the distal bronchioles (Fig. 6.1 C and D). However, in distal IPF lung tissue sections, loss of the normal alveolar architecture and marked remodelling was accompanied by abundant KRT5+ cells that were distributed throughout the distorted tissue in various configurations (Fig. 6.1 E – H). KRT5+ cells were grouped in clusters, dispersed as single cells within fibrotic areas or lining honeycomb cysts either in a simple squamous or pseudostratified organisation. At high-resolution, these 'ectopic' KRT5+ cells appeared morphologically comparable to those KRT5+ cells lining the airways (Fig. 6.1 G and H) but adopted a more flattened morphology when colonising damaged alveolar regions (shown later in Fig. 6.6 A).

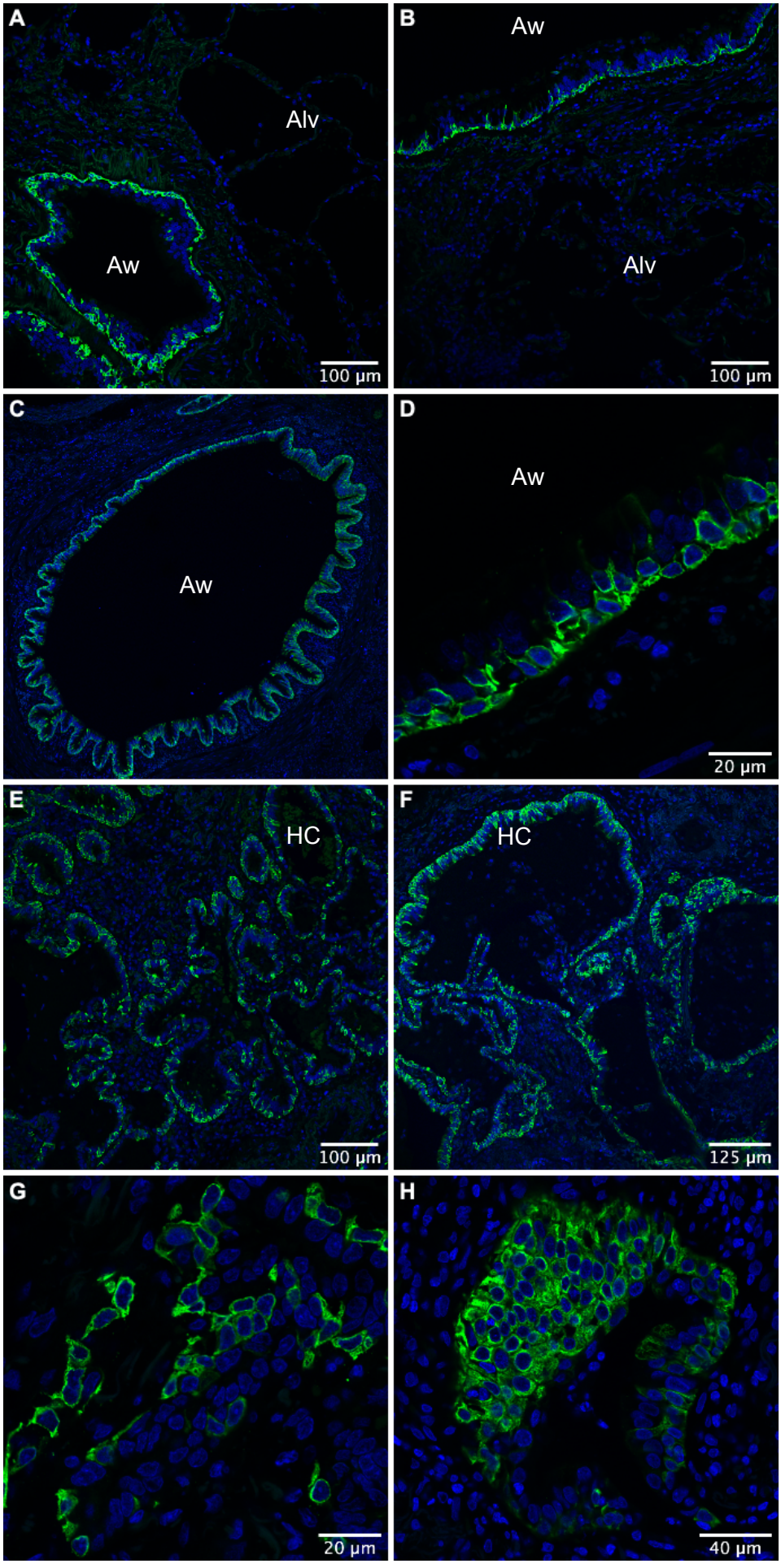

A
Alv
Aw
100 µm
B
Aw
Alv
100 µm
C
Aw
D
Aw
20 µm
E
HC
100 µm
F
HC
125 µm
G
20 µm
H
40 µm

Figure 6.1. Immunofluorescence microscopy of KRT5+ basal cells in distal lung. FFPE sections of distal lung tissue from non-fibrotic controls and IPF patients immunostained for basal cell marker KRT5+ (*green*) and DAPI (*blue*). Representative images taken from control donor 'NORM22' and IPF donor 'IPF06'. In control lung tissue KRT5+ cells are found restricted to the distal airways (Aw) and are absent from the alveolar tissue (Alv) **(A and B)**. In IPF lung tissue, KRT5+ cells are again seen in the distal airways in a typical configuration **(C and D)** but are also located in honeycomb cysts (HC) **(E and F)**, and fibrotic parenchymal tissue **(G and H)**. Images were captured using a Leica SP5 inverted confocal microscope.

6.4.2 Regional collagen organisation in the healthy and fibrotic lung

A unique property of non-centrosymmetric structures such as type I collagen, is the emission of a second harmonic generation (SHG) signal[439]. SHG imaging, a nonlinear microscopy technique has emerged a powerful tool for assessing collagen organisation in the lung and other tissues in health and disease states[439,440]. Quantitative analysis of images acquired by second-harmonic generation (SHG) microscopy was used to better characterise regional differences in fibrillar collagen in normal and IPF lung. Equivalent regions of IPF and control distal lung were compared, with the addition of fibrotic and honeycomb cyst (HC) regions that were clearly exclusive to IPF (Fig. 6.2 A and B). KRT5+ basal cells in different configurations (as described in Section 6.4.1) co-localised with high SHG signal in fibrotic regions and honeycomb cysts in the IPF lung (Fig. 6.2 A - C).

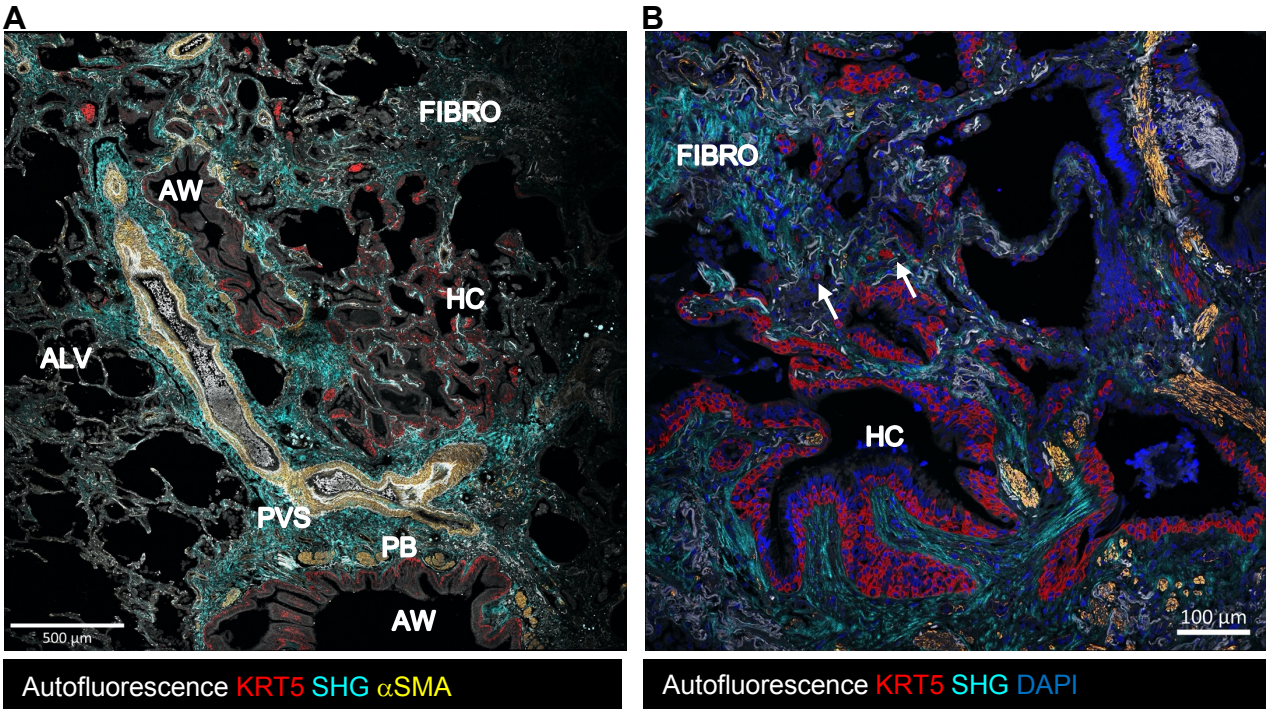


Figure 6.2 Evaluation of regional collagen organisation in IPF and control lung using second harmonic generation (SHG) imaging. Images of distal lung acquired by Dr Fred Fercoq (Cancer Research UK Beatson Institute, Glasgow, UK) using a Zeiss 880 2-photon microscope with spectral detection. Linear unmixing of autofluorescence and SHG signals. Representative images of IPF distal lung (A – C) highlighting regions analysed; airway (AW), peribronchiolar (PB), perivascular (PVS), alveolar (ALV), honeycomb cyst (HC), fibrotic (FIBRO). Arrows marking single KRT5+ basal cells within matrix. Red blood cells - RBCs.

In control lung tissue, there was a progressive decline in SHG signal and collagen density from peribronchiolar region to alveolar region (Fig. 6.3 A and B). In IPF lung tissue there was a similar trend with peribronchiolar collagen density higher than that of normal appearing alveoli, however collagen density was generally higher in each region compared to control (non-significant). IPF fibrotic and honeycomb cyst (HC) regions were characterised by higher collagen density than in the alveolar regions. Collagen fibres had greater SHG mean intensity in fibrotic areas compared to alveolar and HC regions in IPF. Fibre orientation in the fibrotic regions, measured by coherency, was significantly lower than the peribronchiolar region, and comparable to the perivascular region (Fig. 6.3 C).

Image texture analysis using a grey level co-occurrence matrix (GLCM) demonstrated significant inter-regional differences in collagen organisation. Entropy, a statistical measurement of the degree of randomness in the image, was increased in IPF fibrotic areas compared to alveolar and HC regions (Fig. 6.3 D). Collagen in the perivascular region of normal and IPF lung tissue exhibited significantly higher entropy than that of the alveolar region. The perivascular region also demonstrated higher contrast - a measure of local variations in the image - than the alveolar region (Fig. 6.3 E). Correlation between neighbouring pixels was significantly higher in normal compared to IPF alveolar regions (Fig. 6.3 F). Compared to the HC regions, collagen in fibrotic regions demonstrated lower inverse difference moment (IDM), reflecting lower homogeneity in an image, and lower angular second moment (ASM), indicative of lower uniformity of the image (Fig. 6.3 G and H). GLCM homogeneity and uniformity image texture measurements were higher in the alveolar regions, but this may be more reflective of black pixels in aerated regions rather than collagen. No differences were found in other GLCM parameters between normal and IPF lung tissue in the equivalent regions; peribronchiolar, perivascular and alveolar.

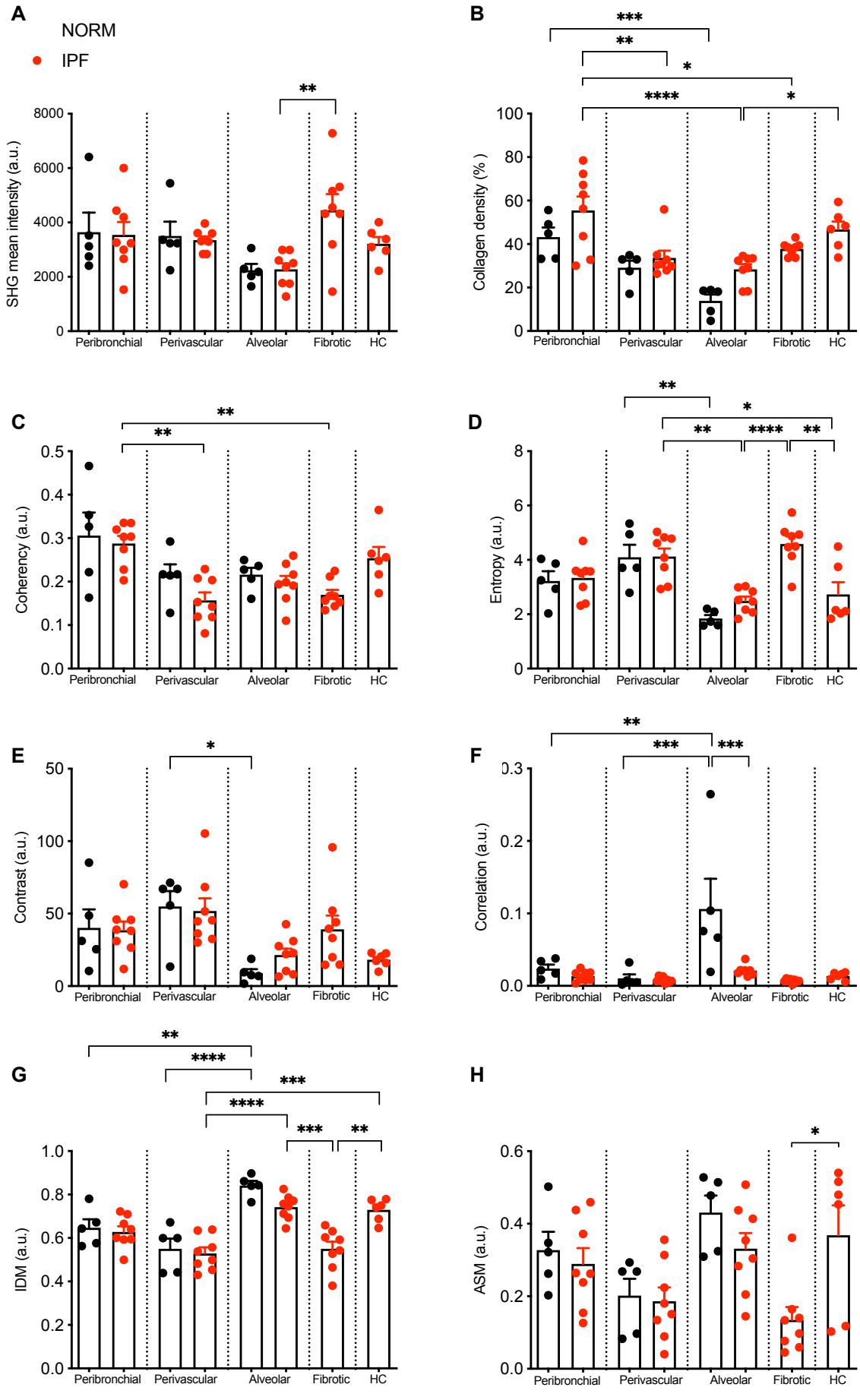

A
NORM
IPF
SHG mean intensity (a.u.)
8000
6000
4000
2000
0
**
Peribronchial
Perivascular
Alveolar
Fibrotic
HC
B
***
**
*
****
*
Collagen density (%)
100
80
60
40
20
0
C
**
**
Coherency (a.u.)
0.5
0.4
0.3
0.2
0.1
0.0
D
**
*
**
****
**
Entropy (a.u.)
8
6
4
2
0
E
*
Contrast (a.u.)
50
100
50
0
F
**
***
***
Correlation (a.u.)
0.3
0.2
0.1
0.0
G
**
****
***
****
***
**
IDM (a.u.)
1.0
0.8
0.6
0.4
0.2
0.0
H
*
ASM (a.u.)
0.6
0.4
0.2
0.0

Figure 6.3 Quantitative analysis of SHG images of distal lung tissue from controls and IPF patients. In each region data is plotted for **(A)** SHG mean signal intensity of collagen fibres **(B)** collagen density (% collagen coverage of area analysed) **(C)** fibre coherency (OrientationJ analysis), and GLCM analysis **(D)** entropy, **(E)** contrast, **(F**) correlation, **(G)** inverse difference moment (IDM), and **(H)** angular second moment (ASM). Data plotted as mean + SEM. Each data point represents the average of 2 – 6 images per area per subject. Ordinary one-way ANOVA with Tukey's multiple comparisons test ****P<0.0001, ***P<0.001, **P<0.01, *p<0.05. Raw data provided by Dr Fred Fercoq (Cancer Research UK Beatson Institute, Glasgow, UK).

Interestingly, principal component analysis (PCA) of regional collagen structural characteristics revealed clustering of the HC and peribronchiolar regions, driven by increased fibre coherency and density (Fig. 6.4 A). IPF fibrotic regions cluster with perivascular regions, characterised by increased entropy. A scree plot showed that the first 3 principal components explained 80.3% of the variance in the dataset (Fig. 6.4 B). The contribution of variables to each principal component (dimension) was highlighted using a corr plot (Fig. 6.4 C), which showed IDM and entropy contribute the most to dimension 1 (dim1), and coherency and density to dimension 2 (dim2).

Differences in collagen structure as assessed by SHG were correlated with KRT5+ basal cell numbers and arrangement, identified by IF staining. In the IPF lung, there were quantifiable regional differences in the abundance of KRT5+ basal cells (Fig. 6.5 A), and a significant positive correlation of KRT5+ basal cell numbers with collagen density (Fig. 6.5 B), and coherency (Fig. 6.5 C) was found.

Interestingly, in the fibrotic, remodelled alveolar region of the IPF lung there was a close spatial relationship between KRT5+ basal cells, myofibroblasts expressing a-SMA and increased SHG signal (Fig. 6.6 A and B).

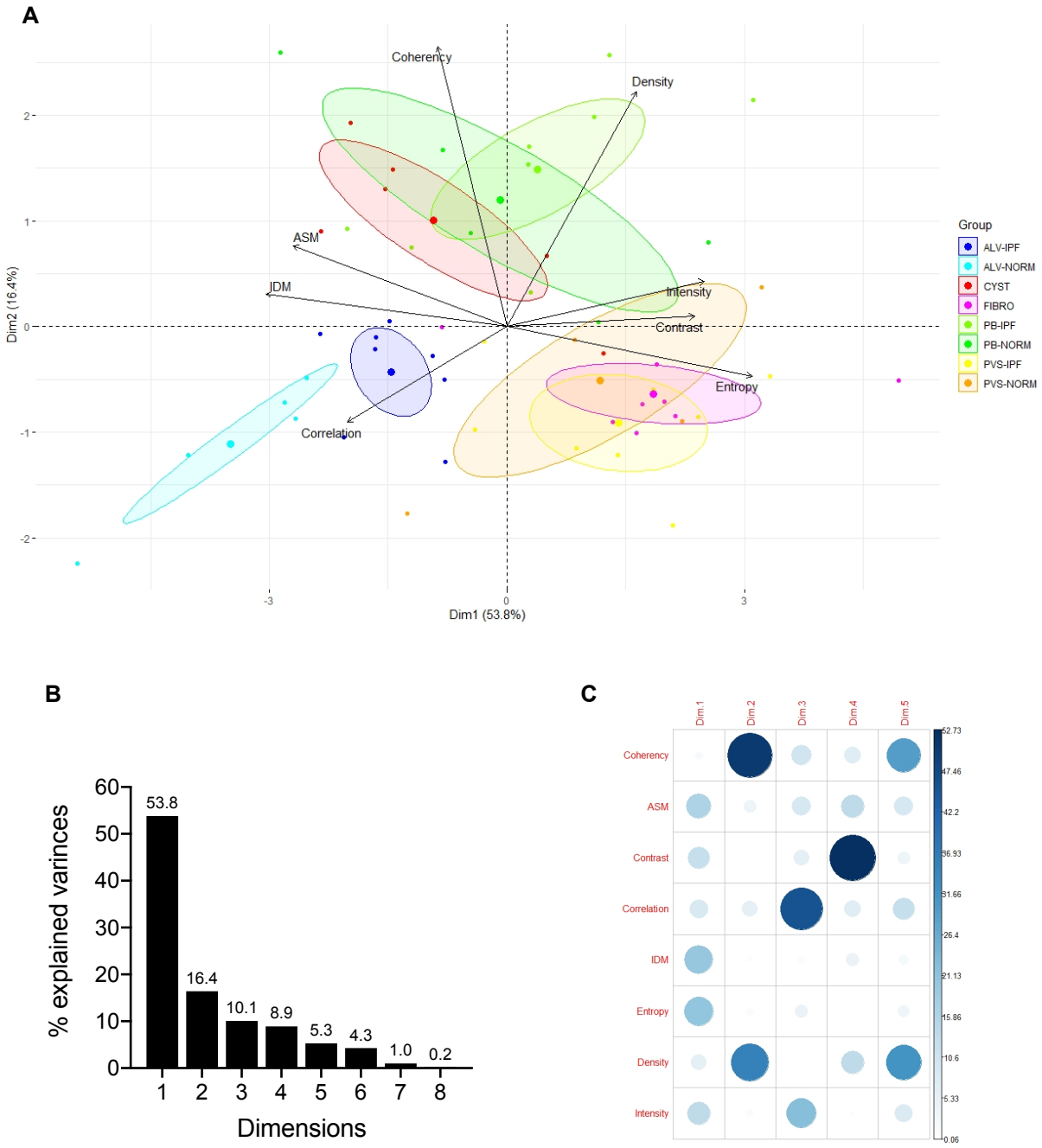


Figure 6.4 Principal component analysis of quantitative image analysis of IPF and control distal lung. **(A)** Principal component analysis (PCA) of collagen intensity, density, orientation and GLCM parameters. Generated by Dr Fred Fercoq (Cancer Research UK, Beatson Institute) using FactoMineR and factorextra packages in R software. **(B)** Scree plot of % variances explained by each component (dimension) **(C)** Corr plot of variables contributing to each dimension. Data provided by Dr Fred Fercoq (Cancer Research UK Beatson Institute, Glasgow, UK).

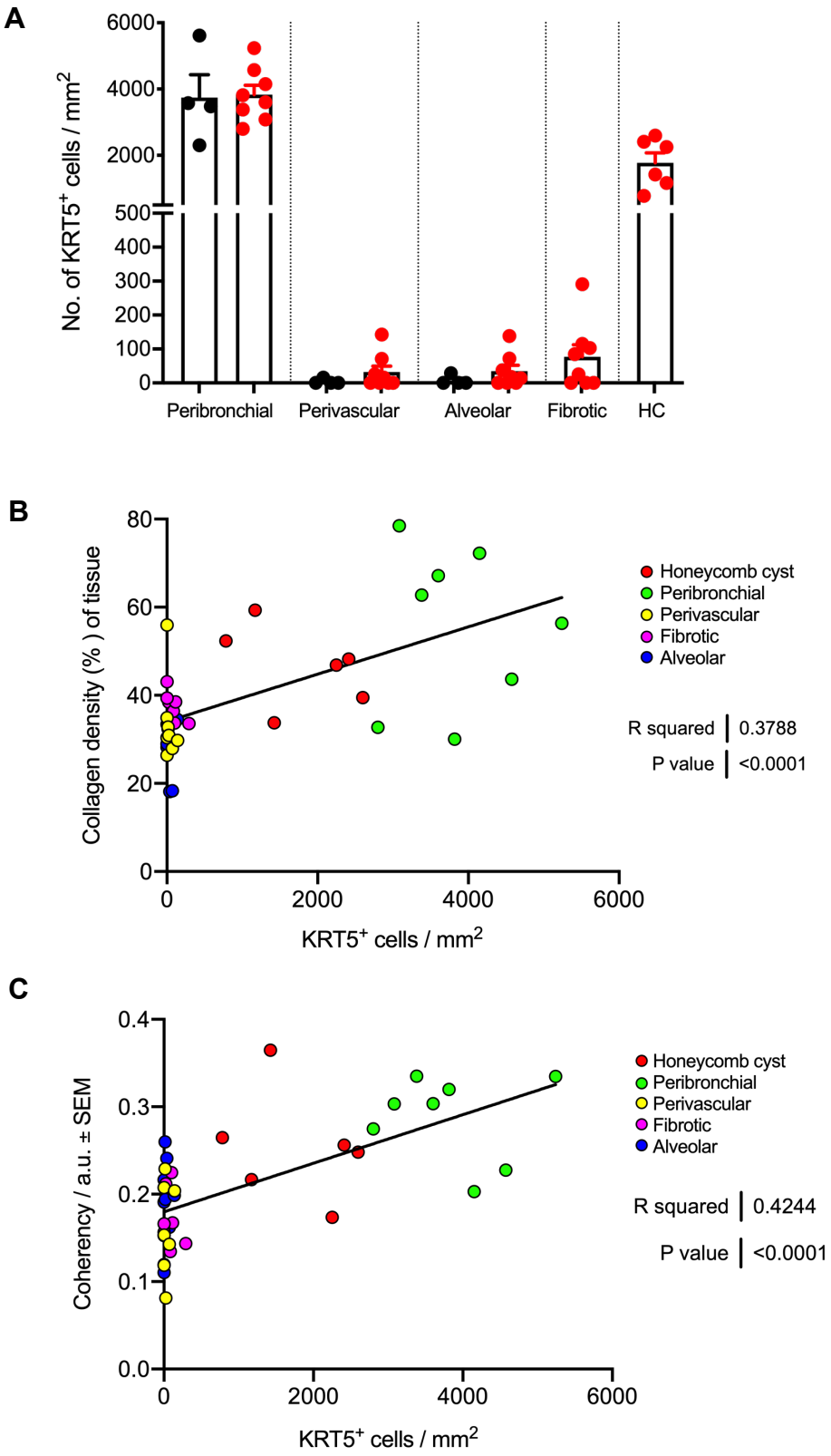
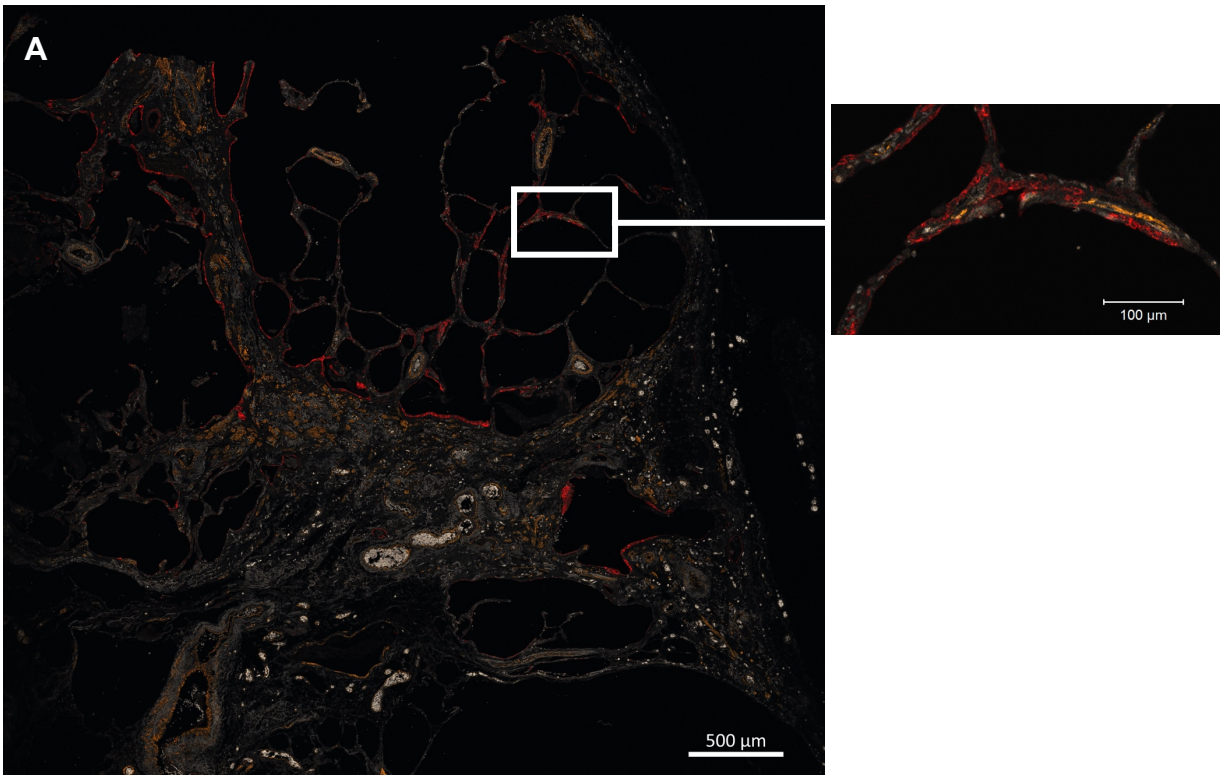


Figure 6.5 KRT5+ basal cell distribution correlates with regional collagen organisation. **(A)** KRT5+ cells counted per image area. Correlation analysis of **(B)** collagen density and KRT5+ cell numbers and **(C)** collagen orientation and KRT5+ numbers. Raw data provided by Dr Fred Fercoq (Cancer Research UK Beatson Institute, Glasgow, UK).



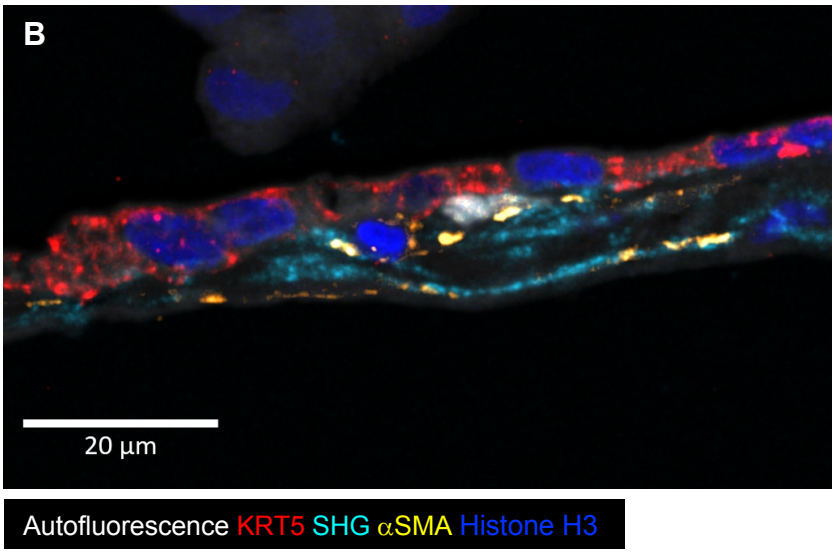


Figure 6.6 KRT5+ basal cells co-localise with fibroblasts in the fibrotic niche. (A) Overview of fibrotic, remodelled alveolar region of IPF lung with high-resolution image (white box) showing co-localisation of KRT5+ basal cells (*red*) with α-SMA+ fibroblasts (*yellow*). **(B)** High resolution image of fibrotic alveolar region of IPF lung showing co-localisation of KRT+ basal cells, α-SMA+ fibroblasts and SHG signal (*cyan*).

6.4.3 Extracellular matrix proteins collagen I and collagen III promote increased motility of healthy and IPF airway basal cells.

Mechanisms directing KRT5+ cell redistribution in the human fibrotic lung have not yet been elucidated. Section 6.4.2 demonstrated close spatial localisation of KRT5+ cells with the surrounding extracellular matrix. Structural cues in the pulmonary microenvironment are critically important in determining the kinetics of cell migration[263,441]. Therefore, the influence of different ECM proteins on basal cell movement was tested in a two-dimensional (2D) cell migration assay. Flow cytometry confirmed all epithelial cells (EpCAM+) in submerged culture at passage 3 expressed basal cell marker KRT5, 95% expressed cytokeratin 14 and 87% expressed TP63 (Fig. 6.7 A and B). Consistent with this, immunofluorescence microscopy demonstrated all cultured cells in the field of view stained positive for epithelial basal cell cytoskeletal marker, KRT5 (Fig. 6.7 C). Typical basal cell morphology and cytoskeletal arrangement was confirmed with high-resolution immunofluorescence microscopy (Fig. 6.7 D). Basal cells from healthy control (n=5) and IPF (n=5) donors (Appendix, Table 7.3) were cultured on ECM proteins known to be present in the fibrotic lung[34]; collagen I, III, IV, versican and fibronectin, all at 50μg/ml, a concentration previously optimised for cell tracking in our lab[263]. Migration was captured using the JuLI live-cell imaging system over 48 hours but only the first 12 hours was analysed because beyond this timepoint, basal cell confluency increased, and tracking became less reliable. Semi-automated analysis of the migration characteristics of individual cells within a field of view was carried out using the Nikon NIS- Elements software platform (Nikon Instruments Inc.). To ensure the data was robust, all automated tracks were rigorously checked for accuracy and manually curated, removing any tracks that were false as described in Section 2.22.1. In total *n*= 1593 healthy and *n*= 1574 IPF basal cells were tracked on the ECM substrates.

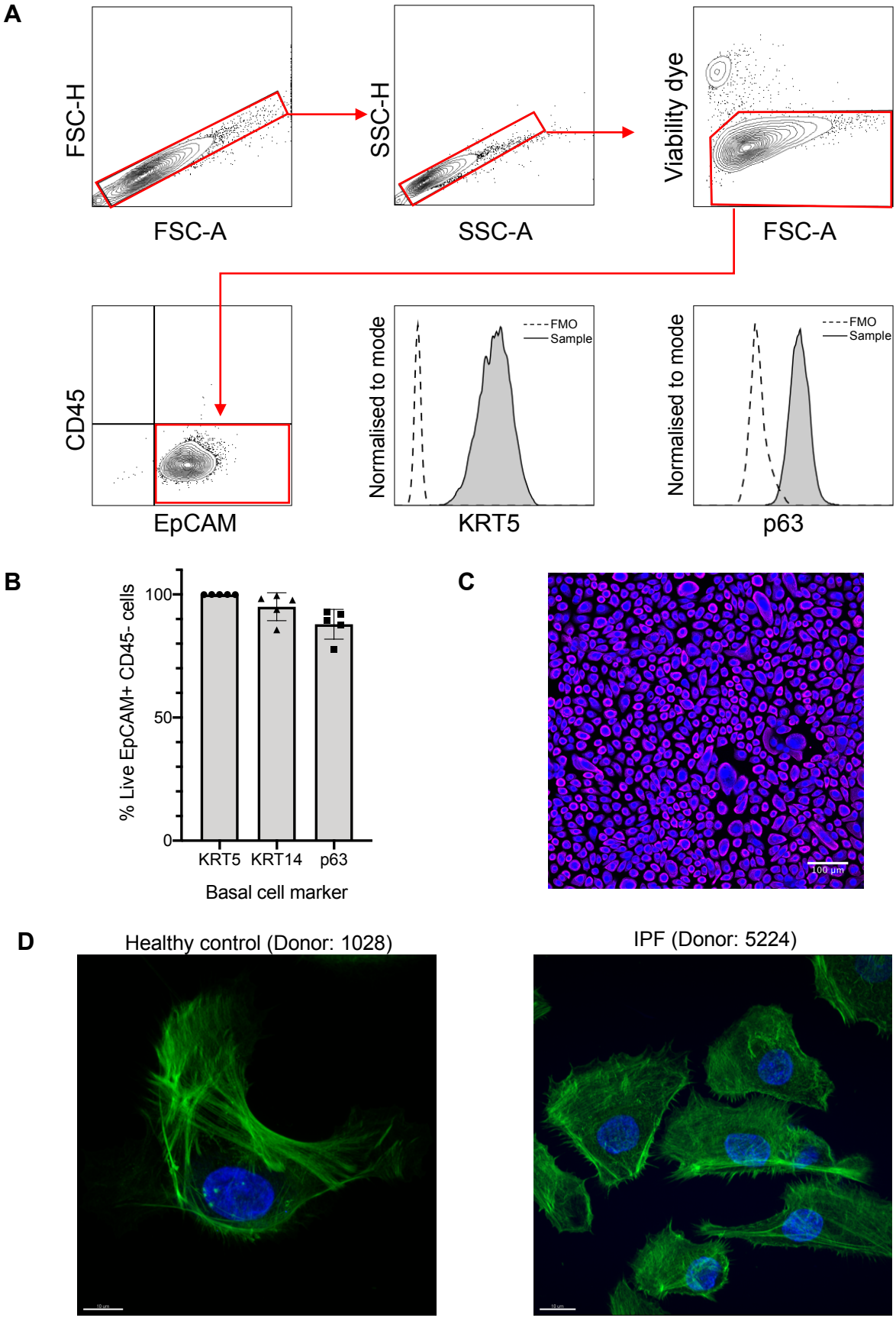


Figure 6.7 Human airway basal cell phenotyping. Primary bronchial epithelial cells (P3) in submerged culture were phenotyped by (A, B) flow cytometry for basal cell markers KRT5, KRT14, and p63, (C) immunofluorescence microscopy for basal cell marker KRT5 (*magenta*) and DAPI (*blue*). (D) maximum intensity projection of z-stack images acquired by Dr Leo Carlin (Beatson Institute) using a Zeiss 880 microscope with Airyscan detectors, showing basal cell morphology and cytoskeletal organisation on collagen I, with F-actin (phalloidin, *green*) and DAPI (*blue*).

Cell migration over the 2D tissue culture surface was observed to occur to varying degrees on all ECM substrates. Basal cells remained viable and could be seen undergoing cell division during the 12- hour period. A mesenchymal mode of migration was observed with lamellipodia protruding at the leading edge of basal cells[437].

Healthy control basal cell migration was enhanced on collagen I and III compared to collagen IV and versican, as evidenced by significantly increased total track length (Fig. 6.8 A), mean track speed (Fig. 6.8 B), and displacement (Fig. 6.8 D). Indices of directional persistence – straightness (directionality ratio) (Fig. 6.8 C) and mean squared displacement (MSD) (Fig. 6.8 E), were also increased on collagen I and III. Basal cells were spatially restricted by collagen IV and versican and appeared to oscillate in one position (Fig. 6.8 F).

Overall, a similar pattern was seen for IPF basal cells (Fig. 6.9 A – F), which also exhibited greater displacement and straighter tracks on collagen I and III, but also fibronectin (*n*= 4 donors tracked). IPF basal cell track length (Fig. 6.9 A), and speed (Fig. 6.9 B), was also significantly increased on collagen III compared to collagen IV. Again, there were striking differences in migratory patterns, dictated by ECM substrate (Fig. 6.9 F). There were no significant differences directly comparing migration between IPF and healthy basal cells on any of the ECM substrates, however there was a modest reduction in track length and speed of IPF basal cells on collagen I compared to healthy control basal cells.

Although initial cell seeding density was controlled, growth rates between donors varied accounting for some becoming overconfluent. Tracking data was not obtained for control subjects 1018, 1028 and 1036 cultured on fibronectin as the cells became too confluent to accurately track. No tracking data was obtained for IPF subject 5374 as the cells did not adhere to the substrate.

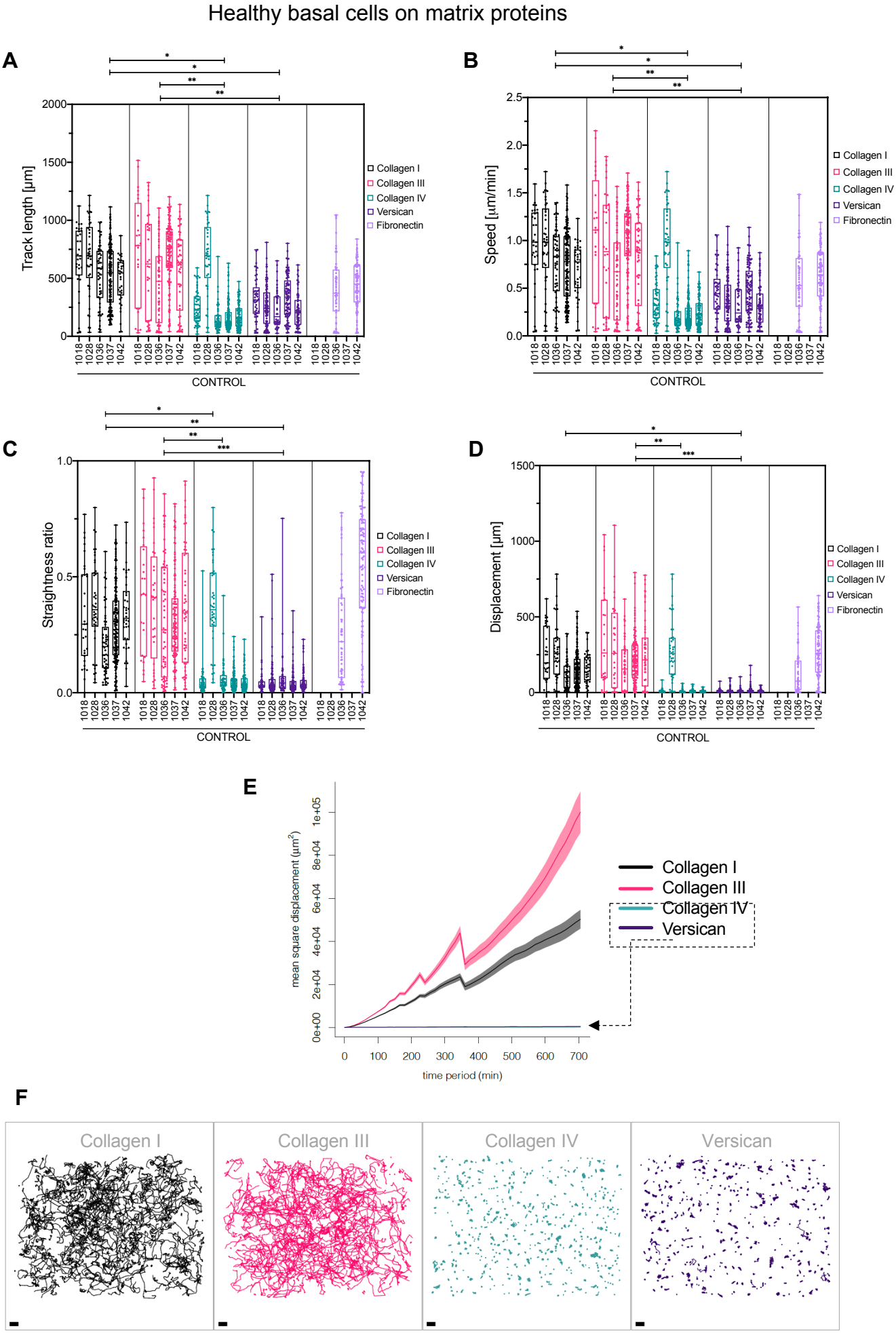

Healthy basal cells on matrix proteins
A
Track length [µm]
2000
1500
1000
500
0
CONTROL
Collagen I
Collagen III
Collagen IV
Versican
Fibronectin
B
Speed [µm/min]
2.5
2.0
1.5
1.0
0.5
0.0
C
Straightness ratio
1.0
0.5
0.0
D
Displacement [µm]
1500
1000
500
0
E
mean square displacement (µm²)
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
0
100
200
300
400
500
600
700
time period (min)
F
Collagen I
Collagen III
Collagen IV
Versican

Figure 6.8 Extracellular matrix proteins modulate migratory dynamics of healthy airway basal cells. Human airway basal cells from healthy donors (n = 5) were seeded on tissue culture plates coated with ECM molecules - collagen I, collagen III, collagen IV, versican or fibronectin. Cell migration was captured using JuLi live-cell imaging system over 12 hours and data analysed using Nikon NIS spot detection and semi-automated cell tracking functions. Data shown as box and whisker dot plots quantifying (**A**) track length (**B**) cell speed (**C**) track straightness and (**D**) displacement. Using MotilityLab software data was plotted as (**E**) mean square displacement per substrate and, (**F**) individual tracks from all donors per substrate. Fibronectin was excluded from E and F because tracking data was only obtained from two donors. Nested one-way ANOVA with Tukey's multiple comparison test, ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05. Scale bars, 100μm.

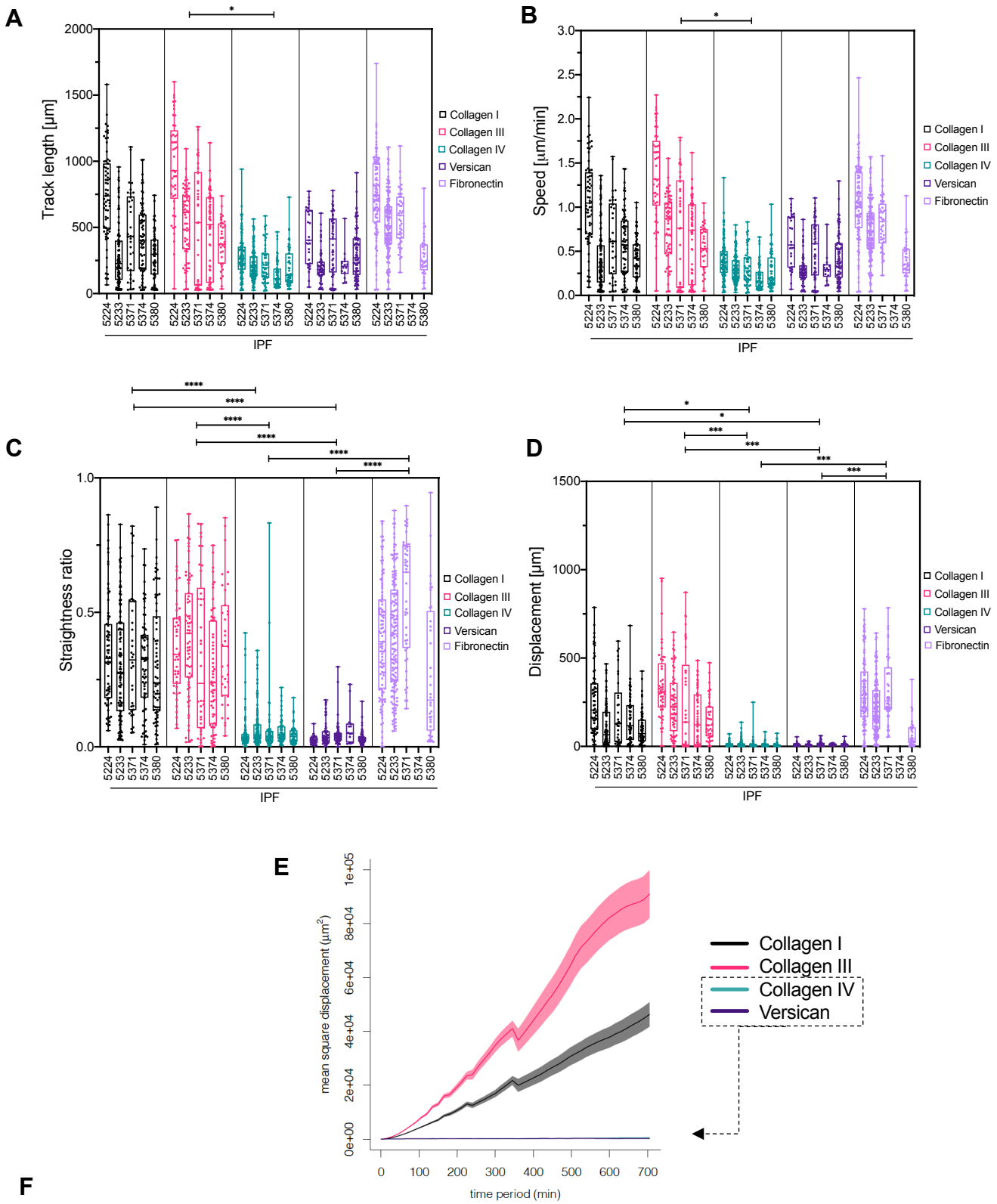

IPF basal cells on matrix proteins
A
B
C
D
E
F
Track length [µm]
Speed [µm/min]
Straightness ratio
Displacement [µm]
mean square displacement (µm²)
time period (min)
IPF
5224
5233
5371
5374
5380
Collagen I
Collagen III
Collagen IV
Versican
Fibronectin
Collagen I
Collagen III
Collagen IV
Versican
Collagen I
Collagen III
Collagen IV
Versican

Figure 6.9 Extracellular matrix proteins modulate migratory dynamics of IPF airway basal cells. Human airway basal cells from IPF donors (n = 5) were seeded on tissue culture plates coated with ECM molecules - collagen I, collagen III, collagen IV, versican or fibronectin. Cell migration was captured using JuLi live-cell imaging system over 12 hours and data analysed using Nikon NIS spot detection and semi-automated cell tracking functions. Data shown as box and whisker dot plots quantifying (**A**) track length (**B**) cell speed (**C**) track straightness and (**D**) displacement. Using MotilityLab software data was plotted as (**E**) mean square displacement per substrate and, (**F**) individual tracks from all donors per substrate. Fibronectin was excluded from E and F because tracking data was only obtained from two donors. Nested one-way ANOVA with Tukey's multiple comparison test, ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05. Scale bars, 100μm.

6.4.4 IPF and healthy basal cells exhibit similar migratory capacity through fibrotic cell-derived matrix

In a 2D culture system, significant migratory differences were driven by changes in the ECM protein substrate, but not by the basal cell disease status (healthy compared to IPF). To test if there were differences between healthy and IPF basal cells migrating through a 3D matrix, recapitulating the complex tissue microenvironment of the IPF distal lung, human lung fibroblasts (HLFs) isolated from IPF lung tissue were used to generate cell-derived matrix (CDM).

IPF HLFs naturally synthesised ECM efficiently to generate a 3D matrix when cultured at high-density for 21 days. Adequate denudation of the fibroblasts without damage to the structural integrity of the CDM was first confirmed by bright-field microscopy (Fig 6.10 A). During the extraction process, despite delicate pipetting, some of the CDMs partially or completely detached from the well and could not be used for further experiments. Immunofluorescence staining of collagen I and fibronectin revealed the structural organisation of the matrix, and absence of phalloidin (F-actin) and DAPI confirmed successful decellularization (Fig 6.10 B). Basal cells from healthy controls (n= 4) and IPF patients (n= 4) (Appendix, Table 7.4) were labelled with a fluorescent dye and seeded on the CDMs. High-resolution live-cell microscopy was used to acquire images at 5-minute intervals for 12 hours. Nikon NIS was used for spot detection and cell tracking. A total of 213 healthy basal cells and 362 IPF basal cells were tracked, and a spectrum of migratory patterns were observed which differed to that seen in the 2D system. In some areas basal cells could be seen adopting an elongated mesenchymal morphology then migrating in a directional manner dictated by the matrix topography visualised by bright-field microscopy. Basal cells could also be seen traversing back and forth over previously made tracks. Other cells appeared spatially restricted and remained anchored in one position within the matrix where they oscillated around that point. Although two clear migratory patterns were observed within the total population of basal cells, there were no significant differences in track length (Fig. 6.11 A), mean speed (Fig. 6.11 B), straightness ratio (Fig. 6.11 C), or track displacement (Fig. 6.11 D), between basal cells from IPF patients and healthy controls. Tracks plotted from centroid (Fig. 6.11 E) and mean square displacement (Fig. 6.11 F), were comparable between healthy and IPF basal cells.

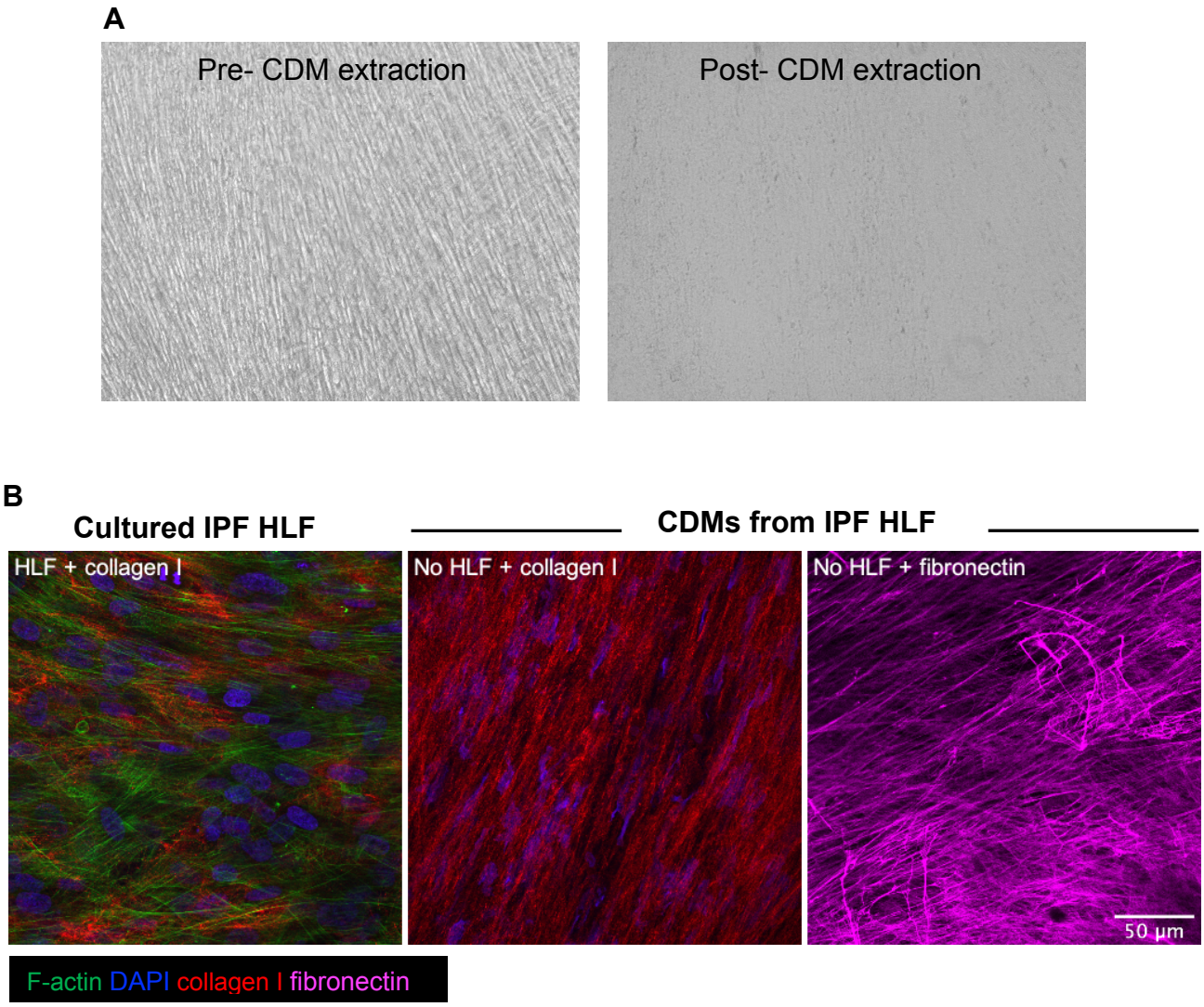


Figure 6.10 Generation of CDMs from primary human lung fibroblasts. Primary HLFs from IPF patients cultured at high-density secrete and organise a natural 3D matrix. (A) Bright-field images of the wells containing HLF and secreted matrix pre- and post- CDM extraction. (B) Immunofluorescence (IF) microscopy showing primary IPF HLF cultured at high-density and extracted CDMs following denudation of HLFs. Matrix components collagen I and fibronectin were stained. IF images were captured using a Leica SP5 inverted confocal microscope.

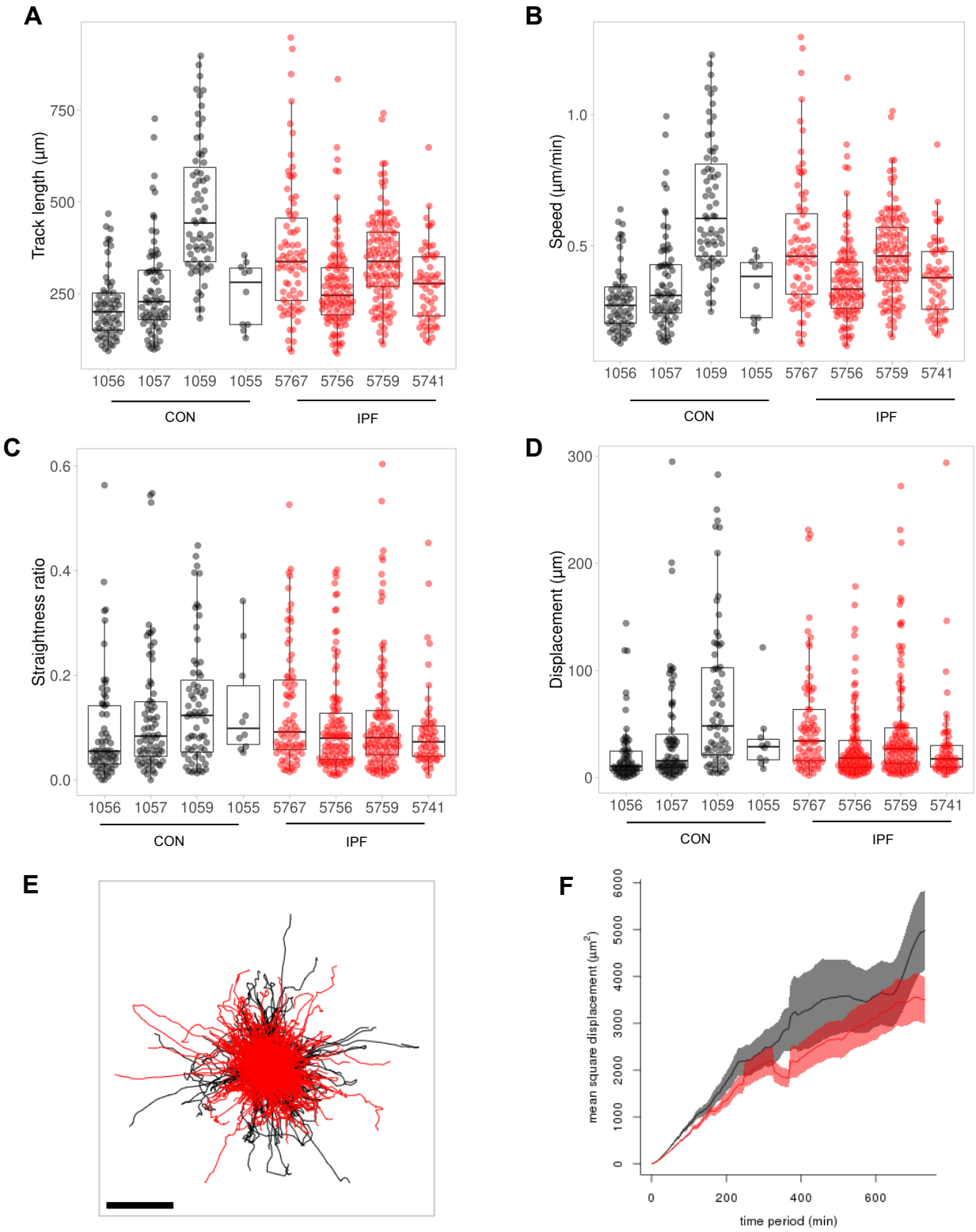


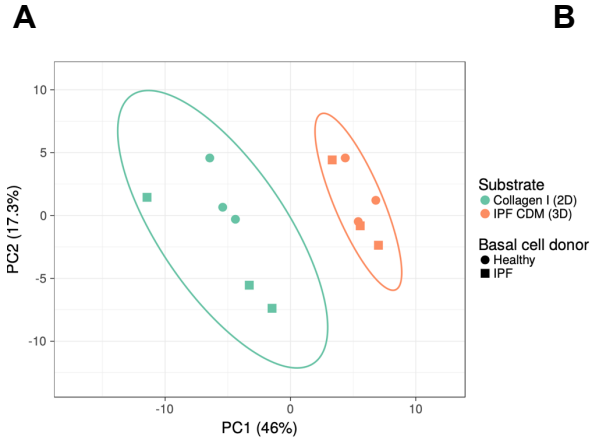
Figure 6.11 Healthy and IPF airway basal cell migration through IPF CDM. Basal cells from healthy controls (n=4) and IPF patients (n=4) were fluorescently labelled and seeded on CDMs generated from IPF HLFs. Cell migration was captured using a Zeiss Axio Observer Inverted Widefield Microscope and data analysed using Nikon NIS spot detection and semi-automated cell tracking functions. Data shown as box and whisker dot plots quantifying (**A**) track length (**B**) speed (**C**) straightness ratio and (**D**) displacement. Using MotilityLab software collated tracks from healthy controls (*black*) and IPF patients (*red*) were plotted from (**E**) centroid and (**F**) mean square displacement was plotted for healthy (*black*) and IPF (*red*) basal cells. Scale bar, 100µm.

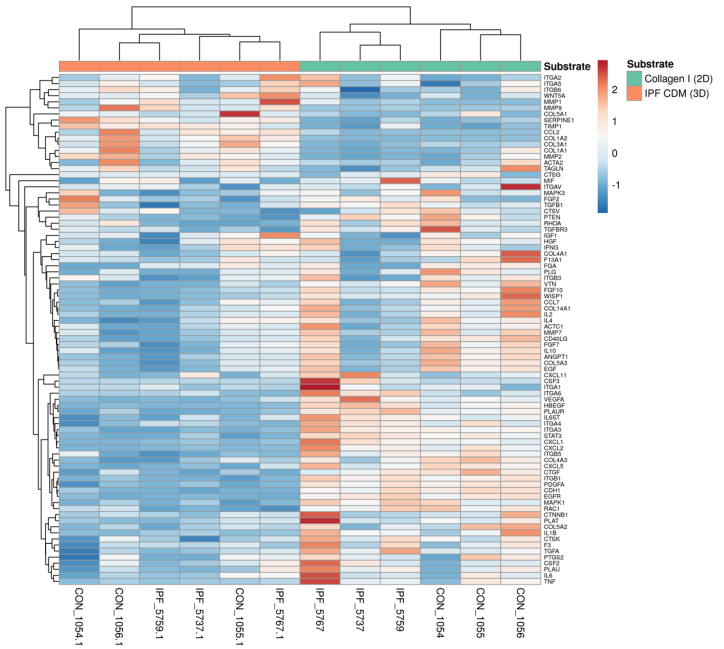
6.4.5 Changes to matrix microenvironment induce differential expression of wound healing genes in basal cells

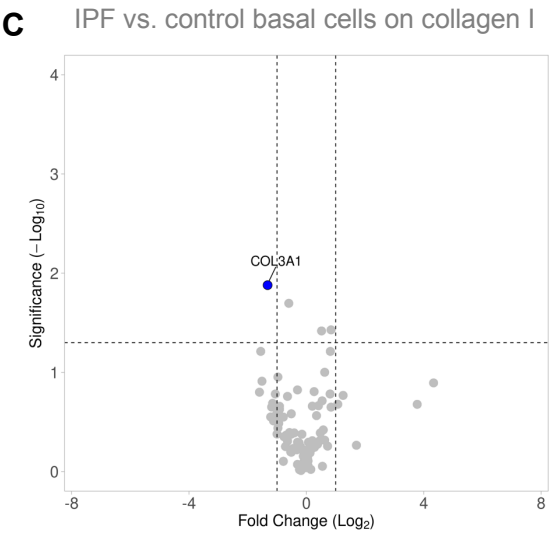
To interrogate relevant molecular changes that occur in basal cells interacting with different matrix substrates, a PCR array measuring the expression of 84 genes linked to wound healing was performed before and after cells migrated through IPF CDMs. Prior to seeding the basal cells on CDMs, they were expanded under standard conditions on collagen I coated plates. RNA was harvested from IPF and healthy basal cells (*n*=3 in each group) (Appendix, Table 7.5) cultured on collagen I (2D) or after migration through IPF CDM (3D). RNA integrity number (RIN) ranged from 7.3 to 10. All 12 samples passed the PCR array reproducibility and reverse transcription efficiency test performed in the Qiagen software. The mean C_T value of three housekeeping genes (*ACTB, B2M* and *RPLP0)* showing least variance in expression between samples were used to calculate normalised gene expression.

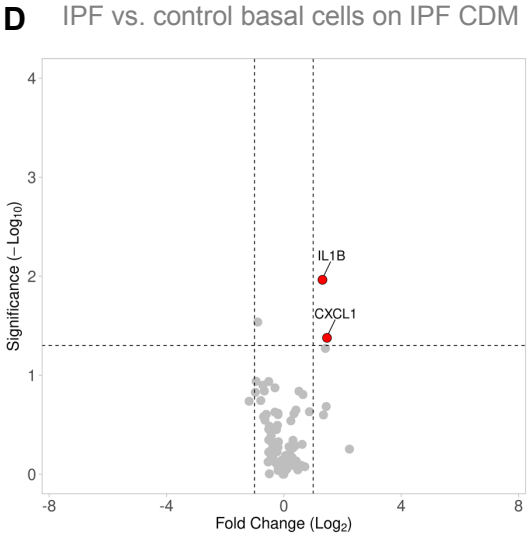
Principal component analysis (PCA) clearly demonstrated the samples clustering according to the substrate cultured on (PC1; accounting for 46% variance), rather than disease status of basal cells (Fig. 6.12 A). Similarly, unsupervised hierarchical clustering showed samples clustering according to matrix substrate (Fig. 6.12 B). Volcano plots were used to visualise significant gene expression changes between two groups. Direct comparison of gene expression changes between IPF and healthy basal cells cultured on either substrate revealed very few significant differences (Fig. 6.12 C and D). Interestingly, on IPF CDM, *IL1B* and *CXCL1*, were significantly upregulated in IPF basal cells compared to controls (Fig. 6.12 D). Strikingly, there were a greater number of significant gene expression differences when analysing basal cells (healthy or IPF) cultured on a 3D IPF CDM compared to a collagen I coated 2D surface (Fig. 6.12 E and F). When comparing the influence of different substrates on fold changes in gene expression, collagen I was assigned as the ‘control group’ and IPF CDM as the ‘test group’. Genes downregulated in basal cells on IPF CDM vs. collagen I (the control), were therefore expressed at higher levels on collagen I compared to IPF CDM. Genes encoding collagen 1 alpha 2 chain (*COL1A2*) and serpin family E family member 1 (*SERPINE1*) were upregulated in healthy and IPF basal cells cultured on IPF CDM. Exclusively upregulated in IPF basal cells were other key collagen genes (*COL3A1, COL1A1*), and MMP9, a matrix metallopeptidase. In

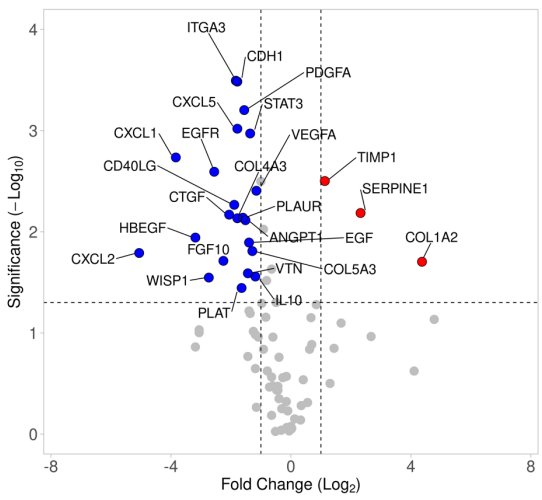
contrast, on collagen I, genes encoding growth factors (*CTGF, HBEGF*), pro-inflammatory chemokines (*CXCL1, CXCL2*) and integrin subunits such as *ITGA3*, were expressed at higher levels than on IPF CDM.


A
PC2 (17.3%)
PC1 (46%)
Substrate
Collagen I (2D)
IPF CDM (3D)
Basal cell donor
Healthy
IPF


B
Substrate
Collagen I (2D)
IPF CDM (3D)
CON_1054.1
CON_1056.1
IPF_5759.1
IPF_5737.1
CON_1055.1
IPF_5767.1
IPF_5767
IPF_5737
IPF_5759
CON_1054
CON_1055
CON_1056


C
IPF vs. control basal cells on collagen I
COL3A1
Significance (−Log10)
Fold Change (Log2)


D
IPF vs. control basal cells on IPF CDM
IL1B
CXCL1
Significance (−Log10)
Fold Change (Log2)


E
IPF CDM (3D) vs collagen I (2D) - Healthy basal cells
ITGA3
CDH1
PDGFA
CXCL5
STAT3
CXCL1
EGFR
VEGFA
CD40LG
COL4A3
TIMP1
SERPINE1
CTGF
PLAUR
HBEGF
ANGPT1
EGF
COL1A2
CXCL2
FGF10
VTN
COL5A3
WISP1
IL10
PLAT
Significance (−Log10)
Fold Change (Log2)

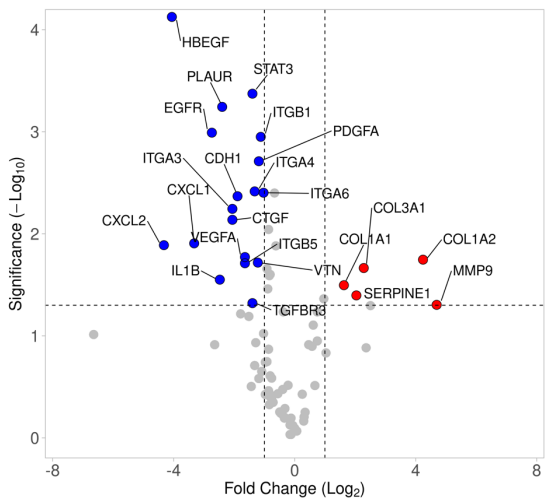

F
IPF CDM (3D) vs. collagen I (2D) – IPF basal cells
HBEGF
PLAUR
STAT3
EGFR
ITGB1
PDGFA
ITGA3
CDH1
ITGA4
CXCL1
ITGA6
COL3A1
CXCL2
CTGF
VEGFA
ITGB5
COL1A1
COL1A2
IL1B
VTN
MMP9
SERPINE1
TGFBR3
Significance (−Log10)
Fold Change (Log2)

Figure 6.12 Extracellular matrix microenvironment drives gene expression changes in basal cells. A Qiagen Wound Healing RT^2 PCR Profiler Array was used to measure gene expression changes in healthy (n=3) and IPF (n=3) airway basal cells cultured on a collagen I coated plate (2D) or IPF CDM (3D) **(A)** Principal component analysis of the gene array data; unit variance scaling was applied to rows; SVD with imputation was used to calculate principal components. Prediction ellipses are such that with probability 0.95, a new observation from the same group will fall inside the ellipse. **(B)** Hierarchical clustering of normalised gene expression. Rows and columns using correlation distance and average linkage. Rows are centred and unit variance scaling is applied. Volcano plots to visualise significant gene expression changes between (**C**) IPF vs. healthy basal cells on collagen I, (**D**) IPF vs. healthy basal cells on IPF CDM, (**E**) healthy basal cells cultured on IPF CDM vs. collagen I, and (**F**) IPF basal cells cultured on IPF CDM vs. collagen I. Upregulated genes (*red)*, downregulated genes (*blue)*. Significance defined as log fold change >2 or <2, $p<0.05$. P values were calculated based on a Student's t-test of the replicate normalised gene expression values (2^(- Delta CT)) for each gene in control and test groups. Housekeeping genes used - *ACTB, B2M* and *RPLP0.*

6.4.6 Basal cell migration is restricted through IPF cell-derived matrix

Collectively this data indicates that disease status does not significantly alter basal cell migratory behaviour, but instead the properties of the matrix play a more significant role. To investigate whether disease-specific changes in the lung matrix microenvironment differentially regulate basal cell migration, healthy basal cells were tracked through cell-derived matrix generated from IPF HLFs (*n*=3) and control HLFs (*n*=3) (Appendix, Table 7.6). In total 301 cells were tracked on control matrix and 257 cells tracked on IPF matrix. There was significant inter-donor variability and as such in a nested analysis, there were no significant differences found between composite cell migration indices such as track length (Fig. 6.13 A), speed (Fig. 6.13 B), straightness (Fig. 6.13 C), or displacement (Fig. 6.13 D). As before, heterogeneity in basal cell migratory behaviour through the CDMs was observed with two broad migratory patterns. Spatially restricted cells remained in a specific location with minimal displacement from their starting position over 12hrs. Other more migratory cells exhibited greater directional persistence with a higher straightness ratio and overall displacement from their starting position. There was a higher frequency of migratory cells on control matrix compared with IPF (Fig. 6.13 E and F). This data could also be visualised by plotting individual basal cell tracks on control and IPF CDMs (Fig. 6.13 G). All tracks plotted from centroid showed more spatially restricted cell tracks on IPF CDM compared to control (Fig. 6.13 H). Overall, mean square displacement was reduced through the IPF CDMs compared to control (Fig. 6.13 I). Immunofluorescence imaging demonstrated basal cells interacting directly with CDMs (Fig. 6.14). Collectively, this data indicates that IPF matrix restrains cell migration relative to control.

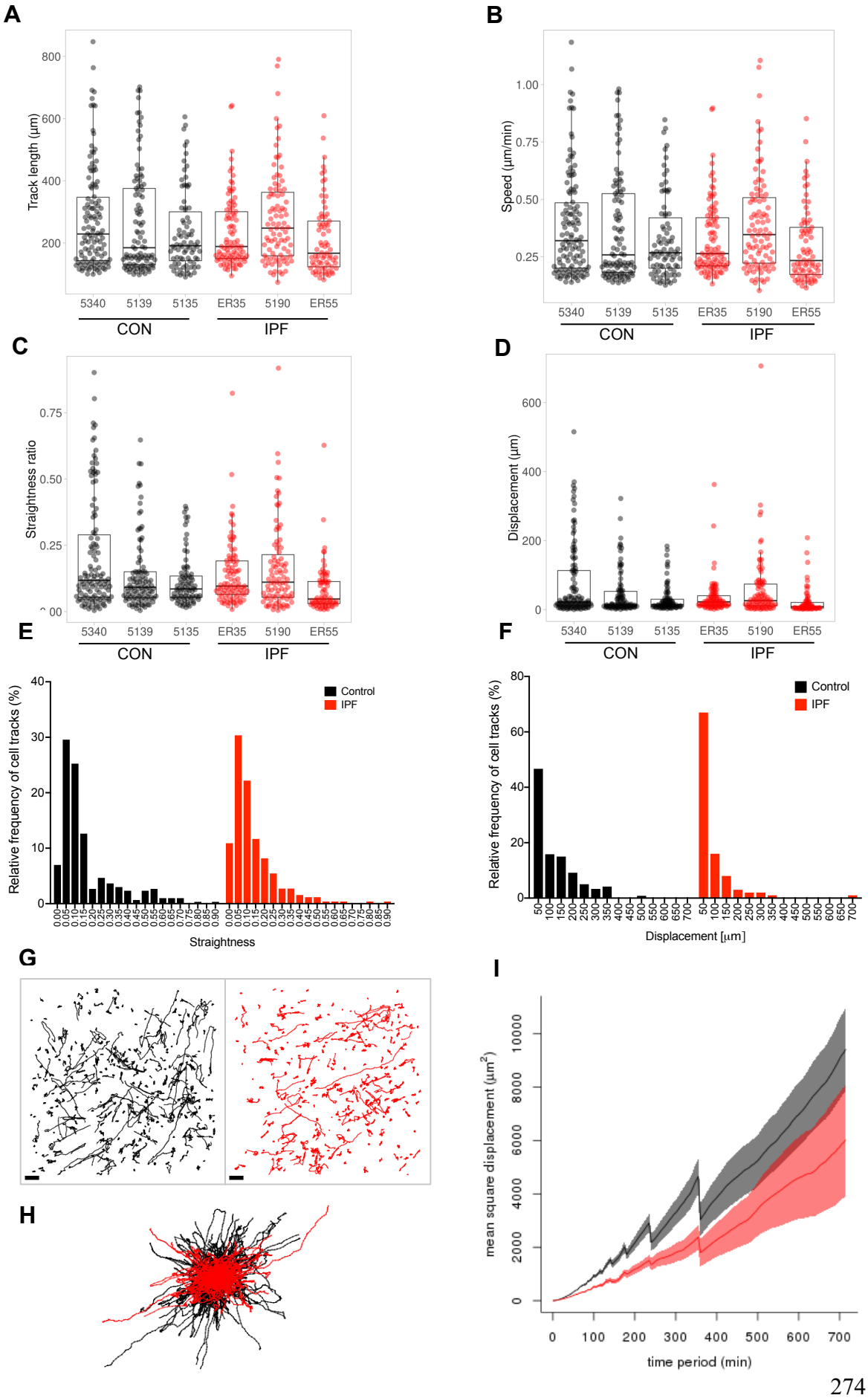

A
Track length (μm)
800
600
400
200
5340
5139
5135
ER35
5190
ER55
CON
IPF
B
Speed (μm/min)
1.00
0.75
0.50
0.25
C
Straightness ratio
0.75
0.50
0.25
D
Displacement (μm)
600
400
200
0
E
Relative frequency of cell tracks (%)
40
30
20
10
Control
IPF
Straightness
F
80
60
40
20
Displacement [μm]
G
H
I
mean square displacement (μm²)
10000
8000
6000
4000
2000
time period (min)

Figure 6.13 Airway basal cell migration through control and IPF CDMs. Healthy airway basal cells were fluorescently labelled and seeded on CDMs generated from control (*n*=3) and IPF HLFs (*n*=3). Cell migration was captured using a Zeiss Axio Observer Inverted Widefield Microscope and data analysed using Nikon NIS spot detection and semi-automated cell tracking functions. Data shown as box and whisker dot plots quantifying (**A**) track length (**B**) speed (**C**) straightness ratio and (**D**) displacement. Histograms of (**E)** track straightness ratio frequency distribution and **(F)** track displacement frequency distribution. Using MotilityLab software tracks on control CDM (*black*) and IPF CDM (*red*) were plotted (**G**) as measured, and (**H**) from centroid. (**E**) Mean square displacement was plotted for control CDM (*black*) and IPF CDM (*red*). Scale bar, 100μm.

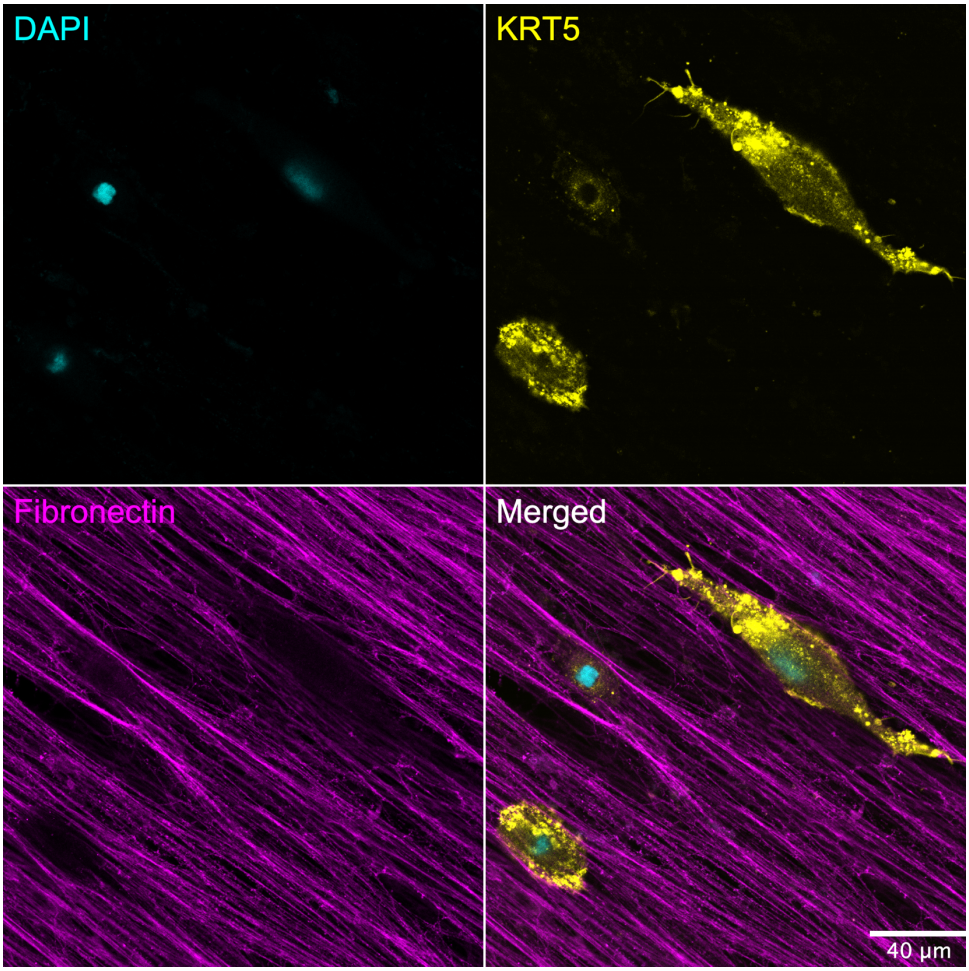


Figure 6.14 Interactions between basal cells and fibronectin in IPF CDMs. Basal cells (phalloidin, *yellow*) within an IPF CDM. Processes can be seen directly interacting with fibronectin fibres (*magenta*). Nucleus counterstained with DAPI (*cyan).* Images were acquired using a Leica SP8 inverted confocal microscope**.**

6.4.7 MS-based proteomics of cell-derived matrices

Differences in basal cell migration were observed on matrix derived from IPF fibroblasts compared to control, therefore mass spectrometry (MS) – based proteomics was used to investigate compositional differences in the matrix which may explain these findings. Fibroblast-deposited ECM from IPF patients (*n*=3 in duplicate) and non-fibrotic controls (*n*=3 in duplicate) was isolated and enriched using a fractionation protocol established by Lennon et al.[266] (Fig. 6.15 A). Western blotting confirmed enrichment of ECM protein, fibronectin in the ECM fraction (fraction 4) and relative depletion of laminin B (nuclear protein) and tubulin (cytoplasmic protein) (Fig. 6.15 B). As detailed in Section 2.23.3 Maryline Fresquet processed the ECM enriched fraction for MS-based proteomics and performed initial data processing.

After taking forward only proteins with a high confidence score, detected with a minimum of 1 unique peptide and present in at least two-thirds of the samples in each group, a total of 2795 proteins were identified by MS. Using annotations from the Gene Ontology (GO) knowledgebase (http://geneontology.org/), Proteome Discoverer categorized 4.56% proteins detected in the ECM enriched fraction, as “extracellular” (Fig. 6.15 C). When cross-referenced with the human matrisome database[154],154 matrix proteins were identified (Table 6.1). The most abundant proteins in the CDMs were ECM glycoproteins and ECM regulators (Fig. 6.15 D). Functional enrichment analysis using the Database for Annotation, Visualisation and Integrated Discovery (DAVID)[442] revealed the top 10 GO Biological Processes included extracellular matrix organisation, cell adhesion, response to wounding and regulation of cell migration (Fig. 6.15 E). The IPF fibroblast secreted- matrisome clustered separately to that generated by control fibroblast in a principal component analysis (PCA) (Fig. 6.16 A). Disease status of fibroblast donor was the main source of variation (PC1 37.9%). PC1-5 explained 90% of the total variance seen in the dataset (Fig. 6.16 B). Similarly, unsupervised hierarchical clustering of the matrisome proteins by normalised abundance, demonstrated clustering of CDMs by fibroblast disease state status (Fig. 6.16 C).

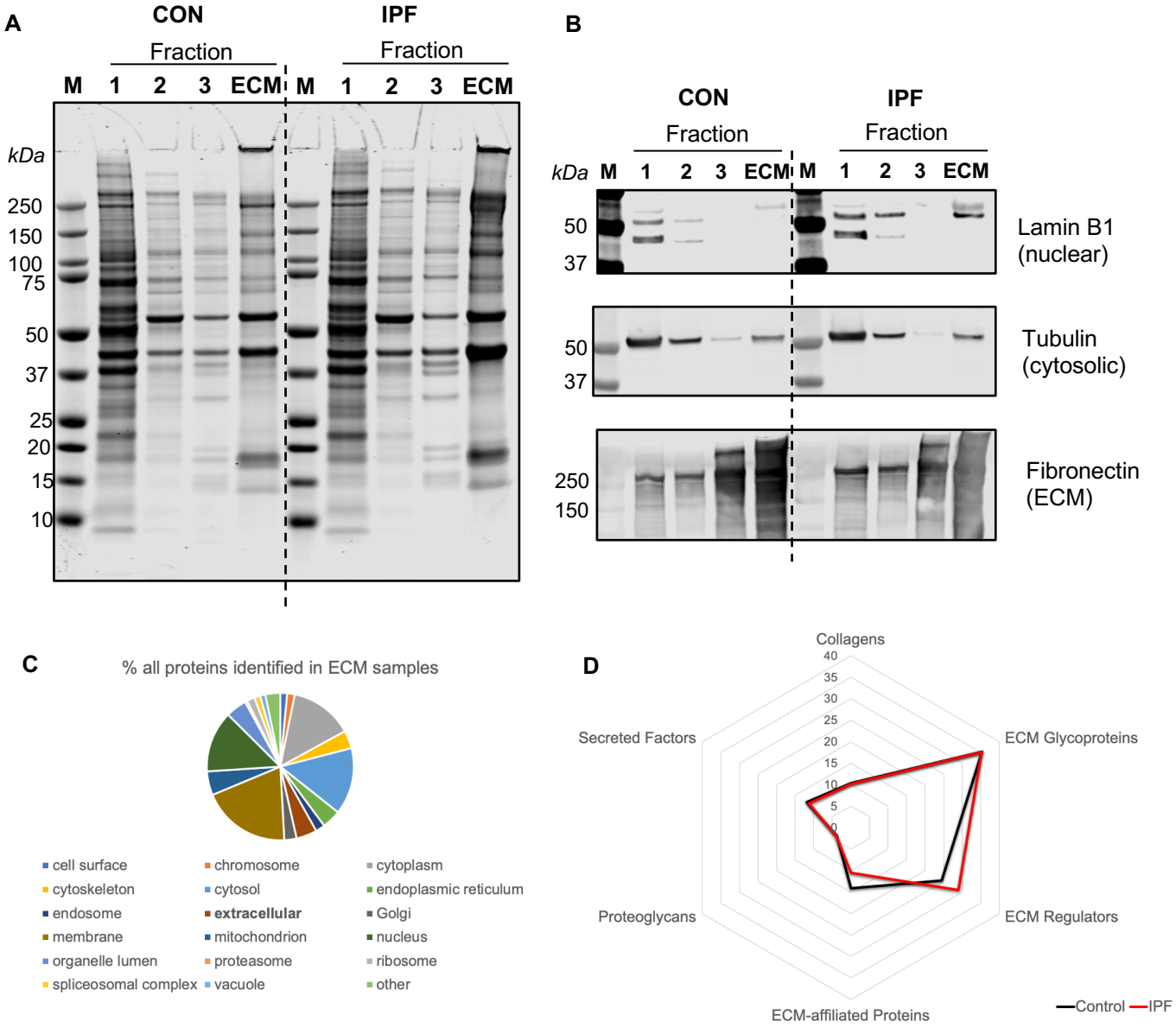


Figure 6.15 Isolation and MS analysis of enriched lung fibroblast-derived ECM. Data acquired and analysed in collaboration with Dr Maryline Fresquet and Prof Rachel Lennon at the Wellcome Trust Centre for Cell-Matrix Research. (**A**) Coomassie staining, and (**B**) Western blotting (WB) of fractions 1–3 and ECM fraction probing for nuclear proteins with lamin B1, intracellular proteins with tubulin and extracellular proteins with fibronectin. Representative example from control CDM and IPF CDM. M, molecular mass marker. (**C**) Pie chart representing proportion of all proteins identified in ECM fraction, characterised by Proteome Discoverer software using annotations from the GO knowledgebase (**D**) Radar plot of protein abundance according to matrisome categories identified in CDMs from IPF (*red)* and (*control)* (**E**) GO Biological Pathways using DAVID v6.8 functional annotation tool for *n*=154 matrisome proteins identified (Benjamini-Hochberg adjusted p-value).

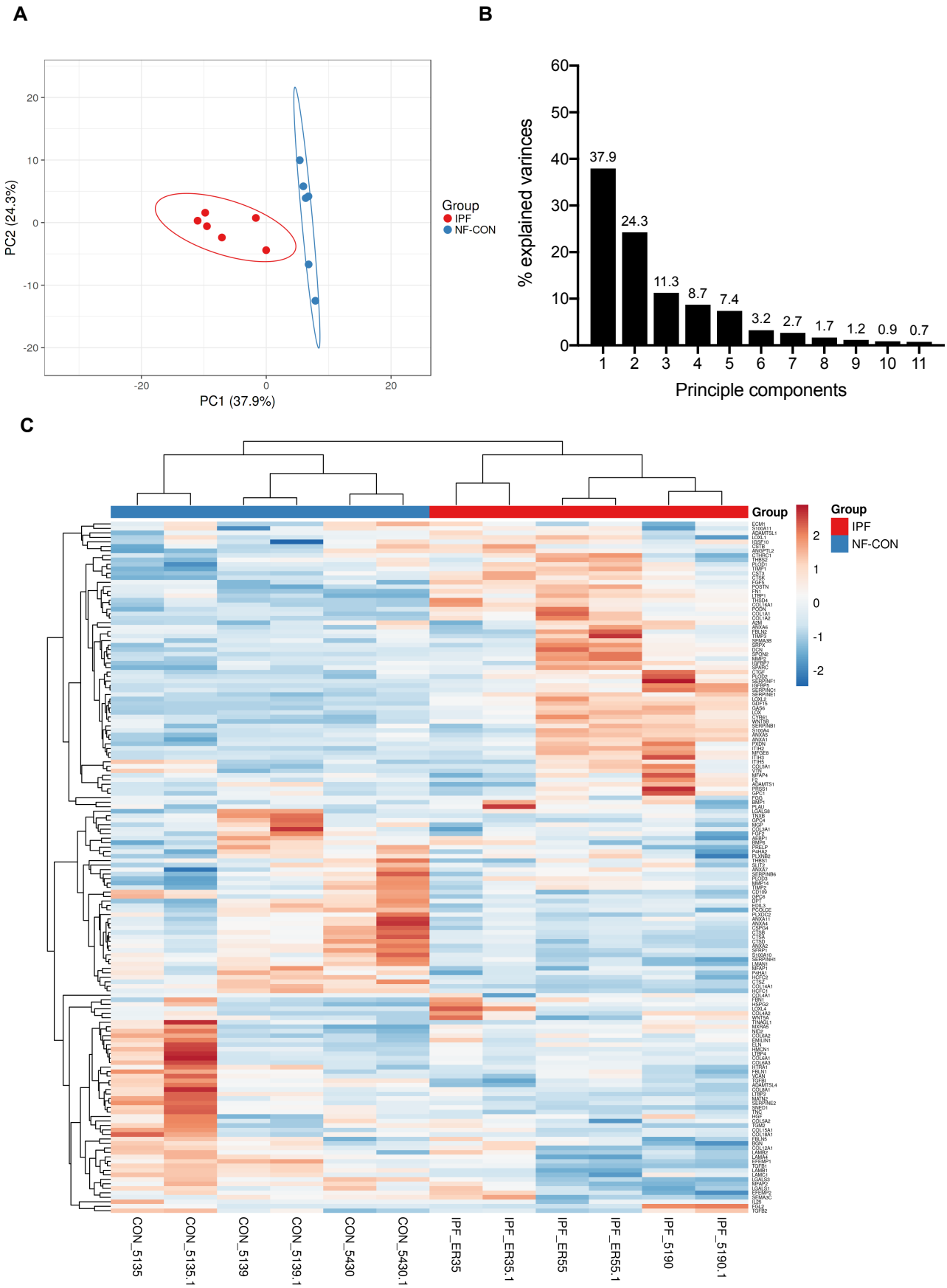

A
PC2 (24.3%)
PC1 (37.9%)
Group
IPF
NF-CON
B
% explained varinces
Principle components
37.9
24.3
11.3
8.7
7.4
3.2
2.7
1.7
1.2
0.9
0.7
C
Group
IPF
NF-CON
CON_5135
CON_5135.1
CON_5139
CON_5139.1
CON_5430
CON_5430.1
IPF_ER35
IPF_ER35.1
IPF_ER55
IPF_ER55.1
IPF_5190
IPF_5190.1

Figure 6.16 Unsupervised clustering of 154 matrisome proteins identified in lung fibroblast generated CDM by MS. Data acquired and analysed in collaboration with Dr Maryline Fresquet and Prof Rachel Lennon at the Wellcome Trust Centre for Cell-Matrix Research (**A)** PCA of matrisome dataset from non-fibrotic control CDM *(blue)* (n=3 in duplicate) and IPF CDM (*red)* (*n*=3 in duplicate). Unit variance scaling is applied to rows; SVD with imputation is used to calculate principal components. Prediction ellipses are such that with probability 0.95, a new observation from the same group will fall inside the ellipse. (**B)** Scree plot of % variances explained by each principal component **(C)** Hierarchical clustering of normalised abundance of each identified matrisome proteins. Rows and columns using correlation distance and average linkage. Rows are centred and unit variance scaling is applied.

In total there were 19 significantly differentially expressed matrix proteins between IPF and control CDMs ($p<0.05$, Log_2 fold change >1), of these 14 were upregulated in IPF CDM compared to control (Fig. 6.17 A). Fourteen of the differentially expressed proteins were included in one or more of three enriched GO Biological Processes that are relevant to cell migration (Fig. 6.17 B). Insulin Like Growth Factor Binding Protein 5 (IGFBP-5) was the most significantly upregulated protein in IPF CDM compared to control (2.61 Log_2 fold change, $p <0.001$) and is linked to cell migration[443]. Other significantly upregulated proteins with roles in the regulation of cell migration and ECM organisation include Growth Arrest Specific Protein-6 (GAS6), Fibronectin (FN1) and Secreted Protein Acidic and Cysteine Rich (SPARC). Analysis using the STRING database showed specific protein-protein interactions between IGFBP-5, GAS6, FN1, SPARC, Milk Fat Globule-EGF Factor 8 (MFGE8) and Fibrillin-1 (FBN1) (FDR stringency: 1%) (Fig. 6.17 C and D).

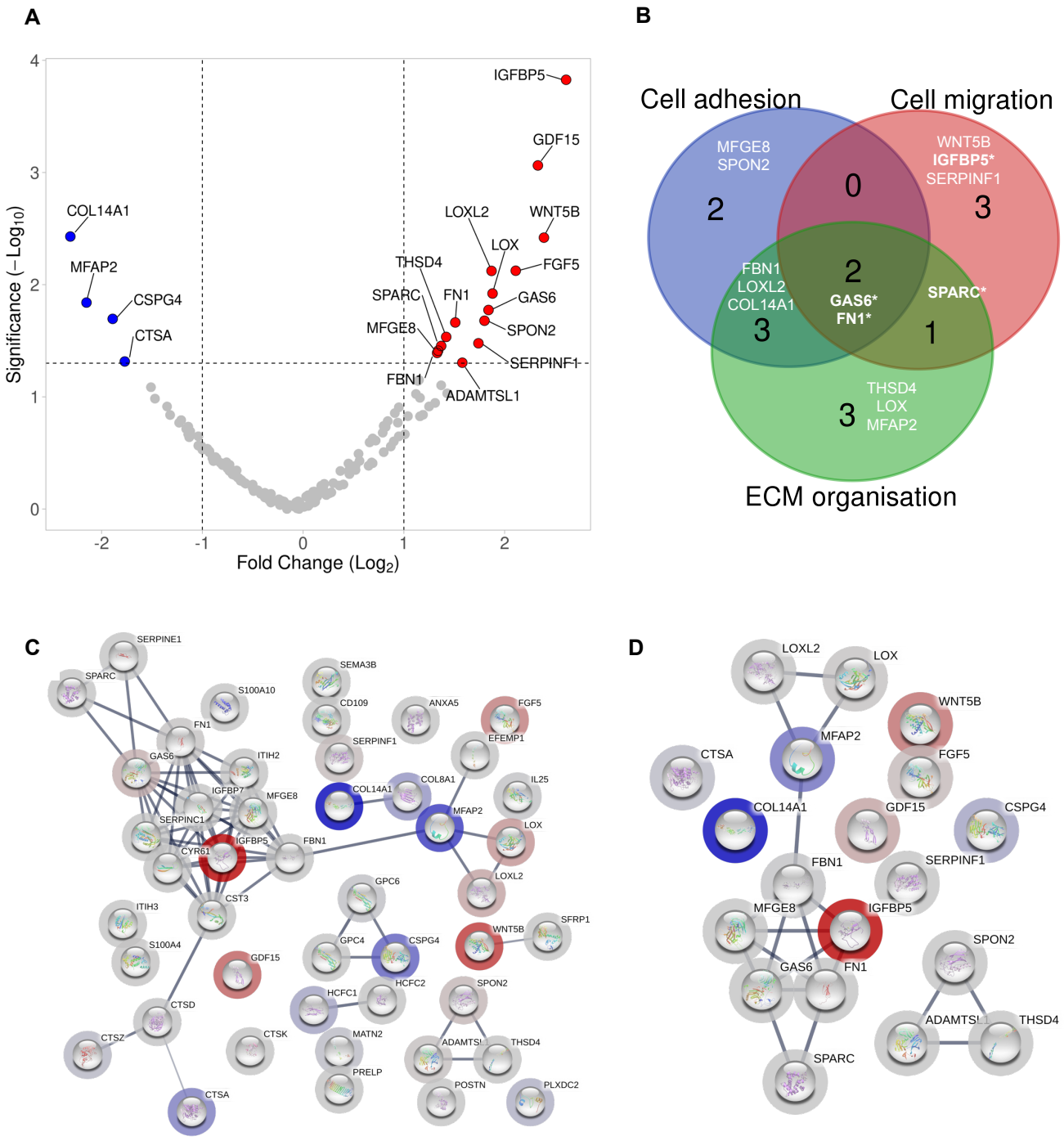


Figure 6.17 IPF and control lung fibroblasts have differential ECM composition. Data acquired and analysed in collaboration with Dr Maryline Fresquet and Prof Rachel Lennon at the Wellcome Trust Centre for Cell-Matrix Research. (**A**) Volcano plot showing significant differentially expressed proteins with between IPF CDMs and control CDMs, upregulated proteins (*red)*, downregulated proteins (*blue)*. Significance defined as a fold change >2 or <2, p<0.05. (**B**) Venn diagram showing shared GO biological processes for significantly differentially expressed proteins (**C**) Protein-protein interaction network using STRING v11 using matrisome proteins with an abundance ratio >2 or <2 fold change or, (**D**) an abundance ration >2 or <2, p<0.05.

Table 6.1 Proteins identified in human lung fibroblast generated CDMs

Protein	Gene Symbol	MW (kDa)	Abundance Ratio (log2)	P-Value	Classification
Insulin-like growth factor-binding protein 5	IGFBP5	30.6	2.61	0.0001	ECM Glycoproteins
Protein Wnt-5b	WNT5B	31.6	2.39	0.0038	Secreted Factors
Growth differentiation factor 15	GDF15	34.1	2.33	0.0009	Secreted Factors
Fibroblast growth factor 5	FGF5	14.5	2.11	0.0075	Secreted Factors
Protein-lysine 6-oxidase	LOX	46.9	1.88	0.0120	ECM Regulators
Lysyl oxidase homolog 2	LOXL2	86.7	1.87	0.0075	ECM Regulators
Growth arrest-specific protein 6	GAS6	79.6	1.84	0.0168	ECM Glycoproteins
Spondin-2	SPON2	35.8	1.8	0.0210	ECM Glycoproteins
Pigment epithelium-derived factor	SERPINF1	46.3	1.74	0.0333	ECM Regulators
ADAMTS-like protein 1	ADAMTSL1	193.3	1.58	0.0496	ECM Regulators
Fibronectin	FN1	262.5	1.51	0.0217	ECM Glycoproteins
Protein S100-A4	S100A4	11.7	1.43	0.0931	Secreted Factors
Thrombospondin type-1 domain-containing protein 4	THSD4	112.4	1.42	0.0292	ECM Glycoproteins
Secreted Protein Acidic And Rich In Cysteine	SPARC	34.6	1.37	0.0352	ECM Glycoproteins
Semaphorin-3B	SEMA3B	83.6	1.37	0.0789	ECM-affiliated Proteins
Milk Fat Globule-EGF Factor 8 Protein	MFGE8	38.4	1.34	0.0388	ECM Glycoproteins
Fibrillin-1	FBN1	312	1.33	0.0406	ECM Glycoproteins
Cystatin-C	CST3	15.8	1.31	0.1133	ECM Regulators
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	75	1.25	0.1219	ECM Regulators
Antithrombin-III	SERPINC1	52.6	1.19	0.1445	ECM Regulators
Cysteine-rich, angiogenic inducer, 61 variant	CYR61	42	1.16	0.1492	ECM Glycoproteins
Plasminogen activator inhibitor 1	SERPINE1	45	1.16	0.1020	ECM Regulators
Periostin	POSTN	90.4	1.15	0.0706	ECM Glycoproteins
Annexin A5	ANXA5	35.9	1.14	0.0721	ECM-affiliated Proteins
Cathepsin K	CTSK	36.9	1.12	0.0943	ECM Regulators
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	106.4	1.04	0.1487	ECM Regulators
Insulin-like growth factor-binding protein 7	IGFBP7	29.1	1.01	0.2156	ECM Glycoproteins
Prothrombin	F2	5.7	0.95	0.2249	ECM Regulators
Decorin	DCN	39.7	0.95	0.1240	Proteoglycans
Metalloproteinase inhibitor 1	TIMP1	10.6	0.94	0.1699	ECM Regulators

Latent-transforming growth factor beta-binding protein 1	LTBP1	186.7	0.91	0.1397	ECM Glycoproteins
Podocan	PODN	45.7	0.9	0.2496	Proteoglycans
Vitronectin	VTN	54.3	0.87	0.2771	ECM Glycoproteins
Matrix-remodeling-associated protein 5	MXRA5	312	0.84	0.1645	ECM Glycoproteins
72 kDa type IV collagenase	MMP2	73.8	0.82	0.1852	ECM Regulators
Connective tissue growth factor	CTGF	38	0.8	0.3112	ECM Glycoproteins
Annexin A1	ANXA1	38.7	0.8	0.1816	ECM-affiliated Proteins
collagen alpha-1(XVI) chain	COL16A1	157.7	0.79	0.2088	Collagens
Collagen alpha-2(IV) chain	COL4A2	167.4	0.72	0.2456	Collagens
Metalloproteinase inhibitor 3	TIMP3	24.1	0.69	0.3904	ECM Regulators
Heparan sulfate proteoglycan 2 (Perlecan), isoform CRA_b	HSPG2	463.7	0.69	0.2350	Proteoglycans
Lysyl oxidase homolog 4	LOXL4	84.4	0.68	0.3846	ECM Regulators
Trypsin-1	PRSS1	28.1	0.66	0.2547	ECM Regulators
Sushi repeat-containing protein SRPX	SRPX	51.5	0.65	0.4117	ECM Glycoproteins
Alpha-2-macroglobulin	A2M	163.2	0.56	0.3727	ECM Regulators
Collagen alpha-1(I) chain	COL1A1	138.9	0.55	0.3220	Collagens
Sushi repeat-containing protein SRPX2	SRPX2	52.9	0.52	0.4484	ECM Glycoproteins
Angiopoietin-like 2	ANGPTL2	57	0.52	0.4125	Secreted Factors
Peroxidasin homolog	PXDN	165.2	0.46	0.4446	ECM Glycoproteins
Metalloendopeptidase	BMP1	90.6	0.46	0.4882	ECM Regulators
Collagen alpha-2(I) chain	COL1A2	129.1	0.45	0.3932	Collagens
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	81.1	0.4	0.6040	ECM Regulators
Glypican-1	GPC1	61.6	0.37	0.4577	ECM-affiliated Proteins
Annexin A6	ANXA6	75.8	0.36	0.6175	ECM-affiliated Proteins
A disintegrin-like and metalloprotease with thrombospondin type 1 motif, 1	ADAMTS1	103.3	0.3	0.5880	ECM Regulators
Cystatin-B	CSTB	11.1	0.28	0.7105	ECM Regulators
Collagen triple helix repeat-containing protein 1	CTHRC1	26.2	0.26	0.7245	ECM Glycoproteins
Urokinase-type plasminogen activator	PLAU	39	0.23	0.7503	ECM Regulators
Fibulin-2	FBLN2	129.3	0.21	0.5965	ECM Glycoproteins
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3	84.7	0.2	0.6397	ECM Regulators
Collagen type V alpha 1	COL5A1	183.5	0.18	0.7056	Collagens
Galectin	LGALS8	38.2	0.18	0.6205	ECM-affiliated Proteins
Thrombospondin-2	THBS2	129.9	0.14	0.7724	ECM Glycoproteins

Collagen alpha-1(IV) chain	COL4A1	160.5	0.13	0.7526	Collagens
TIMP metallopeptidase inhibitor 2, isoform CRA_b	TIMP2	20.2	0.12	0.8623	ECM Regulators
Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	PLOD1	83.5	0.11	0.6978	ECM Regulators
Protein Wnt -5A	WNT5A	24	0.11	0.8737	Secreted Factors
Inter-alpha (globulin) inhibitor H5	ITIH5	78.4	0.1	0.8419	ECM Regulators
Lysyl oxidase homolog 1	LOXL1	63.1	0.1	0.7006	ECM Regulators
Annexin	ANXA7	50.3	0.06	0.9171	ECM-affiliated Proteins
Collagen, type V, alpha 2	COL5A2	86	0.05	0.7910	Collagens
Nidogen-2	NID2	151.2	0.05	0.7525	ECM Glycoproteins
Plexin-B2	PLXNB2	51.7	0.05	0.9385	ECM-affiliated Proteins
COL6A2 protein	COL6A2	47.4	0.03	0.7749	Collagens
Transforming growth factor, beta 2	TGFB2	47.7	0.01	0.9559	Secreted Factors
Latent-transforming growth factor beta-binding protein 4	LTBP4	173.3	-0.03	0.9032	ECM Glycoproteins
Hepatocyte growth factor isoform 1 preproprotein variant	HGF	27.6	-0.04	0.9987	Secreted Factors
Isoform 2 of Fibroblast growth factor 2	FGF2	22.6	-0.05	0.8540	Secreted Factors
Matrix metalloproteinase 14 (membrane-inserted)	MMP14	65.9	-0.06	0.8846	ECM Regulators
Bmp6 protein	BMP6	49.7	-0.06	0.9876	Secreted Factors
Collagen alpha-3(VI) chain	COL6A3	343.5	-0.07	0.8772	Collagens
Fibrinogen-like 2 variant	FGL2	50.2	-0.07	0.8736	ECM Glycoproteins
Protein S100	S100A11	11.7	-0.09	0.9484	Secreted Factors
Microfibrillar-associated protein 1	MFAP1	51.9	-0.14	0.9213	ECM Glycoproteins
EMILIN-1	EMILIN1	106.6	-0.16	0.9838	ECM Glycoproteins
Prolyl 4-Hydroxylase Subunit Alpha 2	P4HA2	60.6	-0.16	0.9983	ECM Regulators
Tubulointerstitial nephritis antigen-like	TINAGL1	52.4	-0.19	0.8576	ECM Glycoproteins
Slit homolog 2 protein	SLIT2	169.8	-0.21	0.8469	ECM Glycoproteins
Matrix Gla protein	MGP	12.3	-0.23	0.9196	ECM Glycoproteins
Immunoglobulin superfamily member 10	IGSF10	290.7	-0.24	0.8191	ECM Glycoproteins
Collagen alpha-1(VI) chain	COL6A1	108.3	-0.31	0.8629	Collagens
thrombospondin-1	THBS1	129.3	-0.31	0.8551	ECM Glycoproteins
Truncated extracellular matrix protein 1	ECM1	45.5	-0.32	0.8235	ECM Glycoproteins
EGF-containing fibulin-like extracellular matrix protein 2	EFEMP2	49.3	-0.32	0.8747	ECM Glycoproteins
Leukocyte elastase inhibitor	SERPINB1	42.7	-0.32	0.7325	ECM Regulators
Collagen alpha-1(III) chain	COL3A1	138.5	-0.34	0.8268	Collagens

Biglycan	BGN	37.6	-0.35	0.8005	Proteoglycans
Elastin	ELN	60.7	-0.38	0.7215	ECM Glycoproteins
Fibulin-1	FBLN1	70.5	-0.38	0.7875	ECM Glycoproteins
Sushi, nidogen and EGF-like domain-containing protein 1	SNED1	152.1	-0.38	0.7174	ECM Glycoproteins
Annexin A11	ANXA11	54.4	-0.42	0.7324	ECM-affiliated Proteins
Serpin B6	SERPINB6	42.6	-0.45	0.6812	ECM Regulators
Collagen, type XVIII, alpha 1	COL18A1	135.4	-0.48	0.6783	Collagens
Dermatopontin	DPT	24	-0.49	0.6914	ECM Glycoproteins
Fibulin-5	FBLN5	50.8	-0.5	0.6061	ECM Glycoproteins
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	61	-0.5	0.6515	ECM Regulators
Serine protease HTRA1	HTRA1	48.4	-0.57	0.5940	ECM Regulators
Tenascin C (hexabrachion)	TNC	240.8	-0.58	0.5852	ECM Glycoproteins
Hemicentin-1	HMCN1	613	-0.61	0.5561	ECM Glycoproteins
EGF-like repeat and discoidin I-like domain-containing protein 3	EDIL3	53.7	-0.62	0.5604	ECM Glycoproteins
Tenascin-X	TNXB	455.9	-0.62	0.5471	ECM Glycoproteins
Galectin-1	LGALS1	14.7	-0.64	0.5298	ECM-affiliated Proteins
Protein-glutamine gamma-glutamyltransferase 2	TGM2	77.3	-0.65	0.5219	ECM Regulators
Laminin subunit beta-2	LAMB2	195.9	-0.66	0.4535	ECM Glycoproteins
Versican core protein	VCAN	372.6	-0.66	0.5105	Proteoglycans
Collagen alpha-1(XV) chain	COL15A1	140	-0.67	0.4738	Collagens
cDNA FLJ78235	CTSB	37.8	-0.68	0.4984	ECM Regulators
Collagen alpha-1(XII) chain	COL12A1	332.9	-0.69	0.4890	Collagens
Laminin subunit gamma-1	LAMC1	177.5	-0.7	0.4905	ECM Glycoproteins
Procollagen C-endopeptidase enhancer	PCOLCE	47.9	-0.73	0.4035	ECM Glycoproteins
Serpin E2	SERPINE2	45.2	-0.73	0.4590	ECM Regulators
Serpin H1	SERPINH1	46.4	-0.75	0.4347	ECM Regulators
Semaphorin-3C	SEMA3C	85.2	-0.76	0.3834	ECM-affiliated Proteins
Laminin subunit beta-1	LAMB1	197.9	-0.8	0.4054	ECM Glycoproteins
Laminin, alpha 4	LAMA4	201.7	-0.81	0.3934	ECM Glycoproteins
Microfibril-associated glycoprotein 4	MFAP4	28.6	-0.82	0.3873	ECM Glycoproteins
Transforming growth factor-beta-induced protein ig-h3	TGFBI	74.6	-0.86	0.3526	ECM Glycoproteins
Latent-transforming growth factor beta-binding protein 2	LTBP2	194.9	-0.87	0.3444	ECM Glycoproteins
Lectin, mannose-binding, 1 variant	LMAN1	57.5	-0.89	0.3737	ECM-affiliated Proteins

Adipocyte enhancer-binding protein 1	AEBP1	130.8	-0.9	0.3522	ECM Glycoproteins
Annexin A4	ANXA4	35.9	-0.93	0.3127	ECM-affiliated Proteins
Galectin	LGALS3	27.1	-0.93	0.3053	ECM-affiliated Proteins
ADAMTS-like protein 4	ADAMTSL4	116.5	-0.94	0.3024	ECM Regulators
Annexin A2	ANXA2	38.6	-0.95	0.2926	ECM-affiliated Proteins
Transforming growth factor beta-1	TGFB1	44.3	-0.95	0.2903	Secreted Factors
CD109 antigen	CD109	161.6	-1	0.2957	ECM Regulators
Protein S100-A10	S100A10	11.2	-1.01	0.2745	Secreted Factors
Prolargin	PRELP	43.8	-1.04	0.2250	Proteoglycans
Host cell factor C2 variant	HCFC2	86.7	-1.04	0.2722	Secreted Factors
Secreted frizzled- related protein 1	SFRP1	35.3	-1.12	0.2032	Secreted Factors
EGF containing fibulin-like extracellular matrix protein 1 isoform 2	EFEMP1	45.7	-1.13	0.2290	ECM Glycoproteins
Cathepsin D	CTSD	44.2	-1.14	0.2155	ECM Regulators
Glypican-4	GPC4	62.4	-1.14	0.1843	ECM-affiliated Proteins
Interleukin-25	IL25	20.3	-1.16	0.1929	Secreted Factors
Glypican-6	GPC6	62.7	-1.21	0.1613	ECM-affiliated Proteins
Matrilin 2	MATN2	104.9	-1.23	0.1701	ECM Glycoproteins
Cathepsin Z	CTSZ	33.8	-1.32	0.1451	ECM Regulators
Plexin domain-containing protein 2	PLXDC2	59.5	-1.35	0.1219	ECM-affiliated Proteins
Host cell factor 1	HCFC1	208.6	-1.47	0.1037	Secreted Factors
Collagen alpha-1(VIII) chain	COL8A1	73.3	-1.51	0.0821	Collagens
Lysosomal protective protein	CTSA	54.2	-1.77	0.0484	ECM Regulators
Chondroitin sulfate proteoglycan 4	CSPG4	250.4	-1.89	0.0202	ECM-affiliated Proteins
Microfibrillar-associated protein 2	MFAP2	20.8	-2.15	0.0144	ECM Glycoproteins
Collagen alpha-1(XIV) chain	COL14A1	193.4	-2.31	0.0037	Collagens

MW – molecular weight. Abundance ratio calculated in Proteome Discoverer software as the median of all possible pairwise peptide ratios calculated between replicates of all connected peptides. PD uses the ANOVA method for calculating the p-value with Benjamini-Hochberg correction for the adjusted p-value.

6.4.8 Secreted Protein Acidic and Cysteine Rich (SPARC) modulates basal cell migration in vitro

Next, we sought to functionally validate key matrix protein candidates identified as being upregulated in IPF CDMs by MS-proteomics. Our hypothesis was that at a higher concentration, one of the proteins identified in the IPF CDMs would influence basal cell migration. To achieve this, migration of basal cells from healthy donors (n=3) was evaluated using live-cell imaging on escalating concentrations of IGFBP-5, GAS6, FN1 and SPARC in a 2D culture environment. The concentration range of each protein tested were determined using previously published literature[263,443-448]. Limited cell migration was observed on an uncoated culture plate, therefore collagen I – previously demonstrated to be a component of the CDMs - was used as a substrate that would promote migration. By testing the influence of each protein on migration at increasing doses, it was evident that IGFBP5, GAS6 and FN1 did not change basal cell migration on collagen I (Fig. 6.18 A-C). SPARC at a concentration of 20µg/ml caused significant restriction in basal cell migration (Fig. 6.18 D-F), toward a level of cell displacement seen on an uncoated plate.

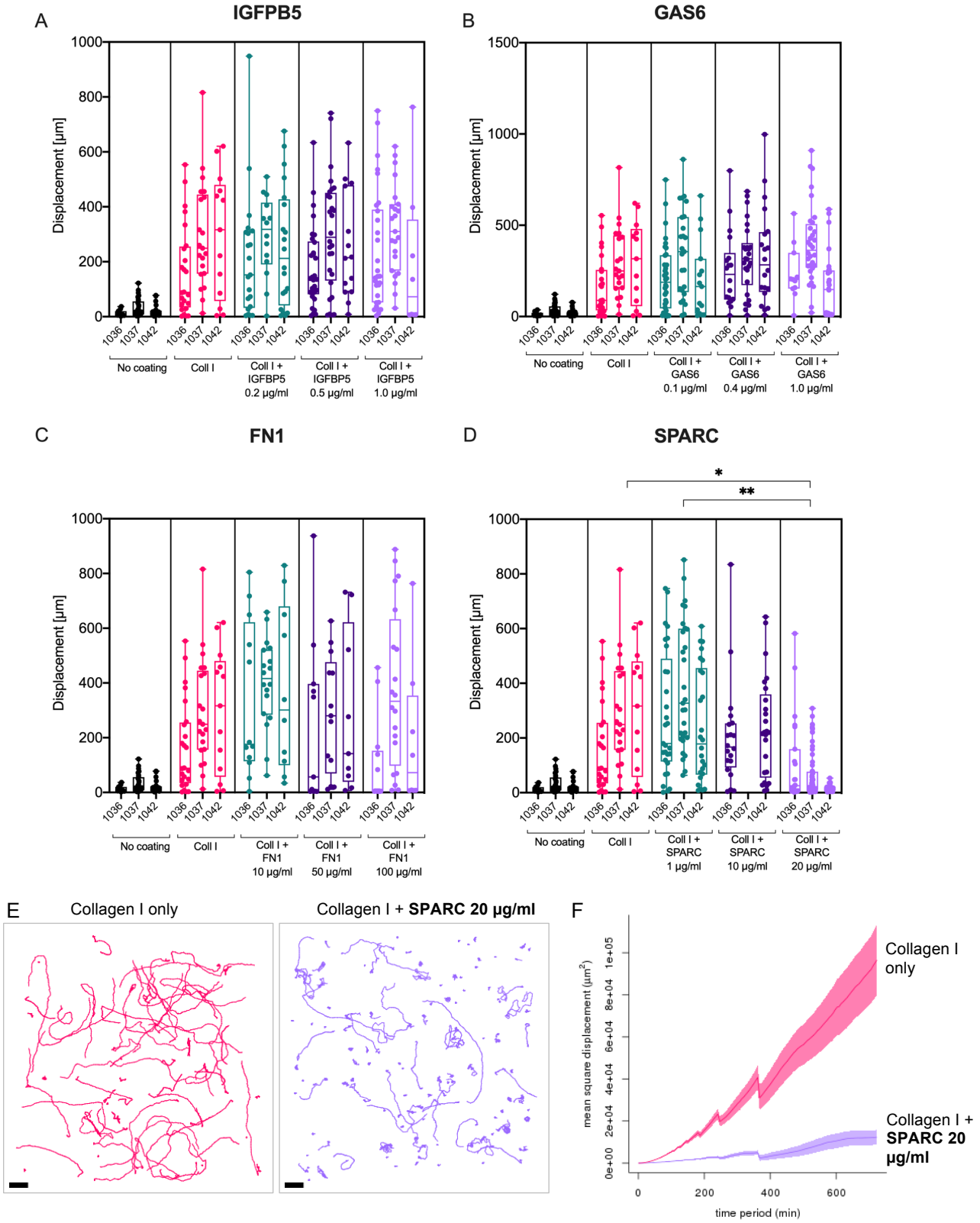


Figure 6.18 SPARC restricts migration of healthy basal cells at higher concentrations. Live cell imaging was used to track migratory dynamics of healthy basal cells (n= 3 donors). Cell migration was captured using a Zeiss Axio Observer Inverted Widefield Microscope and data analysed using Nikon NIS spot detection and semi-automated cell tracking functions. Data shown as box and whisker dot plots quantifying displacement of cells from starting position plotted for basal cells cultured on (**A**) IGFBP5, (**B**) GAS6, (**C**) FN1 and (**D**) SPARC. (**E**) Cell tracks and (**F**) mean squared displacement for all donors combined on collagen I alone compared to collagen I and SPARC 20µg/ml (Plotted using motilitylab.com). A nested one-way anova and Tukey's multiple comparison test used for statistical analysis, **p<0.01, *p<0.05.

6.5 **Discussion**

Ectopic KRT5+ basal cells populate remodelled and fibrotic parenchymal lung tissue in IPF[36,104,187,205,431], but it is unclear how or why this occurs. The principle aim of this chapter was to investigate how changes in the extracellular matrix environment of the distal lung may modulate the activity of KRT5+ basal cells - specifically their migration - which may provide insight into the pathological mechanisms underlying IPF. Collagen structural organisation was examined in defined regions of control and IPF lung tissue sections using SHG imaging. Then, to recapitulate disease-specific changes in ECM and delineate the properties which influence basal cell migratory behaviour, functional *in vitro* migration studies using 3D cell-derived matrix from primary human lung fibroblasts was combined with MS-based proteomics and immunofluorescence imaging.

Other studies have demonstrated by IHC or IF of distal lung tissue specimens that KRT5+ cells accumulate in the fibrotic distal lung, but in this Chapter the corresponding structural properties of the supporting collagen-rich extracellular matrix are described. There is a relative paucity of studies investigating collagen organisation in the human lung. SHG microscopy is a powerful, label-free technique used to assess collagen organisation, that avoids the need for extensive tissue processing required for staining protocols, which may disrupt the ECM architecture[439,440]. Despite the tissue heterogeneity characteristic of IPF, the majority of SHG studies summarise collagen organisation in the whole lung[449-451]. In line with traditional approaches using Sirius red staining[452], SHG imaging has demonstrated a greater density of collagen in IPF lung tissue with a more disorganised structure[449]. Coaligned, wavy fibres were described in IPF lung tissue[449]. Measuring SHG emission directionality as the forward/ backward ratio (F_{SHG}/B_{SHG}) provides information about the sub-resolution microstructure of collagen fibres; the fibril size, packing density and order of fibril assembly[453]. A lower F_{SHG}/B_{SHG} ratio was reported in IPF lung tissue compared to control, suggestive of smaller and less organised collagen fibrils[449]. In agreement with this, Kottman et al. demonstrated that alveolar regions with preserved architecture, fibrotic foci and honeycomb areas of the IPF lung all shared similar microstructural properties as quantified by F_{SHG}/B_{SHG} ratio, which was significantly lower than that

found in healthy control tissue[454]. In contrast to these studies, Tjin et al. showed an increased F_{SHG}/B_{SHG} ratio in IPF lung parenchyma, reflecting increased collagen fibre maturity and thickness[450]. The authors' suggested image sampling and analysis protocols may have led to this discordance with previous findings. SHG texture analysis using a grey level cooccurrence matrix (GLCM) has been employed to study fibrillar collagen organisation in the conducting airways of asthmatics, showing collagen fibres are disorganised with higher entropy, and are less uniform with lower IDM[455].

In this Chapter, a regional approach was taken to analyse collagen structure in the IPF distal lung using a combination of SHG imaging and GLCM texture analysis. Analysis of equivalent regions in the healthy and IPF lung, revealed greater coherence in fibre arrangement in the peribronchiolar region compared with the perivascular regions. Interestingly, in the IPF lung, the structure of collagen was distinct in parenchymal fibrotic regions compared to HC regions. GLCM texture analysis statistics indicated greater disorder in fibrotic regions, mirroring results obtained from the perivascular regions of the healthy and IPF lung. Conversely, the collagen fibres were more orientated and coherent in HC regions, echoing that found in the peribronchiolar regions. Consistent with the notion that there are regional differences in the characteristics of collagen in the lung, an electron microscopy study demonstrated greater fibril thickness in parenchymal interstitial collagen regions compared to basement membrane regions in IPF[456]. Dysregulated collagen cross-linking has been shown to be a key determinant of increased tissue stiffness in the IPF lung[457]. Therefore, it is likely that the regional changes in collagen geometry identified in this Chapter may lead to disruption in local lung biomechanics and impair function.

A positive association was found between KRT5+ basal cell numbers and collagen density and fibre orientation. As anticipated, KRT5+ cells were in highest abundance in the airways overlying the basement membrane, but also in HC regions where collagen organisation was similar. Interestingly, an early IHC study reported positive staining for tenascin-C (TN) – a glycoprotein usually present in the basement membrane of larger airways – in the subepithelial extracellular matrix of HC in IPF[458],

suggesting there are molecular similarities in ECM composition in these regions too. The close spatial relationship between KRT5+ cells and α-SMA myofibroblasts visualised by IF staining in fibrotic HC regions of the IPF lung suggests potential epithelial cell - mesenchymal interactions may be responsible for driving differences in ECM composition, as has been previously reported in co-culture models of the embryonic gut[459], and alveolar niche[286].

The regional approach adopted in this analysis acknowledges the inherent heterogeneity characteristic of IPF lung tissue and provides valuable novel insight. Taken together these findings suggest matrix is deposited and assembled in the distal lung in a way that generates regional differences in collagen density and organisation. In the IPF lung, collagen that accumulates in the fibrotic areas is structurally disordered and associated with lower numbers of KRT5+ basal cells arranged as single cells or clusters. In contrast, in HC regions, where collagen appears more organised, greater numbers of KRT5+ accumulate in an arrangement comparable to that seen in the normal distal airway. Intriguingly, collagen-producing subpopulations have been identified by scRNA-seq residing in different anatomical locations in control and fibrotic lungs[146]. It is plausible that these may be responsible for governing the different regional collagen organisation that were observed by SHG.

The data intimated that regional differences in matrix topography may play a role in dictating the spatial distribution of KRT5+ basal cells in the IPF lung. Highly aligned, wavy collagen fibres in a fabricated 3D biomimetic scaffold - representative of SHG microscopy images of IPF collagen - were shown to enhance migratory dynamics, specifically speed and straightness of human lung fibroblasts[460]. To investigate the extent to which the matrix can influence cell behaviour, the migratory dynamics of KRT5+ basal cells were studied on a 2D tissue culture surface coated in ECM proteins previously identified in the fibrotic lung[34] – collagen I, III, IV, versican and fibronectin. Collagen I, III and fibronectin enhanced cell spreading and migration compared to collagen IV and versican. This recapitulated some of the early findings using bovine bronchial epithelial cells in modified Boydon chambers, showing type I collagen enhanced chemotactic migration of epithelial cells, yet type IV collagen and laminin – both basement membrane components - restricted migration[461]. There are very few

studies which have studied single-cell migration of human airway epithelial cells on different matrix substrates; many focus on migration of epithelial cell sheets during wound healing assays[462-464]. The specific role of ECM components such as fibronectin, have also been studied in the context of collective airway epithelial cell migration during wound healing[448].

It should be noted that despite seeding basal cells at the same density, there was significant variability in cell number and tracks obtained per field of view imaged, both between subjects and on different ECM substrates. With the exception of fibronectin, a sufficient number of cells were tracked per donor on each substrate to provide meaningful and significant results in a nested analysis of the data. A second limitation was the accuracy of the automated tracking software at higher levels of cell confluency. A 12- hour experimental time period was chosen for this reason because beyond this, cells became too confluent to reliably track.

There were no significant basal cell- intrinsic, disease-driven differences in migration on any of the individual ECM substrates. However, emerging data suggests functional differences do exist in basal cell subpopulations in IPF[199]. Single-cell analyses of epithelial cells in the IPF distal lung demonstrated enrichment of genes linked to cell migration[189], and increased KRT5+ basal cell expression of *MEG3*, a long noncoding RNA, which enhanced migration when over-expressed in cultured epithelial cells[465].

The ECM is a complex mixture of collagens, proteoglycans and gycloproteins[153], that is not replicated *in vitro* by using individual collagen proteins (which may not be folded or organised correctly). The majority of interstitial collagen is secreted and organised into sheets and cables by fibroblasts[466]. To better understand basal cell migratory behaviour *in vitro* in a microenvironment which mimics the *in vivo* 3D matrix architecture and molecular composition of lung tissue, fibroblast- derived matrix was generated[262]. Cell-derived matrices (CDMs) produced by fibroblasts cultured at high density have been increasingly employed to study a range of important biological questions, such as how the normal ECM regulates cancer cell proliferation[467]. A key advantage of this *in vitro* system is that patient-derived fibroblasts can be used, providing a model which better recapitulates *in vivo* protein composition[262,468].

In health, basal cells are adherent to a basement membrane, interact with adjacent cells via intercellular junctions, and display apical-basal polarity. When we seeded KRT5+ basal cells on fibroblast generated 3D CDMs, they demonstrated phenotypic features in keeping with epithelial-mesenchymal transition (EMT), a process important in tissue development, repair, fibrosis and cancer[469]. During EMT epithelial cell-cell junctions are dismantled, cells lose polarity and acquire an enhanced migratory and invasive phenotype[470]. Crucially, epithelial cells display remarkable phenotypic plasticity and EMT is often a transient and reversible process[469]. Healthy and IPF basal cells exhibited very similar patterns of migration through an IPF 3D matrix. Within cellular populations from the same donor there was some heterogeneity in migratory behaviour. A mesenchymal mode of migration[437] was observed; cells adopted an elongated morphology with lamellipodia and migrated in a directed manner along matrix fibres. Other cells remained relatively fixed in one position - presumably with stronger adhesions to the ECM - but projected filopodia to explore the local matrix microenvironment. Direct interactions between KRT5+ basal cells and fibronectin fibres were demonstrated by high-resolution IF microscopy. Cell- matrix adhesions are important in modulating 3D cell migration[471,472], and local microenvironmental variations in fibre stiffness in a 3D matrix altered focal adhesion dynamics and fibroblast motility[473]. It is likely that microenvironmental differences in our CDMs were responsible for the distinct migratory patterns observed. The data presented here is significant as represents the first report of KRT5+ basal cells migrating through an IPF fibroblast-derived 3D CDM, and therefore provides valuable, translatable insight into the dynamics of basal cell movement in the fibrotic lung.

Crucially in both the 2D and 3D migration models, there were no significant migratory differences between healthy and IPF basal cells. In agreement with this there were very few gene expression differences between healthy and IPF basal cells. Crucially however, the properties of the matrix microenvironment (2D vs. 3D) played a prominent role in determining gene expression changes in basal cells. Interaction of either healthy or IPF basal cells with matrix derived from IPF fibroblasts induced upregulation of genes *SERPINE1, COLA1A2, and MMP9*, all of which have been implicated in fibrogenesis[53,474,475] and mechanisms regulating cell migration[373,476,477]. Metalloproteinases (MMPs) produced at the cell-surface such as membrane-type-1

matrix metalloproteinase (MT1-MMP/MMP-14) play an important role in local proteolytic remodelling of the 3D matrix to facilitate invasive single- cell migration[478]. Using human bronchial epithelial cells in an *in vitro* wound-repair model, MMP-9 was identified in lamellipodia at the leading edge of migrating cells and was associated with increased cell migration speed[477].

Interestingly, compared to IPF CDM, a larger number of wound healing proinflammatory and growth factor genes were upregulated on the 2D collagen substrate, suggesting basal cells were 'activated' to a greater degree on this surface. Also noteworthy was the upregulation of integrin subunit genes such as *ITGA3* on 2D collagen compared to IPF CDM. Integrins are heterodimeric receptors composed of a combination of α- and β- subunits, which span the cell membrane linking the cytoskeleton to ECM[479]. Integrin-mediated adhesion plays an essential role in directional cell migration driven by ECM ligand density or 'haptotaxis', and stiffness gradient or 'durotaxis'. Airway epithelial cells express a host of integrins with different specificities for ECM substrates[480,481]. Various integrin subunits play a prominent role in airway epithelial cell migration in the context of the airway wound repair process, including the collagen/ laminin - binding integrin α3-subunit[463,464].

Collectively, this data implies that the properties of the ECM play a more significant role in determining the molecular state and migratory capacity of basal cells than cell-intrinsic factors. To take this concept further, basal cell migration was studied in fibroblast generated CDMs from IPF and control subjects. Interestingly, greater restriction in cell displacement was observed in basal cells migrating through IPF CDMs compared to controls. A variety of ECM features govern 3D cell migration including molecular composition, topography, stiffness and cell confinement[437]. MS-based proteomics was utilised to define the molecular composition of IPF and control CDMs. Remarkably, there were distinct differences in the matrisome of CDMs generated over 21 days by HLFs from IPF patients compared to controls. Importantly, proteins that were significantly upregulated in our IPF CDMs have been implicated in the pathogenesis of the disease including IGFBP-5[482], GDF-15[483], WNT5B[484], LOXL2[450], GAS6[485], FN[34] and SPARC[287,486]. Pertinent to findings from our functional

in vitro studies, many of these proteins play a role in ECM organisation, cell adhesion and cell migration.

IGFBP-5 - a protein significantly overexpressed in IPF CDMs compared to control CDMs - was present at high levels in IPF lung tissue[482], and *in vitro,* induced production of ECM components by normal lung fibroblasts[482], and the migration of peripheral blood mononuclear cells and human lung fibroblasts[487].

GAS6, is a secreted protein with high affinity for AXL which belongs to the TAM (TYRO3, AXL and MER) receptor tyrosine kinase family[488]. AXL is expressed by human TP63+ airway basal cells, and the number of AXL-expressing basal cells was significantly increased in COPD small airways compared to healthy controls[489]. GAS6 has multiple biological functions[490], and its expression was increased in IPF lung tissue and cultured fibroblasts compared to controls[485]. Targeting the pathway with small-molecule TAM inhibitors reduced IPF fibroblast invasion and activation[485]. AXL activity was increased in the hyperplastic epithelial cells adjacent to fibrotic areas in IPF, and AXL kinase inhibition was associated in decreased migration of human lung multipotent cells, and acquisition of epithelial junctions[491]. Airway basal cell proliferation was triggered in response to the recognition of apoptotic cells by AXL receptor[489].

Fibronectin (FN) polymerisation is critically important for the deposition and fibrillar organisation of collagen I and thrombospondin-1 in the ECM[492]. An intact FN matrix can be assembled rapidly by α-SMA myofibroblasts[493]. FN has been shown to promote the migration of small airway epithelial cells at low concentration, but at higher concentrations, results in a dose- dependent reduction in cell migration[494]. Therefore, the higher expression of fibronectin in our IPF CDMs may be a critical factor restricting basal cell movement.

Also important in regulating collagen deposition in the ECM is the multifunctional glycoprotein SPARC – also known as osteonectin - which has been localised by IHC to the IPF fibroblast foci[458]. SPARC was secreted at significantly higher levels by cultured IPF fibroblasts than control and promoted AT2 cell migration in a wound

healing assay[287]. In this Chapter I demonstrate for the first time that higher concentrations of SPARC significantly restrict the migration of basal cells on collagen I. The observed difference in effect on basal cell migration compared to that previously described on AT2 cells[287], could indicate SPARC influences migration in a cell-type specific manner. These results require validation with a greater number of healthy donors to confirm the effect.

Therefore, the proteins differentially expressed in IPF CDMs may directly modulate basal cell migration by altering the matrix configuration, adhesive properties, stiffness or even through signalling pathways which directly change cell migratory phenotype. Almost half of the proteins detected in our IPF fibroblast generated CDMs could be identified in published IPF lung proteomic datasets[157,495-497]. Discordance in protein identification and expression between IPF CDM and lung tissue datasets may be due to numerous cell types contributing to the matrix *in vivo*, or differences in detergents used for protein extraction[498], high-throughput technique for protein characterisation or bioinformatic pipeline.

It should be noted that the biomechanical and biophysical properties of CDMs such as stiffness and elasticity[499], were not measured in the experiments described in this Chapter. These properties can also influence cell migration[500], and therefore could be evaluated in future work.

6.6 **Conclusion**

The data described in this chapter provides novel evidence that KRT5+ basal cell migratory behaviour is modulated by properties of the extracellular matrix in the IPF distal lung. On this basis, the Chapter 6 hypothesis was accepted. Both the structure and composition of the matrix play a critical role. Interestingly in an IPF fibroblast-secreted, cell-free 3D matrix, basal cells movement is more restricted than in a matrix derived from control fibroblasts. Higher levels of glycoprotein, SPARC were detected in IPF CDMs and live- cell tracking suggests this may play a role in restricting basal cell migration.

Regional differences in collagen organisation and the association with basal cell distribution suggest reciprocal epithelial-mesenchymal interactions may occur; fibroblast generated ECM modulates the behaviour of basal cells and in turn basal cells may regulate ECM production by fibroblasts, but this hypothesis will require further investigation.

CHAPTER 7

General discussion and future work

7 General discussion & future work

The overarching aim of this project was to understand how epithelial cells interact with their local tissue microenvironment, and how this may drive the pathogenesis of IPF. Animal models have helped develop our understanding of mechanisms that drive disease, but none recapitulate progressive fibrosis or the histological UIP pattern characteristic of IPF[213]. Thus, there is an unmet need for experimental work using primary cells and tissues from well-defined cohorts of patients with this incurable and ultimately fatal, fibrosing lung disease[268]. I have explored the molecular phenotype and function of airway epithelial cells in IPF patients and healthy subjects. I have characterised the cellular landscape of the airway mucosa and exposed epithelial-immune crosstalk in IPF. I have examined interactions between basal epithelial cells and the extracellular matrix, to gain insight into the pathological tissue remodelling that occurs in the IPF distal lung.

To date, mechanistic studies of IPF have focused on the role of alveolar epithelial cells which are co-located with areas of fibrosis[501], rather than proximal airway epithelial cells. Greater accessibility to explant lung tissue samples from diseased donors has fuelled studies of the distal lung environment in end-stage IPF[53,163]. However, the association of the airway mucin *MUC5B* variant rs35705950 with disease susceptibility[87], and the expansion of basal cell populations in the distal lung[205], have ignited interest in the potential role of airway epithelial cells in this devastating disease[210]. Indirect evidence suggesting an imbalance in airway epithelial immune homeostasis in IPF has arisen from studies reporting an altered microbiome and significantly increased bacterial burden in disease[31,73,74]. It remains unknown whether the interplay between the airways and alveoli in IPF is protective or pathogenic[298].

For the experimental work carried out in this project, airway epithelial cells were harvested via bronchoscopy from a cohort of newly diagnosed IPF patients. Whilst the radiological extent of fibrosis varied between cases, these patients were all able to undergo bronchoscopy and had, on average, preserved lung physiology. Therefore, this thesis provides novel sequencing data on a cohort of IPF patients with less advanced disease than previously studied[53,145,163]. Greater understanding of

pathogenic events taking place during the earlier phases of disease may translate into more tractable therapeutic targets[49]. From a diagnostic and disease biomarker perspective, findings from easily accessible airway- derived samples (rather than peripheral lung tissue samples) are highly relevant for future clinical trials in IPF.

To gain a comprehensive insight into the full spectrum of disease and study remodelling in the distal lung, human tissue samples from advanced cases were used to study basal cell-matrix interactions.

7.1 Airway epithelial immune homeostasis is disturbed at a molecular level in IPF

The data presented in Chapter 3 demonstrate that fully- differentiated IPF airway epithelial cells cultured ex vivo at ALI, do not exhibit a distinct pathological phenotype. Instead, I found they provided comparable barrier function and mucociliary clearance to healthy airway epithelial cells. These observations were made under homeostatic growth conditions, without specific stimulus. Stimulation with a relevant exposure is critical in eliciting differences in immune response and function between cells from diseased and healthy donors in other settings. For example, rhinovirus (RV)-16 infection of asthmatic PBECs revealed an impaired interferon-β response compared to healthy PBECs[277]. The specific environmental insults which catalyse IPF disease pathogenesis remain unknown, making these experiments are more challenging. An influenza A model of injury was studied in this thesis, but equally relevant and worthy of future investigation would be the IPF airway epithelial response to cigarette smoke or pollutants. Lack of any statistically significant differences in the changes observed may be attributable to inter-individual variability and an insufficient number of donors tested. Nevertheless, these findings align with our expectations because defects in epithelial barrier and ciliary function would likely result in chronic airway inflammation and recurrent infection which are not typical features of IPF. Airway epithelial cells from COPD[272] and asthma[277] retain disease-specific phenotypes in cultured systems, but it appears the IPF airway epithelial phenotype is more subtle in culture – at least under homeostatic conditions or following influenza virus infection.

The *in vitro* microenvironment is artificial and does not represent the *in vivo* airway epithelial niche. Airway epithelial cell phenotype and function can be shaped by interacting immune cells, the microbiome, environmental conditions (temperature differences, bidirectional airflow) and inhaled antigens[52]. Thus, the situation *in vivo* is more complex and studying human cells in this context (or as close to it as possible) takes into account these critical microenvironmental signals. To capture the molecular state of cells in the airway mucosa I performed RNA-sequencing of airway brushings. It became evident that these brushings sampled not only epithelial cell populations but also closely associated immune cell populations including macrophages and T cells, offering valuable perspective on the local immune landscape. Bulk RNA-sequencing of brushes demonstrated significant upregulation in chemokine genes *PPBP/ CXCL7, CCL13* and *CCL24* in IPF patients compared to controls. These genes showed significant correlation with the extent of disease in the IPF cohort quantified by composite physiological index. Therefore, whilst IPF proximal airway epithelial cells appeared normal in both microscopy and *in vitro* functional studies, a state of altered innate immune signalling was evident by RNA-seq. Luzina et al. showed that akin to fibrotic areas, the normal-appearing areas of IPF explant lung had a distinct transcriptomic profile to that of healthy control lung, suggesting involvement of the entire lung in the disease process[331]. Similarly, bulk RNA-seq analysis of nasal epithelial samples demonstrated upregulation of inflammatory pathways such as interferon signalling in IPF compared to age-matched controls[332]. Airway-wide transcriptomic changes may exist in IPF and serve as a useful biomarker of disease. This aligns with the “field of injury” concept in which lung cancer-specific molecular changes can be detected in cytologically normal airway epithelial cells, potentially due to airway-wide exposure to cigarette smoke[502]. The genes upregulated in our proximal airway samples differed to those described in the nasal epithelial samples by Sala et al.[332], likely reflecting regional differences in epithelial function known to exist in the healthy airways[179].

Through single-cell transcriptomic analysis it was possible to resolve discrete epithelial and immune cell populations in the airway brushings. Basal cell subclustering exposed a subpopulation with a transcriptomic profile comparable to that previously described as ‘activated basal cells’[199], which was expanded in IPF. This basal cell subtype was

enriched in genes linked to pathways involved in the integrated stress response and regulation of fibroblast activity. Further work should seek to explore whether there is overlap in the transcriptomic signature between these cells and the expanded KRT5^{+} populations in the distal lung in IPF. Carraro et al. took a similar approach and discovered enrichment of a "secretory-primed" basal cell subpopulation in the IPF distal lung[199].

Of all the airway epithelial cell populations, secretory cells exhibited the greatest gene expression changes in IPF compared to control. Interrogation of ligand-receptor pairs in the dataset using the CellPhoneDB tool[261], predicted increased interactions between secretory cells and airway macrophage populations in IPF compared to healthy controls. Whilst direct interactions are challenging to prove in studies using human samples, I demonstrated a close spatial relationship between CD206^{+} airway macrophages and airway epithelial cells by immunofluorescence imaging of precision cut lung slices. Live, *in situ* confocal microscopy of mouse lungs has elegantly revealed that alveolar macrophages directly attach to the alveolar epithelium via connexin 43 (Cx43)-containing gap junction channels[181]. Critically this close interaction was immunomodulatory and suppressed cytokine release associated with lipopolysaccharide (LPS)-induced inflammation[181]. The extent to which this direct epithelial-macrophage intercommunication occurs in the human airway microenvironment, and how this differs in health and disease, warrants further research.

Epithelial cell-macrophage co-culture models have been utilised to study the innate immune response to ozone (O_3) pollutant exposure[503]. Bauer et al. showed macrophages co-cultured with a human bronchial epithelial cell line (16HBE) displayed decreased phagocytic capacity and enhanced CD206 and scavenger receptor A1 (SRA1) expression after O_3 exposure[503]. This approach could be used to disentangle epithelial-derived signals that modulate macrophage phenotype and response, in IPF airway samples.

Evidence has emerged of a reciprocal interaction in which macrophage populations can influence epithelial cell behaviour and vice versa. Following epithelial injury, lung

macrophages were shown induce proliferation of epithelial cells via a Trefoil factor 2 (TFF2)/ Wnt axis[504]. In a separate study, an increase in alternatively activated M2- like macrophages - some of which were monocyte derived - were found in the mouse trachea following polidocanol injury[384]. The data suggested CCR2+ airway-associated myeloid populations influence basal cell growth and mediate epithelial repair following injury[384]. These processes are likely to be highly relevant in the pathogenesis of IPF, and in the future my dataset could be used to mine these interactions.

7.2 Airway macrophage populations are expanded in IPF and exhibit a unique transcriptomic signature

Airway macrophages (AMs) are a key cellular component of the mucosal ecosystem and reside in close association with epithelial cells[169]. The data presented in Chapter 4 demonstrated $CD206^{+}$ AMs from newly diagnosed cases of IPF, herald a transcriptomic profile distinct to that of healthy volunteers. The IPF AM transcriptome was enriched in genes linked to the pathogenesis of pulmonary fibrosis. Compared to healthy control AMs, there was upregulation of genes associated with cell cycle regulation, matrix organisation and cellular movement. GM-CSF (CSF2) was the upstream regulator predicted to drive these changes in gene expression. This cytokine plays a pivotal role in AM homeostasis and can induce proliferation, differentiation and activation of myeloid cells in inflamed tissue[505]. Although not proven by this work, airway epithelial cells may represent the cellular source of this mediator in IPF. Interestingly, GM-CSF was produced by bronchial epithelial cells *in vitro* following exposure to diesel exhaust particles[506]. Future work should aim to determine if GM-CSF is produced by IPF airway epithelial cells, and whether this mediator is responsible for the transcriptomic changes observed in IPF AMs.

Bulk transcriptomic analysis provided global insight into the phenotype and function of AMs expressing marker CD206 but did not capture the subtleties of AM heterogeneity in the IPF airways. In our laboratory, multiparametric flow cytometry has revealed a distinct macrophage subset in the IPF airways, characterised by impaired iron handling[30].This indicates there is functional heterogeneity in AMs present in the airway which requires delineation. The bulk RNA-seq data generated in my study was from BAL samples, and therefore reflects luminal macrophage state in both the airway and

alveolar compartments. Bronchial brushings were thus acquired to facilitate exploration of the proximal airways at single-cell resolution (Chapter 5).

Single-cell RNA-seq studies have exposed macrophage heterogeneity in fibrosis affecting other organ systems including the liver[507] and kidneys[508]. Proximal airway macrophage populations have been characterised in healthy volunteers[179], asthmatics[168], and patients with cystic fibrosis[167], but until now, have not been studied in IPF. I identified three discrete clusters of transcriptionally distinct macrophages in the healthy and IPF proximal airways, likely representing a spectrum of macrophage states[382]. The largest population of AMs (cluster 6) had an M2- like immunophenotype, expressed genes of mature macrophage function, and were expanded in IPF. There were two smaller populations: “monocyte-derived macrophages” in cluster 11 which co-expressed macrophage and monocyte markers (making annotation challenging), and “proliferating macrophages” cluster 12 which were characterised by an abundance of cell cycle genes. Genes linked to cell adhesion and migration pathways were also enriched cluster 11 supporting the idea that these are a recruited AM population present in health and disease. This aligns with a previous study from our laboratory which demonstrated in healthy individuals, AMs are monocyte- derived[174]. At other epithelial surfaces for example the intestinal mucosa, murine studies revealed macrophages are constantly replaced by circulating Ly6C^{hi} monocytes to maintain local homeostasis[509]. Therefore, it is likely that the populations I have described in this work represent different stages of monocyte – to – macrophage differentiation. The relationship between these populations requires further delineation. This could be achieved using trajectory analysis, as was nicely demonstrated by Schupp et al. who studied myeloid populations in cystic fibrosis sputum samples[167]. Intriguingly, I discovered a strong association between the ‘proliferating macrophage’ cluster and disease severity characterised by composite physiological index (CPI). It is conceivable that a higher proportion of proliferating macrophages could differentiate into AMs with a more fibrotic phenotype, which participate in the disease process and predict more severe disease i.e. a higher CPI. In contrast to previously published studies[53,163,164], I did not find a unique IPF-specific macrophage population at the clustering resolution level I applied. Rather, I found an IPF-driven expansion of pre-existing AM populations.

Questions remain over the relationship between alveolar and airway macrophage populations. The microenvironment of the airway differs from that of the alveoli, and this is likely to influence the phenotype of macrophages, which are known to exhibit marked plasticity[170]. In mice, scRNA-seq has demonstrated airway-associated macrophages are molecularly distinct from alveolar macrophages[384]. A study profiling populations of mouse macrophages from different tissues demonstrated that the environment is pivotal in regulating the chromatin landscape through tissue-restricted transcription factor expression[165]. Epigenetic modification by environmental signals determined macrophage identity and function[165]. A similar experimental approach combining RNA-seq, chromatin immunoprecipitation (ChIP-seq) and an assay for transposase-accessible chromatin (ATAC-seq) could be employed in the future to study how changes in the IPF airway and alveolar environment influence AM identity.

There are distinct alterations to the microbiome and a higher bacterial load in the IPF airway compared to healthy controls[31]. Interestingly in the gut, activation of epithelial toll-like receptor 4 (TLR4) by lipopolysaccharide (LPS) derived from the microbiota, upregulated CCL2 resulting in recruitment of monocyte-like macrophages[510]. The extent to which microbial dysbiosis in the IPF airway modulates epithelial and macrophage activity, and how this alters disease course, warrants further investigation.

There is currently a lack of consensus on the specific marker genes required to assign identities to the cell clusters resolved using scRNA-seq. As such, a direct comparison of the AM subsets described in this thesis with those previously reported in the lung[53,163,168,179,511] is challenging. The Human Lung Cell Atlas project endeavours to provide clarity on cellular identities using single-cell molecular studies[216]. Integration of multiple scRNA-seq datasets from different tissues and patient cohorts has been achieved in a meta-analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) entry genes[512].

To further delineate cellular relationships and spatial organisation in the airway microenvironmental niche, spatial transcriptomic tools such as Slide-seq[513] could be used in future studies of the IPF lung.

7.3 Cues from the fibrotic extracellular matrix dictate basal cell activity in the distal lung

The extracellular matrix (ECM) provides structural support to the lung, maintaining the tissue architecture and biomechanical properties necessary for optimal function[411]. Crucially, the lung ECM also provides instructive signals to the cells closely associated with it, influencing cell phenotype and behaviour[514]. There is a growing body of evidence suggesting the ECM regulates profibrotic processes in IPF[515].

In Chapter 6, I used state-of-the-art imaging techniques to better characterise the ECM surrounding expanded populations of KRT5$^+$ basal cells in the fibrotic lung. The purpose of this work was not to delineate the cell-of-origin of these aberrantly located basal cells - which is debated[208] - but rather, to understand how they interact with the local matrix microenvironment. SHG imaging and GLCM texture analysis revealed regional differences in collagen organisation between fibrotic areas and honeycomb cysts, which influenced the number and configuration of KRT5$^+$ basal cells. In some areas of the imaged lung, a close relationship was observed between KRT5$^+$ basal cells, α -SMA$^+$ fibroblasts and collagen (SHG signal). Basal cell- stromal crosstalk warrants further investigation as this may drive some of the pathological remodelling that occurs in this disease.

Ultimately, these imaging studies provide a valuable, ‘snapshot’ perspective on the spatial reorganisation of basal cells in the fibrotic niche. Tissue remodelling is a dynamic process[127], and as such these studies do not address the temporal question of whether the emergence of KRT5$^+$ basal cells precede or follow matrix remodelling. Insights have been drawn from live-imaging and lineage-tracing studies in murine influenza challenge models[104]. Vaughan et al. showed mobilisation of progenitor cells in the injured distal lung[104], therefore I explored the role of ECM changes on the migratory dynamics of human KRT5$^+$ basal cells *in vitro*. Cell-derived matrices (CDMs)

from primary human lung fibroblasts were used to recapitulate the *in vivo* microenvironment of the distal lung. These data revealed that the properties of the matrix had a profound influence on the molecular state and migratory capacity of basal cells. IPF CDM's restrained the migration of KRT5+ basal cells relative to control CDM's. An ECM glycoprotein, SPARC, was more abundant in the IPF matrices relative to control and restricted basal cell migration at higher concentrations. The biophysical properties of the ECM determine interstitial cell migration[516]. For example, adult bovine meniscal tissue contains dense, highly organised collagen fibres that are more prohibitive to interstitial cell migration than in fetal tissue[517]. After treating adult meniscal tissue with collagenase to reduce density and stiffness, cells migrated to a similar degree as through immature, fetal tissue[517]. Addition of enzymes to selectively degrade different components of the matrix may be a useful future approach to assess the effect of modifying the structural properties of IPF CDMs on cell migration. The IPF matrisome was distinct from that of control fibroblasts, being enriched in proteins involved in ECM organisation, cell adhesion and cell migration. Further *in vitro* studies to assess the impact of individual proteins on basal cell behaviour are required. One approach would be through siRNA silencing of specific ECM genes in fibroblasts[518], during CDM generation.

Under specific culture conditions, human airway basal cells can be rapidly expanded and differentiated to repopulate tracheal scaffolds with the potential to be used in airway reconstruction[519]. The use of similar cell-based strategies in the dramatically remodelled fibrotic lung is likely to be more challenging. The observation that $KRT5^{+}$ cells accumulate in various configurations in areas where the alveoli are severely destroyed, suggests an alternative epithelial repair pathway is activated[104]. There are likely to be numerous cues to this process, including signals derived from the ECM. Rather than forming a specialised gas exchange surface, basal cells form cystic structures with a mucociliary, airway- like epithelium[36]. It appears therefore, that the presence of basal cells in the fibrotic lung are the sequelae of a regenerative response gone awry[210]. Clinically, a basal cell signature in BAL infers a poorer prognosis[212].

7.4 **Strengths of the research**

This study provides novel data on two distinct microenvironments in pulmonary fibrosis. It provides the first single-cell atlas of the proximal airway niche in newly diagnosed cases of IPF. This captured cellular changes occurring in a different region of the lung and at an earlier stage of disease than has been previously investigated by IPF scRNA-seq studies. The individual subject contribution to this dataset was rigorously evaluated to ensure the trends observed were genuinely a disease-phenomenon rather than donor-effect. Secondly, the study demonstrated for the first time using human cells and cell-derived matrix, that the migratory dynamics of basal cells are influenced by the properties of the fibrotic matrix. I used advanced microscopy techniques to demonstrate the relevance of basal cell-matrix interactions in the context of the pathological remodelling that occurs in IPF lung tissue specimens. Data on how the lung extracellular matrix influences cell migration is likely to also be of interest to those studying the dissemination of tumour cells leading to metastases in lung cancer.

7.5 **Limitations of the research**

Due to the inherent variability associated with human samples, there was a relatively broad spread of data in our *in vitro* findings. Increasing the sample size (based on a power calculation guided by this data) will permit validation of these results in the future.

Inclusion of a control group with shared risk factors for example, chronic obstructive pulmonary disease (COPD) patients, or with other forms of pulmonary fibrosis such as Rheumatoid arthritis- associated ILD, may have provided further useful mechanistic insight.

The single-cell transcriptomics data provided high-resolution insight into the heterogeneity in epithelial cell populations present from the human airways. This technique was not used to evaluate the cultured epithelial cell populations in our ALI model, therefore the extent to which this heterogeneity is preserved *in vitro* is unknown. Moreover, the discrete cell clusters identified by scRNA-seq are yet to be

validated by immunofluorescence imaging. The direct functional relevance of multiple clusters of the same cell-type for example ciliated cells, remains unknown and warrants investigation. Finally, computational tools were used for trajectory inference and prediction of cell-cell interactions using scRNA-seq data, and these have recognised limitations. For example, 'CellPhoneDB' does not have a complete list of all ligand-receptor pairs nor consider the spatial aspect of interacting cells.

7.6 **Future work**

Validation of the work described in this thesis in a separate cohort of IPF patients and healthy controls to add statistical power to our findings, would be beneficial given the variability of some of the data.

In vitro functional work to ascertain the nature of epithelial – macrophage interactions in the airway would be a logical next experiment. This could involve use of a co-culture model and a commercially available inhibitor to block one of the well-characterised ligand-receptor pairs outlined in the 'CellPhoneDB' results section (Chapter 5).

For the first time, the work in this thesis has provided clues on how KRT5+ basal cells may traverse fibrotic matrix in the distal lung. The next step will be understanding why these cells move and what signals they are responding to. Integration of techniques that permit the spatial interrogation of lung tissue specimens to ascertain potential signals in each microenvironment will be pursued next.

7.7 **Concluding remarks**

It is evident from the body of work described in this thesis that epithelial interactions with their microenvironment in the proximal airway and distal lung are integral to the pathogenesis of IPF. These interactions take place in newly diagnosed cases of IPF, and those with advanced fibrosis and tissue remodelling. In the proximal airway, there was perturbation of immune homeostasis and increased crosstalk between secretory epithelial cells and closely associated airway macrophage populations. There was expansion of two airway macrophage subsets in IPF and activation of pathways

relevant to fibrosis. In the IPF distal lung, changes in the matrix microenvironment influenced basal cell migratory capacity and organisation. It is possible that basal cell-matrix interactions drive distinct remodelling programmes in an attempt to regenerate the fibrotic lung.

On the basis of the work presented in this thesis, the overall hypothesis was accepted. Studying airway epithelial cells in IPF has provided valuable insight with direct relevance to disease pathology. The data generated in this project provides a rich resource with which to examine the molecular and cellular changes that occur in the airway epithelial niche in IPF. It is hoped that this will accelerate research to examine whether modulating the microenvironment may restore normal epithelial cell phenotype and function, and ultimately change the course of this devastating disease.

References

1 Raghu, G. *et al.* Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *American journal of respiratory and critical care medicine* **198**, e44-e68, doi:10.1164/rccm.201807-1255ST (2018).

2 Lederer, D. J. & Martinez, F. J. Idiopathic Pulmonary Fibrosis. *The New England journal of medicine* **378**, 1811-1823, doi:10.1056/NEJMra1705751 (2018).

3 American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. *American journal of respiratory and critical care medicine* **165**, 277-304, doi:10.1164/ajrccm.165.2.ats01 (2002).

4 Travis, W. D. *et al.* An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *American journal of respiratory and critical care medicine* **188**, 733-748, doi:10.1164/rccm.201308-1483ST (2013).

5 Wolters, P. J. *et al.* Time for a change: is idiopathic pulmonary fibrosis still idiopathic and only fibrotic? *The Lancet. Respiratory medicine* **6**, 154-160, doi:10.1016/s2213-2600(18)30007-9 (2018).

6 Martinez, F. J. *et al.* Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* **3**, 17074, doi:10.1038/nrdp.2017.74 (2017).

7 Olson, A. L., Gifford, A. H., Inase, N., Fernández Pérez, E. R. & Suda, T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. *Eur Respir Rev* **27**, doi:10.1183/16000617.0077-2018 (2018).

8 Ley, B. *et al.* Code-based Diagnostic Algorithms for Idiopathic Pulmonary Fibrosis. Case Validation and Improvement. *Ann Am Thorac Soc* **14**, 880-887, doi:10.1513/AnnalsATS.201610-764OC (2017).

9 Strongman, H., Kausar, I. & Maher, T. M. Incidence, Prevalence, and Survival of Patients with Idiopathic Pulmonary Fibrosis in the UK. *Adv Ther* **35**, 724-736, doi:10.1007/s12325-018-0693-1 (2018).

10 Navaratnam, V. *et al.* The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax* **66**, 462-467, doi:10.1136/thx.2010.148031 (2011).

11 Hutchinson, J., Fogarty, A., Hubbard, R. & McKeever, T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *The European respiratory journal* **46**, 795-806, doi:10.1183/09031936.00185114 (2015).

12 Snell, N. *et al.* P272 Epidemiology of idiopathic pulmonary fibrosis in the uk: findings from the british lung foundation's 'respiratory health of the nation' project. *Thorax* **71**, A236-A236, doi:10.1136/thoraxjnl-2016-209333.415 (2016).

13 Navaratnam, V. & Hubbard, R. B. The Mortality Burden of Idiopathic Pulmonary Fibrosis in the United Kingdom. *American journal of respiratory and critical care medicine* **200**, 256-258, doi:10.1164/rccm.201902-0467LE (2019).

14 Behr, J. *et al.* Management of patients with idiopathic pulmonary fibrosis in clinical practice: the INSIGHTS-IPF registry. *The European respiratory journal* **46**, 186-196, doi:10.1183/09031936.00217614 (2015).

15 Raghu, G., Weycker, D., Edelsberg, J., Bradford, W. Z. & Oster, G. Incidence and prevalence of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **174**, 810-816, doi:10.1164/rccm.200602-163OC (2006).

16 Raghu, G. *et al.* Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11. *The Lancet. Respiratory medicine* **2**, 566-572, doi:10.1016/s2213-2600(14)70101-8 (2014).

17 Ryerson, C. J. *et al.* Cough predicts prognosis in idiopathic pulmonary fibrosis. *Respirology* **16**, 969-975, doi:10.1111/j.1440-1843.2011.01996.x (2011).

18 Raghu, G. Idiopathic pulmonary fibrosis: lessons from clinical trials over the past 25 years. *The European respiratory journal* **50**, doi:10.1183/13993003.01209-2017 (2017).

19 Wells, A. U. *et al.* Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *American journal of respiratory and critical care medicine* **167**, 962-969, doi:10.1164/rccm.2111053 (2003).

20 Lama, V. N. *et al.* Prognostic value of desaturation during a 6-minute walk test in idiopathic interstitial pneumonia. *American journal of respiratory and critical care medicine* **168**, 1084-1090, doi:10.1164/rccm.200302-219OC (2003).

21 Ley, B., Collard, H. R. & King, T. E., Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **183**, 431-440, doi:10.1164/rccm.201006-0894CI (2011).

22 Vancheri, C., Failla, M., Crimi, N. & Raghu, G. Idiopathic pulmonary fibrosis: a disease with similarities and links to cancer biology. *The European respiratory journal* **35**, 496-504, doi:10.1183/09031936.00077309 (2010).

23 Olson, A. L. *et al.* Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. *American journal of respiratory and critical care medicine* **176**, 277-284, doi:10.1164/rccm.200701-044OC (2007).

24 Ley, B. *et al.* A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* **156**, 684-691, doi:10.7326/0003-4819-156-10-201205150-00004 (2012).

25 Russell, A. M., Ripamonti, E. & Vancheri, C. Qualitative European survey of patients with idiopathic pulmonary fibrosis: patients' perspectives of the disease and treatment. *BMC Pulm Med* **16**, 10, doi:10.1186/s12890-016-0171-y (2016).

26 Morell, F. *et al.* Chronic hypersensitivity pneumonitis in patients diagnosed with idiopathic pulmonary fibrosis: a prospective case-cohort study. *The Lancet. Respiratory medicine* **1**, 685-694, doi:10.1016/s2213-2600(13)70191-7 (2013).

27 Hunninghake, G. W. *et al.* Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **164**, 193-196, doi:10.1164/ajrccm.164.2.2101090 (2001).

28 Meyer, K. C. *et al.* An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *American journal of respiratory and critical care medicine* **185**, 1004-1014, doi:10.1164/rccm.201202-0320ST (2012).

29 Raghu, G. *et al.* Diagnosis of Hypersensitivity Pneumonitis in Adults. An Official ATS/JRS/ALAT Clinical Practice Guideline. *American journal of respiratory and critical care medicine* **202**, e36-e69, doi:10.1164/rccm.202005-2032ST (2020).

30 Allden, S. J. *et al.* The Transferrin Receptor CD71 Delineates Functionally Distinct Airway Macrophage Subsets during Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **200**, 209-219, doi:10.1164/rccm.201809-1775OC (2019).
31 Molyneaux, P. L. *et al.* The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **190**, 906-913, doi:10.1164/rccm.201403-0541OC (2014).
32 Leslie, K. O. My approach to interstitial lung disease using clinical, radiological and histopathological patterns. *J Clin Pathol* **62**, 387-401, doi:10.1136/jcp.2008.059782 (2009).
33 Jones, M. G. *et al.* Three-dimensional characterization of fibroblast foci in idiopathic pulmonary fibrosis. *JCI insight* **1**, doi:10.1172/jci.insight.86375 (2016).
34 Herrera, J. *et al.* Registration of the extracellular matrix components constituting the fibroblastic focus in idiopathic pulmonary fibrosis. *JCI insight* **4**, doi:10.1172/jci.insight.125185 (2019).
35 Selman, M. & Pardo, A. Role of epithelial cells in idiopathic pulmonary fibrosis: from innocent targets to serial killers. *Proc Am Thorac Soc* **3**, 364-372, doi:10.1513/pats.200601-003TK (2006).
36 Seibold, M. A. *et al.* The idiopathic pulmonary fibrosis honeycomb cyst contains a mucocilary pseudostratified epithelium. *PloS one* **8**, e58658, doi:10.1371/journal.pone.0058658 (2013).
37 Flaherty, K. R. *et al.* Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *American journal of respiratory and critical care medicine* **170**, 904-910, doi:10.1164/rccm.200402-147OC (2004).
38 Walsh, S. L. F. *et al.* Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *The Lancet. Respiratory medicine* **4**, 557-565, doi:10.1016/s2213-2600(16)30033-9 (2016).
39 Raghu, G., Anstrom, K. J., King, T. E., Jr., Lasky, J. A. & Martinez, F. J. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *The New England journal of medicine* **366**, 1968-1977, doi:10.1056/NEJMoa1113354 (2012).
40 Raghu, G. *et al.* An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *American journal of respiratory and critical care medicine* **192**, e3-19, doi:10.1164/rccm.201506-1063ST (2015).
41 Selman, M., King, T. E. & Pardo, A. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Ann Intern Med* **134**, 136-151, doi:10.7326/0003-4819-134-2-200101160-00015 (2001).
42 Schaefer, C. J., Ruhrmund, D. W., Pan, L., Seiwert, S. D. & Kossen, K. Antifibrotic activities of pirfenidone in animal models. *Eur Respir Rev* **20**, 85-97, doi:10.1183/09059180.00001111 (2011).
43 Conte, E. *et al.* Effect of pirfenidone on proliferation, TGF-β-induced myofibroblast differentiation and fibrogenic activity of primary human lung fibroblasts. *Eur J Pharm Sci* **58**, 13-19, doi:10.1016/j.ejps.2014.02.014 (2014).

44 Noble, P. W. *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet (London, England)* **377**, 1760-1769, doi:10.1016/s0140-6736(11)60405-4 (2011).
45 King, T. E., Jr. *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *The New England journal of medicine* **370**, 2083-2092, doi:10.1056/NEJMoa1402582 (2014).
46 Wollin, L. *et al.* Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *The European respiratory journal* **45**, 1434-1445, doi:10.1183/09031936.00174914 (2015).
47 Richeldi, L. *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *The New England journal of medicine* **370**, 2071-2082, doi:10.1056/NEJMoa1402584 (2014).
48 Galli, J. A. *et al.* Pirfenidone and nintedanib for pulmonary fibrosis in clinical practice: Tolerability and adverse drug reactions. *Respirology* **22**, 1171-1178, doi:10.1111/resp.13024 (2017).
49 Hewitt, R. J. & Maher, T. M. Idiopathic Pulmonary Fibrosis: New and Emerging Treatment Options. *Drugs & aging* **36**, 485-492, doi:10.1007/s40266-019-00647-y (2019).
50 George, P. M., Patterson, C. M., Reed, A. K. & Thillai, M. Lung transplantation for idiopathic pulmonary fibrosis. *The Lancet. Respiratory medicine* **7**, 271-282, doi:10.1016/s2213-2600(18)30502-2 (2019).
51 Kistler, K. D., Nalysnyk, L., Rotella, P. & Esser, D. Lung transplantation in idiopathic pulmonary fibrosis: a systematic review of the literature. *BMC Pulm Med* **14**, 139, doi:10.1186/1471-2466-14-139 (2014).
52 Hewitt, R. J. & Lloyd, C. M. Regulation of immune responses by the airway epithelial cell landscape. *Nat Rev Immunol*, 1-16, doi:10.1038/s41577-020-00477-9 (2021).
53 Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983, doi:10.1126/sciadv.aba1983 (2020).
54 Pardo, A. & Selman, M. The Interplay of the Genetic Architecture, Aging, and Environmental Factors in the Pathogenesis of Idiopathic Pulmonary Fibrosis. *Am J Respir Cell Mol Biol* **64**, 163-172, doi:10.1165/rcmb.2020-0373PS (2021).
55 Sack, C. & Raghu, G. Idiopathic pulmonary fibrosis: unmasking cryptogenic environmental factors. *The European respiratory journal* **53**, doi:10.1183/13993003.01699-2018 (2019).
56 Taskar, V. S. & Coultas, D. B. Is idiopathic pulmonary fibrosis an environmental disease? *Proc Am Thorac Soc* **3**, 293-298, doi:10.1513/pats.200512-131TK (2006).
57 Steele, M. P. *et al.* Clinical and pathologic features of familial interstitial pneumonia. *American journal of respiratory and critical care medicine* **172**, 1146-1152, doi:10.1164/rccm.200408-1104OC (2005).
58 Park, Y., Ahn, C. & Kim, T. H. Occupational and environmental risk factors of idiopathic pulmonary fibrosis: a systematic review and meta-analyses. *Sci Rep* **11**, 4318, doi:10.1038/s41598-021-81591-z (2021).
59 Baumgartner, K. B. *et al.* Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am J Epidemiol* **152**, 307-315, doi:10.1093/aje/152.4.307 (2000).

60 Johannson, K. A. *et al.* Acute exacerbation of idiopathic pulmonary fibrosis associated with air pollution exposure. *The European respiratory journal* **43**, 1124-1131, doi:10.1183/09031936.00122213 (2014).
61 Sesé, L. *et al.* Role of atmospheric pollution on the natural history of idiopathic pulmonary fibrosis. *Thorax* **73**, 145-150, doi:10.1136/thoraxjnl-2017-209967 (2018).
62 Raghu, G. *et al.* High prevalence of abnormal acid gastro-oesophageal reflux in idiopathic pulmonary fibrosis. *The European respiratory journal* **27**, 136-142, doi:10.1183/09031936.06.00037005 (2006).
63 Johannson, K. A. *et al.* Antacid therapy in idiopathic pulmonary fibrosis: more questions than answers? *The Lancet. Respiratory medicine* **5**, 591-598, doi:10.1016/s2213-2600(17)30219-9 (2017).
64 Lee, J. S. *et al.* Anti-acid treatment and disease progression in idiopathic pulmonary fibrosis: an analysis of data from three randomised controlled trials. *The Lancet. Respiratory medicine* **1**, 369-376, doi:10.1016/s2213-2600(13)70105-x (2013).
65 Kreuter, M. *et al.* Antacid therapy and disease outcomes in idiopathic pulmonary fibrosis: a pooled analysis. *The Lancet. Respiratory medicine* **4**, 381-389, doi:10.1016/s2213-2600(16)00067-9 (2016).
66 Raghu, G. *et al.* Laparoscopic anti-reflux surgery for the treatment of idiopathic pulmonary fibrosis (WRAP-IPF): a multicentre, randomised, controlled phase 2 trial. *The Lancet. Respiratory medicine* **6**, 707-714, doi:10.1016/s2213-2600(18)30301-1 (2018).
67 Mackintosh, J. A. *et al.* In patients with idiopathic pulmonary fibrosis the presence of hiatus hernia is associated with disease progression and mortality. *The European respiratory journal* **53**, doi:10.1183/13993003.02412-2018 (2019).
68 Molyneaux, P. L. & Maher, T. M. The role of infection in the pathogenesis of idiopathic pulmonary fibrosis. *Eur Respir Rev* **22**, 376-381, doi:10.1183/09059180.00000713 (2013).
69 Sheng, G. *et al.* Viral Infection Increases the Risk of Idiopathic Pulmonary Fibrosis: A Meta-Analysis. *Chest* **157**, 1175-1187, doi:10.1016/j.chest.2019.10.032 (2020).
70 Kropski, J. A. *et al.* Extensive phenotyping of individuals at risk for familial interstitial pneumonia reveals clues to the pathogenesis of interstitial lung disease. *American journal of respiratory and critical care medicine* **191**, 417-426, doi:10.1164/rccm.201406-1162OC (2015).
71 Yin, Q. *et al.* Assessment of viral RNA in idiopathic pulmonary fibrosis using RNA-seq. *BMC Pulm Med* **20**, 81, doi:10.1186/s12890-020-1114-1 (2020).
72 Hilty, M. *et al.* Disordered microbial communities in asthmatic airways. *PloS one* **5**, e8578, doi:10.1371/journal.pone.0008578 (2010).
73 Molyneaux, P. L. *et al.* Host-Microbial Interactions in Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **195**, 1640-1650, doi:10.1164/rccm.201607-1408OC (2017).
74 Hewitt, R. J. & Molyneaux, P. L. The respiratory microbiome in idiopathic pulmonary fibrosis. *Ann Transl Med* **5**, 250, doi:10.21037/atm.2017.01.56 (2017).
75 Wilson, A. M. *et al.* Effect of Co-trimoxazole (Trimethoprim-Sulfamethoxazole) vs Placebo on Death, Lung Transplant, or Hospital Admission in Patients With Moderate and Severe Idiopathic Pulmonary Fibrosis: The EME-TIPAC Randomized Clinical Trial. *Jama* **324**, 2282-2291, doi:10.1001/jama.2020.22960 (2020).

76 Armanios, M. Y. *et al.* Telomerase mutations in families with idiopathic pulmonary fibrosis. *The New England journal of medicine* **356**, 1317-1326, doi:10.1056/NEJMoa066157 (2007).
77 Cogan, J. D. *et al.* Rare variants in RTEL1 are associated with familial interstitial pneumonia. *American journal of respiratory and critical care medicine* **191**, 646-655, doi:10.1164/rccm.201408-1510OC (2015).
78 Wang, Y. *et al.* Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am J Hum Genet* **84**, 52-59, doi:10.1016/j.ajhg.2008.11.010 (2009).
79 Nogee, L. M. *et al.* A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *The New England journal of medicine* **344**, 573-579, doi:10.1056/nejm200102223440805 (2001).
80 Stuart, B. D. *et al.* Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. *The Lancet. Respiratory medicine* **2**, 557-565, doi:10.1016/s2213-2600(14)70124-9 (2014).
81 Petrovski, S. *et al.* An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **196**, 82-93, doi:10.1164/rccm.201610-2088OC (2017).
82 Mathai, S. K., Newton, C. A., Schwartz, D. A. & Garcia, C. K. Pulmonary fibrosis in the era of stratified medicine. *Thorax* **71**, 1154-1160, doi:10.1136/thoraxjnl-2016-209172 (2016).
83 Fingerlin, T. E. *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nature genetics* **45**, 613-620, doi:10.1038/ng.2609 (2013).
84 Noth, I. *et al.* Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *The Lancet. Respiratory medicine* **1**, 309-317, doi:10.1016/s2213-2600(13)70045-6 (2013).
85 Allen, R. J. *et al.* Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *The Lancet. Respiratory medicine* **5**, 869-880, doi:10.1016/s2213-2600(17)30387-9 (2017).
86 Allen, R. J. *et al.* Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **201**, 564-574, doi:10.1164/rccm.201905-1017OC (2020).
87 Mathai, S. K., Schwartz, D. A. & Warg, L. A. Genetic susceptibility and pulmonary fibrosis. *Curr Opin Pulm Med* **20**, 429-435, doi:10.1097/mcp.0000000000000074 (2014).
88 Roy, M. G. *et al.* Muc5b is required for airway defence. *Nature* **505**, 412-416, doi:10.1038/nature12807 (2014).
89 Whitsett, J. A. & Alenghat, T. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nature immunology* **16**, 27-35, doi:10.1038/ni.3045 (2015).
90 Seibold, M. A. *et al.* A common MUC5B promoter polymorphism and pulmonary fibrosis. *The New England journal of medicine* **364**, 1503-1512, doi:10.1056/NEJMoa1013660 (2011).
91 Peljto, A. L. *et al.* Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *Jama* **309**, 2232-2239, doi:10.1001/jama.2013.5827 (2013).

92 Nakano, Y. *et al.* MUC5B Promoter Variant rs35705950 Affects MUC5B Expression in the Distal Airways in Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **193**, 464-466, doi:10.1164/rccm.201509-1872LE (2016).
93 Stock, C. J. *et al.* Interaction between the promoter MUC5B polymorphism and mucin expression: is there a difference according to ILD subtype? *Thorax* **75**, 901-903, doi:10.1136/thoraxjnl-2020-214579 (2020).
94 Hancock, L. A. *et al.* Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. *Nature communications* **9**, 5363, doi:10.1038/s41467-018-07768-9 (2018).
95 Hunninghake, G. M. *et al.* MUC5B promoter polymorphism and interstitial lung abnormalities. *The New England journal of medicine* **368**, 2192-2200, doi:10.1056/NEJMoa1216076 (2013).
96 Putman, R. K. *et al.* Association Between Interstitial Lung Abnormalities and All-Cause Mortality. *Jama* **315**, 672-681, doi:10.1001/jama.2016.0518 (2016).
97 Evans, C. M. *et al.* Idiopathic Pulmonary Fibrosis: A Genetic Disease That Involves Mucociliary Dysfunction of the Peripheral Airways. *Physiol Rev* **96**, 1567-1591, doi:10.1152/physrev.00004.2016 (2016).
98 Moore, C. *et al.* Resequencing Study Confirms That Host Defense and Cell Senescence Gene Variants Contribute to the Risk of Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **200**, 199-208, doi:10.1164/rccm.201810-1891OC (2019).
99 Basil, M. C. *et al.* The Cellular and Physiological Basis for Lung Repair and Regeneration: Past, Present, and Future. *Cell stem cell* **26**, 482-502, doi:10.1016/j.stem.2020.03.009 (2020).
100 Zepp, J. A. & Morrisey, E. E. Cellular crosstalk in the development and regeneration of the respiratory system. *Nat Rev Mol Cell Biol* **20**, 551-566, doi:10.1038/s41580-019-0141-3 (2019).
101 Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J Clin Invest* **123**, 3025-3036, doi:10.1172/jci68782 (2013).
102 Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251-255, doi:10.1038/nature25786 (2018).
103 Kim, C. F. *et al.* Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121**, 823-835, doi:10.1016/j.cell.2005.03.032 (2005).
104 Vaughan, A. E. *et al.* Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* **517**, 621-625, doi:10.1038/nature14112 (2015).
105 Zuo, W. *et al.* p63(+)Krt5(+) distal airway stem cells are essential for lung regeneration. *Nature* **517**, 616-620, doi:10.1038/nature13903 (2015).
106 Zepp, J. A. *et al.* Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. *Cell* **170**, 1134-1148.e1110, doi:10.1016/j.cell.2017.07.034 (2017).
107 Katzenstein, A. L. Pathogenesis of "fibrosis" in interstitial pneumonia: an electron microscopic study. *Human pathology* **16**, 1015-1024 (1985).
108 Uhal, B. D. *et al.* Alveolar epithelial cell death adjacent to underlying myofibroblasts in advanced fibrotic human lung. *Am J Physiol* **275**, L1192-1199, doi:10.1152/ajplung.1998.275.6.L1192 (1998).

109 Qunn, L. *et al.* Hyperplastic epithelial foci in honeycomb lesions in idiopathic pulmonary fibrosis. *Virchows Arch* **441**, 271-278, doi:10.1007/s00428-002-0618-9 (2002).
110 Korfei, M. *et al.* Epithelial endoplasmic reticulum stress and apoptosis in sporadic idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **178**, 838-846, doi:10.1164/rccm.200802-313OC (2008).
111 Sisson, T. H. *et al.* Targeted injury of type II alveolar epithelial cells induces pulmonary fibrosis. *American journal of respiratory and critical care medicine* **181**, 254-263, doi:10.1164/rccm.200810-1615OC (2010).
112 Tsujino, K. *et al.* Tetraspanin CD151 protects against pulmonary fibrosis by maintaining epithelial integrity. *American journal of respiratory and critical care medicine* **186**, 170-180, doi:10.1164/rccm.201201-0117OC (2012).
113 Mora, A. L., Rojas, M., Pardo, A. & Selman, M. Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. *Nature reviews. Drug discovery* **16**, 755-772, doi:10.1038/nrd.2017.170 (2017).
114 van Deursen, J. M. The role of senescent cells in ageing. *Nature* **509**, 439-446, doi:10.1038/nature13193 (2014).
115 Yao, C. *et al.* Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **203**, 707-717, doi:10.1164/rccm.202004-1274OC (2021).
116 Disayabutr, S. *et al.* miR-34 miRNAs Regulate Cellular Senescence in Type II Alveolar Epithelial Cells of Patients with Idiopathic Pulmonary Fibrosis. *PloS one* **11**, e0158367, doi:10.1371/journal.pone.0158367 (2016).
117 Lee, J. S. *et al.* Molecular Markers of Telomere Dysfunction and Senescence are Common Findings in the Usual Interstitial Pneumonia Pattern of Lung Fibrosis. *Histopathology*, doi:10.1111/his.14334 (2021).
118 Alder, J. K. *et al.* Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci U S A* **112**, 5099-5104, doi:10.1073/pnas.1504780112 (2015).
119 Naikawadi, R. P. *et al.* Telomere dysfunction in alveolar epithelial cells causes lung remodeling and fibrosis. *JCI insight* **1**, e86704, doi:10.1172/jci.insight.86704 (2016).
120 Bueno, M., Calyeca, J., Rojas, M. & Mora, A. L. Mitochondria dysfunction and metabolic reprogramming as drivers of idiopathic pulmonary fibrosis. *Redox Biol* **33**, 101509, doi:10.1016/j.redox.2020.101509 (2020).
121 Bueno, M. *et al.* PINK1 deficiency impairs mitochondrial homeostasis and promotes lung fibrosis. *J Clin Invest* **125**, 521-538, doi:10.1172/jci74942 (2015).
122 Liang, J. *et al.* Hyaluronan and TLR4 promote surfactant-protein-C-positive alveolar progenitor cell renewal and prevent severe pulmonary fibrosis in mice. *Nat Med* **22**, 1285-1293, doi:10.1038/nm.4192 (2016).
123 Wu, H. *et al.* Progressive Pulmonary Fibrosis Is Caused by Elevated Mechanical Tension on Alveolar Stem Cells. *Cell* **180**, 107-121.e117, doi:10.1016/j.cell.2019.11.027 (2020).
124 Selman, M. & Pardo, A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *American journal of respiratory and critical care medicine* **189**, 1161-1172, doi:10.1164/rccm.201312-2221PP (2014).
125 King, T. E., Jr., Pardo, A. & Selman, M. Idiopathic pulmonary fibrosis. *Lancet (London, England)* **378**, 1949-1961, doi:10.1016/s0140-6736(11)60052-4 (2011).

126 Zuo, F. *et al.* Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A* **99**, 6292-6297, doi:10.1073/pnas.092134099 (2002).
127 Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15**, 786-801, doi:10.1038/nrm3904 (2014).
128 Khalil, N. *et al.* Increased production and immunohistochemical localization of transforming growth factor-beta in idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* **5**, 155-162, doi:10.1165/ajrcmb/5.2.155 (1991).
129 Eickelberg, O. *et al.* Extracellular matrix deposition by primary human lung fibroblasts in response to TGF-beta1 and TGF-beta3. *Am J Physiol* **276**, L814-824, doi:10.1152/ajplung.1999.276.5.L814 (1999).
130 Xu, Y. D. *et al.* Release of biologically active TGF-beta1 by alveolar epithelial cells results in pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* **285**, L527-539, doi:10.1152/ajplung.00298.2002 (2003).
131 Horan, G. S. *et al.* Partial inhibition of integrin alpha(v)beta6 prevents pulmonary fibrosis without exacerbating inflammation. *American journal of respiratory and critical care medicine* **177**, 56-65, doi:10.1164/rccm.200706-805OC (2008).
132 Willis, B. C. *et al.* Induction of epithelial-mesenchymal transition in alveolar epithelial cells by transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. *The American journal of pathology* **166**, 1321-1332, doi:10.1016/s0002-9440(10)62351-6 (2005).
133 Königshoff, M. *et al.* Functional Wnt signaling is increased in idiopathic pulmonary fibrosis. *PloS one* **3**, e2142, doi:10.1371/journal.pone.0002142 (2008).
134 Kneidinger, N. *et al.* Activation of the WNT/β-catenin pathway attenuates experimental emphysema. *American journal of respiratory and critical care medicine* **183**, 723-733, doi:10.1164/rccm.200910-1560OC (2011).
135 Baarsma, H. A. & Königshoff, M. 'WNT-er is coming': WNT signalling in chronic lung diseases. *Thorax* **72**, 746-759, doi:10.1136/thoraxjnl-2016-209753 (2017).
136 Bolaños, A. L. *et al.* Role of Sonic Hedgehog in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* **303**, L978-990, doi:10.1152/ajplung.00184.2012 (2012).
137 Katzen, J. & Beers, M. F. Contributions of alveolar epithelial cell quality control to pulmonary fibrosis. *J Clin Invest* **130**, 5088-5099, doi:10.1172/jci139519 (2020).
138 Tomasek, J. J., Gabbiani, G., Hinz, B., Chaponnier, C. & Brown, R. A. Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nat Rev Mol Cell Biol* **3**, 349-363, doi:10.1038/nrm809 (2002).
139 Hinz, B. *et al.* The myofibroblast: one function, multiple origins. *The American journal of pathology* **170**, 1807-1816, doi:10.2353/ajpath.2007.070112 (2007).
140 Rock, J. R. *et al.* Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci U S A* **108**, E1475-1483, doi:10.1073/pnas.1117988108 (2011).
141 Yang, J. *et al.* Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **21**, 341-352, doi:10.1038/s41580-020-0237-9 (2020).
142 Yao, L. *et al.* Paracrine signalling during ZEB1-mediated epithelial-mesenchymal transition augments local myofibroblast differentiation in lung fibrosis. *Cell Death Differ* **26**, 943-957, doi:10.1038/s41418-018-0175-7 (2019).

143 Heukels, P. *et al.* Fibrocytes are increased in lung and peripheral blood of patients with idiopathic pulmonary fibrosis. *Respiratory research* **19**, 90, doi:10.1186/s12931-018-0798-8 (2018).
144 Croft, A. P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**, 246-251, doi:10.1038/s41586-019-1263-7 (2019).
145 Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* **6**, eaba1972, doi:10.1126/sciadv.aba1972 (2020).
146 Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nature communications* **11**, 1920, doi:10.1038/s41467-020-15647-5 (2020).
147 Desmoulière, A., Redard, M., Darby, I. & Gabbiani, G. Apoptosis mediates the decrease in cellularity during the transition between granulation tissue and scar. *The American journal of pathology* **146**, 56-66 (1995).
148 Maher, T. M. *et al.* Diminished prostaglandin E2 contributes to the apoptosis paradox in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **182**, 73-82, doi:10.1164/rccm.200905-0674OC (2010).
149 Borok, Z. *et al.* Augmentation of functional prostaglandin E levels on the respiratory epithelial surface by aerosol administration of prostaglandin E. *Am Rev Respir Dis* **144**, 1080-1084, doi:10.1164/ajrccm/144.5.1080 (1991).
150 Kulasekaran, P. *et al.* Endothelin-1 and transforming growth factor-beta1 independently induce fibroblast resistance to apoptosis via AKT activation. *Am J Respir Cell Mol Biol* **41**, 484-493, doi:10.1165/rcmb.2008-0447OC (2009).
151 Xia, H. *et al.* Pathological integrin signaling enhances proliferation of primary lung fibroblasts from patients with idiopathic pulmonary fibrosis. *J Exp Med* **205**, 1659-1672, doi:10.1084/jem.20080001 (2008).
152 Nho, R. S. & Hergert, P. IPF fibroblasts are desensitized to type I collagen matrix-induced cell death by suppressing low autophagy via aberrant Akt/mTOR kinases. *PloS one* **9**, e94616, doi:10.1371/journal.pone.0094616 (2014).
153 Hynes, R. O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol* **4**, a004903, doi:10.1101/cshperspect.a004903 (2012).
154 Naba, A. *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* **11**, M111.014647, doi:10.1074/mcp.M111.014647 (2012).
155 Haak, A. J., Tan, Q. & Tschumperlin, D. J. Matrix biomechanics and dynamics in pulmonary fibrosis. *Matrix Biol* **73**, 64-76, doi:10.1016/j.matbio.2017.12.004 (2018).
156 Philp, C. J. *et al.* Extracellular Matrix Cross-Linking Enhances Fibroblast Growth and Protects against Matrix Proteolysis in Lung Fibrosis. *Am J Respir Cell Mol Biol* **58**, 594-603, doi:10.1165/rcmb.2016-0379OC (2018).
157 Booth, A. J. *et al.* Acellular normal and fibrotic human lung matrices as a culture system for in vitro investigation. *American journal of respiratory and critical care medicine* **186**, 866-876, doi:10.1164/rccm.201204-0754OC (2012).
158 Huang, X. *et al.* Matrix stiffness-induced myofibroblast differentiation is mediated by intrinsic mechanotransduction. *Am J Respir Cell Mol Biol* **47**, 340-348, doi:10.1165/rcmb.2012-0050OC (2012).

159 Parker, M. W. *et al.* Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J Clin Invest* **124**, 1622-1635, doi:10.1172/JCI71386 (2014).
160 Byrne, A. J., Maher, T. M. & Lloyd, C. M. Pulmonary Macrophages: A New Therapeutic Pathway in Fibrosing Lung Disease? *Trends Mol Med* **22**, 303-316, doi:10.1016/j.molmed.2016.02.004 (2016).
161 Schyns, J. *et al.* Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. *Nature communications* **10**, 3964, doi:10.1038/s41467-019-11843-0 (2019).
162 Misharin, A. V. *et al.* Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med* **214**, 2387-2404, doi:10.1084/jem.20162152 (2017).
163 Reyfman, P. A. *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *American journal of respiratory and critical care medicine*, doi:10.1164/rccm.201712-2410OC (2018).
164 Morse, C. *et al.* Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *The European respiratory journal* **54**, doi:10.1183/13993003.02441-2018 (2019).
165 Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312-1326, doi:10.1016/j.cell.2014.11.018 (2014).
166 Yu, Y. R. *et al.* Flow Cytometric Analysis of Myeloid Cells in Human Blood, Bronchoalveolar Lavage, and Lung Tissues. *Am J Respir Cell Mol Biol* **54**, 13-24, doi:10.1165/rcmb.2015-0146OC (2016).
167 Schupp, J. C. *et al.* Single-Cell Transcriptional Archetypes of Airway Inflammation in Cystic Fibrosis. *American journal of respiratory and critical care medicine* **202**, 1419-1429, doi:10.1164/rccm.202004-0991OC (2020).
168 Vieira Braga, F. A. *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* **25**, 1153-1163, doi:10.1038/s41591-019-0468-5 (2019).
169 Puttur, F., Gregory, L. G. & Lloyd, C. M. Airway macrophages as the guardians of tissue repair in the lung. *Immunology and cell biology* **97**, 246-257, doi:10.1111/imcb.12235 (2019).
170 Hussell, T. & Bell, T. J. Alveolar macrophages: plasticity in a tissue-specific context. *Nat Rev Immunol* **14**, 81-93, doi:10.1038/nri3600 (2014).
171 Snelgrove, R. J. *et al.* A critical function for CD200 in lung immune homeostasis and the severity of influenza infection. *Nature immunology* **9**, 1074-1083, doi:10.1038/ni.1637 (2008).
172 Guilliams, M. *et al.* Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *J Exp Med* **210**, 1977-1992, doi:10.1084/jem.20131199 (2013).
173 Hashimoto, D. *et al.* Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* **38**, 792-804, doi:10.1016/j.immuni.2013.04.004 (2013).
174 Byrne, A. J. *et al.* Dynamics of human monocytes and airway macrophages during healthy aging and after transplant. *J Exp Med* **217**, doi:10.1084/jem.20191236 (2020).

175 Osterholzer, J. J. *et al.* Implicating exudate macrophages and Ly-6C(high) monocytes in CCR2-dependent lung fibrosis following gene-targeted alveolar injury. *J Immunol* **190**, 3447-3457, doi:10.4049/jimmunol.1200604 (2013).
176 Gibbons, M. A. *et al.* Ly6Chi monocytes direct alternatively activated profibrotic macrophage regulation of lung fibrosis. *American journal of respiratory and critical care medicine* **184**, 569-581, doi:10.1164/rccm.201010-1719OC (2011).
177 Mills, C. D., Kincaid, K., Alt, J. M., Heilman, M. J. & Hill, A. M. M-1/M-2 macrophages and the Th1/Th2 paradigm. *J Immunol* **164**, 6166-6173, doi:10.4049/jimmunol.164.12.6166 (2000).
178 Tardelli, M. *et al.* Osteopontin is a key player for local adipose tissue macrophage proliferation in obesity. *Mol Metab* **5**, 1131-1137, doi:10.1016/j.molmet.2016.09.003 (2016).
179 Deprez, M. *et al.* A Single-cell Atlas of the Human Healthy Airways. *American journal of respiratory and critical care medicine*, doi:10.1164/rccm.201911-2199OC (2020).
180 Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol* **38**, 970-979, doi:10.1038/s41587-020-0602-4 (2020).
181 Westphalen, K. *et al.* Sessile alveolar macrophages communicate with alveolar epithelium to modulate immunity. *Nature* **506**, 503-506, doi:10.1038/nature12902 (2014).
182 Raredon, M. S. B. *et al.* Single-cell connectomic analysis of adult mammalian lungs. *Science Advances* **5**, eaaw3851, doi:10.1126/sciadv.aaw3851 (2019).
183 Wynn, T. A. & Vannella, K. M. Macrophages in Tissue Repair, Regeneration, and Fibrosis. *Immunity* **44**, 450-462, doi:10.1016/j.immuni.2016.02.015 (2016).
184 Shook, B. A. *et al.* Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair. *Science* **362**, doi:10.1126/science.aar2971 (2018).
185 Pakshir, P. *et al.* Dynamic fibroblast contractions attract remote macrophages in fibrillar collagen matrix. *Nature communications* **10**, 1850, doi:10.1038/s41467-019-09709-6 (2019).
186 Lodyga, M. *et al.* Cadherin-11-mediated adhesion of macrophages to myofibroblasts establishes a profibrotic niche of active TGF-β. *Sci Signal* **12**, doi:10.1126/scisignal.aao3469 (2019).
187 Plantier, L. *et al.* Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. *Thorax* **66**, 651-657, doi:10.1136/thx.2010.151555 (2011).
188 Chilosi, M. *et al.* Abnormal re-epithelialization and lung remodeling in idiopathic pulmonary fibrosis: the role of deltaN-p63. *Lab Invest* **82**, 1335-1345, doi:10.1097/01.lab.0000032380.82232.67 (2002).
189 Xu, Y. *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI insight* **1**, e90558, doi:10.1172/jci.insight.90558 (2016).
190 Rhodin, J. A. The ciliated cell. Ultrastructure and function of the human tracheal mucosa. *Am Rev Respir Dis* **93**, Suppl:1-15, doi:10.1164/arrd.1966.93.3P2.1 (1966).
191 Breeze, R. G. & Wheeldon, E. B. The cells of the pulmonary airways. *Am Rev Respir Dis* **116**, 705-777, doi:10.1164/arrd.1977.116.4.705 (1977).

192 Widdicombe, J. H. Early studies on the surface epithelium of mammalian airways. *Am J Physiol Lung Cell Mol Physiol* **317**, L486-L495, doi:10.1152/ajplung.00240.2019 (2019).
193 Fahy, J. V. & Dickey, B. F. Airway mucus function and dysfunction. *The New England journal of medicine* **363**, 2233-2247, doi:10.1056/NEJMra0910061 (2010).
194 Rock, J. R., Randell, S. H. & Hogan, B. L. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis Model Mech* **3**, 545-556, doi:10.1242/dmm.006031 (2010).
195 Rock, J. R. *et al.* Notch-dependent differentiation of adult airway basal stem cells. *Cell stem cell* **8**, 639-648, doi:10.1016/j.stem.2011.04.003 (2011).
196 Nakajima, M. *et al.* Immunohistochemical and ultrastructural studies of basal cells, Clara cells and bronchiolar cuboidal cells in normal human airways. *Pathol Int* **48**, 944-953, doi:10.1111/j.1440-1827.1998.tb03865.x (1998).
197 Watson, J. K. *et al.* Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium. *Cell Rep* **12**, 90-101, doi:10.1016/j.celrep.2015.06.011 (2015).
198 Goldfarbmuren, K. C. *et al.* Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nature communications* **11**, 2485, doi:10.1038/s41467-020-16239-z (2020).
199 Carraro, G. *et al.* Single Cell Reconstruction of Human Basal Cell Diversity in Normal and IPF Lung. *American journal of respiratory and critical care medicine*, doi:10.1164/rccm.201904-0792OC (2020).
200 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324, doi:10.1038/s41586-018-0393-7 (2018).
201 Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381, doi:10.1038/s41586-018-0394-6 (2018).
202 Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* **15**, e8746, doi:10.15252/msb.20188746 (2019).
203 Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547-554, doi:10.1038/s41587-019-0071-9 (2019).
204 Ruiz García, S. *et al.* Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development* **146**, doi:10.1242/dev.177428 (2019).
205 Smirnova, N. F. *et al.* Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs. *Respiratory research* **17**, 83, doi:10.1186/s12931-016-0404-x (2016).
206 Kumar, P. A. *et al.* Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell* **147**, 525-538, doi:10.1016/j.cell.2011.10.001 (2011).
207 Ray, S. *et al.* Rare SOX2(+) Airway Progenitor Cells Generate KRT5(+) Cells that Repopulate Damaged Alveolar Parenchyma following Influenza Virus Infection. *Stem Cell Reports* **7**, 817-825, doi:10.1016/j.stemcr.2016.09.010 (2016).

208 Fernanda de Mello Costa, M., Weiner, A. I. & Vaughan, A. E. Basal-like Progenitor Cells: A Review of Dysplastic Alveolar Regeneration and Remodeling in Lung Repair. *Stem Cell Reports* **15**, 1015-1025, doi:10.1016/j.stemcr.2020.09.006 (2020).
209 Chilosi, M. *et al.* Migratory marker expression in fibroblast foci of idiopathic pulmonary fibrosis. *Respiratory research* **7**, 95, doi:10.1186/1465-9921-7-95 (2006).
210 Shaykhiev, R. Basal-like Cells in the BAL Fluid: An Echo of Regenerative Crisis in Idiopathic Pulmonary Fibrosis Lungs. *American journal of respiratory and critical care medicine* **199**, 555-557, doi:10.1164/rccm.201808-1557ED (2019).
211 Xi, Y. *et al.* Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nat Cell Biol* **19**, 904-914, doi:10.1038/ncb3580 (2017).
212 Prasse, A. *et al.* BAL Cell Gene Expression Is Indicative of Outcome and Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **199**, 622-630, doi:10.1164/rccm.201712-2551OC (2019).
213 B, B. M. *et al.* Animal models of fibrotic lung disease. *Am J Respir Cell Mol Biol* **49**, 167-179, doi:10.1165/rcmb.2013-0094TR (2013).
214 Zhao, J. *et al.* Smad3 deficiency attenuates bleomycin-induced pulmonary fibrosis in mice. *Am J Physiol Lung Cell Mol Physiol* **282**, L585-593, doi:10.1152/ajplung.00151.2001 (2002).
215 Jenkins, R. G. *et al.* An Official American Thoracic Society Workshop Report: Use of Animal Models for the Preclinical Assessment of Potential Therapies for Pulmonary Fibrosis. *Am J Respir Cell Mol Biol* **56**, 667-679, doi:10.1165/rcmb.2017-0096ST (2017).
216 Schiller, H. B. *et al.* Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol Syst Biol* **11**, 819, doi:10.15252/msb.20156123 (2015).
217 Beura, L. K. *et al.* Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* **532**, 512-516, doi:10.1038/nature17655 (2016).
218 Basil, M. C. & Morrisey, E. E. Lung regeneration: a tale of mice and men. *Semin Cell Dev Biol* **100**, 88-100, doi:10.1016/j.semcdb.2019.11.006 (2020).
219 Invernizzi, R. *et al.* Bacterial burden in the lower airways predicts disease progression in idiopathic pulmonary fibrosis and is independent of radiological disease extent. *The European respiratory journal* **55**, doi:10.1183/13993003.01519-2019 (2020).
220 Du Rand, I. A. *et al.* Summary of the British Thoracic Society guideline for diagnostic flexible bronchoscopy in adults. *Thorax* **68**, 786-787, doi:10.1136/thoraxjnl-2013-203629 (2013).
221 Molyneaux, P. L. *et al.* Bronchoalveolar Lavage is Safe and Well Tolerated in Individuals with Idiopathic Pulmonary Fibrosis: An Analysis of the PROFILE Study. *American journal of respiratory and critical care medicine*, doi:10.1164/rccm.202004-1138LE (2020).
222 Ogger, P. P. *et al.* Itaconate controls the severity of pulmonary fibrosis. *Sci Immunol* **5**, doi:10.1126/sciimmunol.abc1884 (2020).
223 Yang, I. V., Fingerlin, T. E., Evans, C. M., Schwarz, M. I. & Schwartz, D. A. MUC5B and Idiopathic Pulmonary Fibrosis. *Ann Am Thorac Soc* **12 Suppl 2**, S193-199, doi:10.1513/AnnalsATS.201503-110AW (2015).
224 Keerthisingam, C. B. *et al.* Cyclooxygenase-2 deficiency results in a loss of the anti-proliferative response to transforming growth factor-beta in human fibrotic lung

fibroblasts and promotes bleomycin-induced pulmonary fibrosis in mice. *The American journal of pathology* **158**, 1411-1422, doi:10.1016/s0002-9440(10)64092-8 (2001).
225 Stewart, C. E., Torr, E. E., Mohd Jamili, N. H., Bosquillon, C. & Sayers, I. Evaluation of differentiated human bronchial epithelial cell culture systems for asthma research. *J Allergy (Cairo)* **2012**, 943982, doi:10.1155/2012/943982 (2012).
226 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
227 Entwistle, L. J., Puttur, F., Gregory, L. G. & Lloyd, C. M. Group 2 ILC Functional Assays in Allergic Airway Inflammation. *Methods Mol Biol* **2121**, 99-114, doi:10.1007/978-1-0716-0338-3_10 (2020).
228 Srinivasan, B. *et al.* TEER measurement techniques for in vitro barrier model systems. *J Lab Autom* **20**, 107-126, doi:10.1177/2211068214561025 (2015).
229 Chilvers, M. A. & O'Callaghan, C. Analysis of ciliary beat pattern and beat frequency using digital high speed imaging: comparison with the photomultiplier and photodiode methods. *Thorax* **55**, 314-317, doi:10.1136/thorax.55.4.314 (2000).
230 Scull, M. A. *et al.* Avian Influenza virus glycoproteins restrict virus replication and spread through human airway epithelium at temperatures of the proximal airways. *PLoS Pathog* **5**, e1000424, doi:10.1371/journal.ppat.1000424 (2009).
231 Thompson, C. I., Barclay, W. S., Zambon, M. C. & Pickles, R. J. Infection of human airway epithelium by human and avian strains of influenza a virus. *J Virol* **80**, 8060-8068, doi:10.1128/jvi.00384-06 (2006).
232 Maestre-Batlle, D. *et al.* Novel flow cytometry approach to identify bronchial epithelial cells from healthy human airways. *Sci Rep* **7**, 42214, doi:10.1038/srep42214 (2017).
233 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).
234 Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res* **43**, W566-570, doi:10.1093/nar/gkv468 (2015).
235 Goedhart, J. & Luijsterburg, M. S. VolcaNoseR is a web app for creating, exploring, labeling and sharing volcano plots. *Sci Rep* **10**, 20560, doi:10.1038/s41598-020-76603-3 (2020).
236 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
237 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
238 Ma, S. & Dai, Y. Principal component analysis based methods in bioinformatics studies. *Brief Bioinform* **12**, 714-722, doi:10.1093/bib/bbq090 (2011).
239 Jackson, N. D., Ringel, L. & Seibold, M. A. Computational Analysis of RNA-Seq Data from Airway Epithelial Cells for Studying Lung Disease. *Methods Mol Biol* **1809**, 203-235, doi:10.1007/978-1-4939-8570-8_15 (2018).
240 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

241 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
242 The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-d334, doi:10.1093/nar/gkaa1113 (2021).
243 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25-29, doi:10.1038/75556 (2000).
244 Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **8**, 1551-1566, doi:10.1038/nprot.2013.092 (2013).
245 Krämer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530, doi:10.1093/bioinformatics/btt703 (2014).
246 Calvano, S. E. *et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032-1037, doi:10.1038/nature03985 (2005).
247 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
248 Davis, A. P. *et al.* The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* **47**, D948-d954, doi:10.1093/nar/gky868 (2019).
249 Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* **2016**, doi:10.1093/database/baw100 (2016).
250 Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782, doi:10.1038/s41587-019-0114-2 (2019).
251 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).
252 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
253 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e324, doi:10.1016/j.cels.2019.03.003 (2019).
254 Shao, X. *et al.* scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data. *iScience* **23**, 100882, doi:10.1016/j.isci.2020.100882 (2020).
255 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* **34**, 267-273, doi:10.1038/ng1180 (2003).
256 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).
257 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

258 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477, doi:10.1186/s12864-018-4772-0 (2018).
259 Zaragosi, L. E., Deprez, M. & Barbry, P. Using single-cell RNA sequencing to unravel cell lineage relationships in the respiratory tract. *Biochem Soc Trans* **48**, 327-336, doi:10.1042/bst20191010 (2020).
260 Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347-353, doi:10.1038/s41586-018-0698-6 (2018).
261 Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506, doi:10.1038/s41596-020-0292-x (2020).
262 Kaukonen, R., Jacquemet, G., Hamidi, H. & Ivaska, J. Cell-derived matrices for studying cell proliferation and directional migration in a complex 3D microenvironment. *Nat Protoc* **12**, 2376-2390, doi:10.1038/nprot.2017.107 (2017).
263 Puttur, F. *et al.* Pulmonary environmental cues drive group 2 innate lymphoid cell dynamics in mice and humans. *Sci Immunol* **4**, doi:10.1126/sciimmunol.aav7638 (2019).
264 Wang, Y., Jeong, Y., Jhiang, S. M., Yu, L. & Menq, C. H. Quantitative characterization of cell behaviors through cell cycle progression via automated cell tracking. *PloS one* **9**, e98762, doi:10.1371/journal.pone.0098762 (2014).
265 de Chaumont, F. *et al.* Icy: an open bioimage informatics platform for extended reproducible research. *Nat Methods* **9**, 690-696, doi:10.1038/nmeth.2075 (2012).
266 Lennon, R. *et al.* Global analysis reveals the complexity of the human glomerular extracellular matrix. *J Am Soc Nephrol* **25**, 939-951, doi:10.1681/asn.2013030233 (2014).
267 Selman, M. & Pardo, A. The leading role of epithelial cells in the pathogenesis of idiopathic pulmonary fibrosis. *Cell Signal* **66**, 109482, doi:10.1016/j.cellsig.2019.109482 (2020).
268 Jenkins, G. *et al.* In search of the fibrotic epithelial cell: opportunities for a collaborative network. *Thorax* **67**, 179-182, doi:10.1136/thoraxjnl-2011-200195 (2012).
269 Marmai, C. *et al.* Alveolar epithelial cells express mesenchymal proteins in patients with idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* **301**, L71-78, doi:10.1152/ajplung.00212.2010 (2011).
270 Gindele, J. A. *et al.* Intermittent exposure to whole cigarette smoke alters the differentiation of primary small airway epithelial cells in the air-liquid interface culture. *Sci Rep* **10**, 6257, doi:10.1038/s41598-020-63345-5 (2020).
271 Mertens, T. C. J., Karmouty-Quintana, H., Taube, C. & Hiemstra, P. S. Use of airway epithelial cell culture to unravel the pathogenesis and study treatment in obstructive airway diseases. *Pulm Pharmacol Ther* **45**, 101-113, doi:10.1016/j.pupt.2017.05.008 (2017).
272 Heijink, I. H., Noordhoek, J. A., Timens, W., van Oosterhout, A. J. & Postma, D. S. Abnormalities in airway epithelial junction formation in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* **189**, 1439-1442, doi:10.1164/rccm.201311-1982LE (2014).

273 Hackett, T. L. *et al.* Intrinsic phenotypic differences of asthmatic epithelium and its inflammatory responses to respiratory syncytial virus and air pollution. *Am J Respir Cell Mol Biol* **45**, 1090-1100, doi:10.1165/rcmb.2011-0031OC (2011).
274 Hiemstra, P. S., Tetley, T. D. & Janes, S. M. Airway and alveolar epithelial cells in culture. *The European respiratory journal* **54**, doi:10.1183/13993003.00742-2019 (2019).
275 Ghosh, B. *et al.* Strong correlation between air-liquid interface cultures and in vivo transcriptomics of nasal brush biopsy. *Am J Physiol Lung Cell Mol Physiol* **318**, L1056-l1062, doi:10.1152/ajplung.00050.2020 (2020).
276 Dvorak, A., Tilley, A. E., Shaykhiev, R., Wang, R. & Crystal, R. G. Do airway epithelium air-liquid cultures represent the in vivo airway epithelium transcriptome? *Am J Respir Cell Mol Biol* **44**, 465-473, doi:10.1165/rcmb.2009-0453OC (2011).
277 Wark, P. A. *et al.* Asthmatic bronchial epithelial cells have a deficient innate immune response to infection with rhinovirus. *J Exp Med* **201**, 937-947, doi:10.1084/jem.20041901 (2005).
278 Zhu, N. *et al.* Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway epithelial cells. *Nature communications* **11**, 3910, doi:10.1038/s41467-020-17796-z (2020).
279 Maunders, H., Patwardhan, S., Phillips, J., Clack, A. & Richter, A. Human bronchial epithelial cell transcriptome: gene expression changes following acute exposure to whole cigarette smoke in vitro. *Am J Physiol Lung Cell Mol Physiol* **292**, L1248-1256, doi:10.1152/ajplung.00290.2006 (2007).
280 Zarcone, M. C. *et al.* Cellular response of mucociliary differentiated primary bronchial epithelial cells to diesel exhaust. *Am J Physiol Lung Cell Mol Physiol* **311**, L111-123, doi:10.1152/ajplung.00064.2016 (2016).
281 Wawrzyniak, P. *et al.* Regulation of bronchial epithelial barrier integrity by type 2 cytokines and histone deacetylases in asthmatic patients. *J Allergy Clin Immunol* **139**, 93-103, doi:10.1016/j.jaci.2016.03.050 (2017).
282 Amatngalim, G. D. *et al.* Cigarette Smoke Modulates Repair and Innate Immunity following Injury to Airway Epithelial Cells. *PloS one* **11**, e0166255, doi:10.1371/journal.pone.0166255 (2016).
283 Oltean, A., Schaffer, A. J., Bayly, P. V. & Brody, S. L. Quantifying Ciliary Dynamics during Assembly Reveals Stepwise Waveform Maturation in Airway Cells. *Am J Respir Cell Mol Biol* **59**, 511-522, doi:10.1165/rcmb.2017-0436OC (2018).
284 Osei, E. T. *et al.* Interleukin-1α drives the dysfunctional cross-talk of the airway epithelium and lung fibroblasts in COPD. *The European respiratory journal* **48**, 359-369, doi:10.1183/13993003.01911-2015 (2016).
285 Prasad, S., Hogaboam, C. M. & Jarai, G. Deficient repair response of IPF fibroblasts in a co-culture model of epithelial injury and repair. *Fibrogenesis & tissue repair* **7**, 7, doi:10.1186/1755-1536-7-7 (2014).
286 Correll, K. A. *et al.* Transitional human alveolar type II epithelial cells suppress extracellular matrix and growth factor gene expression in lung fibroblasts. *Am J Physiol Lung Cell Mol Physiol* **317**, L283-l294, doi:10.1152/ajplung.00337.2018 (2019).
287 Conforti, F. *et al.* Paracrine SPARC signaling dysregulates alveolar epithelial barrier integrity and function in lung fibrosis. *Cell Death Discov* **6**, 54, doi:10.1038/s41420-020-0289-9 (2020).

288 Watson, J. H. & Brinkman, G. L. ELECTRON MICROSCOPY OF THE EPITHELIAL CELLS OF NORMAL AND BRONCHITIC HUMAN BRONCHUS. *Am Rev Respir Dis* **90**, 851-866, doi:10.1164/arrd.1964.90.6.851 (1964).
289 Mercer, R. R., Russell, M. L., Roggli, V. L. & Crapo, J. D. Cell number and distribution in human and rat airways. *American Journal of Respiratory Cell and Molecular Biology* **10**, 613-624, doi:10.1165/ajrcmb.10.6.8003339 (1994).
290 Hicks, W., Jr. *et al.* Isolation and characterization of basal cells from human upper respiratory epithelium. *Exp Cell Res* **237**, 357-363, doi:10.1006/excr.1997.3796 (1997).
291 Crosby, L. M. & Waters, C. M. Epithelial repair mechanisms in the lung. *Am J Physiol Lung Cell Mol Physiol* **298**, L715-731, doi:10.1152/ajplung.00361.2009 (2010).
292 Jonkman, J. E. *et al.* An introduction to the wound healing assay using live-cell microscopy. *Cell Adh Migr* **8**, 440-451, doi:10.4161/cam.36224 (2014).
293 Comer, D. M., Kidney, J. C., Ennis, M. & Elborn, J. S. Airway epithelial cell apoptosis and inflammation in COPD, smokers and nonsmokers. *The European respiratory journal* **41**, 1058-1067, doi:10.1183/09031936.00063112 (2013).
294 Chilvers, M. A., Rutman, A. & O'Callaghan, C. Functional analysis of cilia and ciliated epithelial ultrastructure in healthy children and young adults. *Thorax* **58**, 333-338, doi:10.1136/thorax.58.4.333 (2003).
295 Garnero, P. *et al.* The collagenolytic activity of cathepsin K is unique among mammalian proteinases. *J Biol Chem* **273**, 32347-32352, doi:10.1074/jbc.273.48.32347 (1998).
296 Sakai, N. & Tager, A. M. Fibrosis of two: Epithelial cell-fibroblast interactions in pulmonary fibrosis. *Biochim Biophys Acta* **1832**, 911-921, doi:10.1016/j.bbadis.2013.03.001 (2013).
297 Sirianni, F. E., Chu, F. S. & Walker, D. C. Human alveolar wall fibroblasts directly link epithelial type 2 cells to capillary endothelium. *American journal of respiratory and critical care medicine* **168**, 1532-1537, doi:10.1164/rccm.200303-371OC (2003).
298 Ryu, C., Homer, R. J. & Herzog, E. L. The Airway in Idiopathic Pulmonary Fibrosis: Protecting the Lung or Promoting Disease? *American journal of respiratory and critical care medicine* **193**, 1081-1082, doi:10.1164/rccm.201601-0055ED (2016).
299 Reeves, S. R. *et al.* Stability of gene expression by primary bronchial epithelial cells over increasing passage number. *BMC Pulm Med* **18**, 91, doi:10.1186/s12890-018-0652-2 (2018).
300 Crystal, R. G. Airway basal cells. The "smoking gun" of chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* **190**, 1355-1362, doi:10.1164/rccm.201408-1492PP (2014).
301 Gruenert, D. C., Finkbeiner, W. E. & Widdicombe, J. H. Culture and transformation of human airway epithelial cells. *Am J Physiol* **268**, L347-360, doi:10.1152/ajplung.1995.268.3.L347 (1995).
302 Aghapour, M., Raee, P., Moghaddam, S. J., Hiemstra, P. S. & Heijink, I. H. Airway Epithelial Barrier Dysfunction in Chronic Obstructive Pulmonary Disease: Role of Cigarette Smoke Exposure. *Am J Respir Cell Mol Biol* **58**, 157-169, doi:10.1165/rcmb.2017-0200TR (2018).

303 Mathai, S. K. *et al.* Desmoplakin Variants Are Associated with Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **193**, 1151-1160, doi:10.1164/rccm.201509-1863OC (2016).
304 Daccord, C. & Maher, T. M. Recent advances in understanding idiopathic pulmonary fibrosis. *F1000Res* **5**, doi:10.12688/f1000research.8209.1 (2016).
305 Moore, B. B. & Moore, T. A. Viruses in Idiopathic Pulmonary Fibrosis. Etiology and Exacerbation. *Ann Am Thorac Soc* **12 Suppl 2**, S186-192, doi:10.1513/AnnalsATS.201502-088AW (2015).
306 Benam, K. H., Denney, L. & Ho, L. P. How the Respiratory Epithelium Senses and Reacts to Influenza Virus. *Am J Respir Cell Mol Biol* **60**, 259-268, doi:10.1165/rcmb.2018-0247TR (2019).
307 Jolly, L. *et al.* Influenza promotes collagen deposition via αvβ6 integrin-mediated transforming growth factor β activation. *J Biol Chem* **289**, 35246-35263, doi:10.1074/jbc.M114.582262 (2014).
308 Marshall, L. J., Perks, B., Ferkol, T. & Shute, J. K. IL-8 released constitutively by primary bronchial epithelial cells in culture forms an inactive complex with secretory component. *J Immunol* **167**, 2816-2823, doi:10.4049/jimmunol.167.5.2816 (2001).
309 Ziegenhagen, M. W., Zabel, P., Zissel, G., Schlaak, M. & Müller-Quernheim, J. Serum level of interleukin 8 is elevated in idiopathic pulmonary fibrosis and indicates disease activity. *American journal of respiratory and critical care medicine* **157**, 762-768, doi:10.1164/ajrccm.157.3.9705014 (1998).
310 Yang, L. *et al.* IL-8 mediates idiopathic pulmonary fibrosis mesenchymal progenitor cell fibrogenicity. *Am J Physiol Lung Cell Mol Physiol* **314**, L127-l136, doi:10.1152/ajplung.00200.2017 (2018).
311 Leung, C., Wadsworth, S. J., Yang, S. J. & Dorscheid, D. R. Structural and functional variations in human bronchial epithelial cells cultured in air-liquid interface using different growth media. *Am J Physiol Lung Cell Mol Physiol* **318**, L1063-l1073, doi:10.1152/ajplung.00190.2019 (2020).
312 Richter, A. *et al.* Autocrine ligands for the epidermal growth factor receptor mediate interleukin-8 release from bronchial epithelial cells in response to cigarette smoke. *Am J Respir Cell Mol Biol* **27**, 85-90, doi:10.1165/ajrcmb.27.1.4789 (2002).
313 Rosas, I. O. *et al.* MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med* **5**, e93, doi:10.1371/journal.pmed.0050093 (2008).
314 Gharib, S. A. *et al.* Matrix metalloproteinase-7 coordinates airway epithelial injury response and differentiation of ciliated cells. *Am J Respir Cell Mol Biol* **48**, 390-396, doi:10.1165/rcmb.2012-0083OC (2013).
315 Yu, H. *et al.* Transgelin is a direct target of TGF-beta/Smad3-dependent epithelial cell migration in lung fibrosis. *Faseb j* **22**, 1778-1789, doi:10.1096/fj.07-083857 (2008).
316 Zhao, F. Y. *et al.* G-CSF Inhibits Pulmonary Fibrosis by Promoting BMSC Homing to the Lungs via SDF-1/CXCR4 Chemotaxis. *Sci Rep* **10**, 10515, doi:10.1038/s41598-020-65580-2 (2020).
317 Zhang, F. *et al.* Mobilization of bone marrow cells by CSF3 protects mice from bleomycin-induced lung injury. *Respiration* **82**, 358-368, doi:10.1159/000328762 (2011).

318 Ashitani, J. *et al.* Granulocyte-colony stimulating factor levels in bronchoalveolar lavage fluid from patients with idiopathic pulmonary fibrosis. *Thorax* **54**, 1015-1020, doi:10.1136/thx.54.11.1015 (1999).
319 Peng, T. *et al.* Hedgehog actively maintains adult lung quiescence and regulates repair and regeneration. *Nature* **526**, 578-582, doi:10.1038/nature14984 (2015).
320 Blokland, K. E. C. *et al.* Senescence of IPF Lung Fibroblasts Disrupt Alveolar Epithelial Cell Proliferation and Promote Migration in Wound Healing. *Pharmaceutics* **12**, doi:10.3390/pharmaceutics12040389 (2020).
321 Morishima, Y. *et al.* Triggering the induction of myofibroblast and fibrogenesis by airway epithelial shedding. *Am J Respir Cell Mol Biol* **24**, 1-11, doi:10.1165/ajrcmb.24.1.4040 (2001).
322 Boyd, D. F. *et al.* Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* **587**, 466-471, doi:10.1038/s41586-020-2877-5 (2020).
323 Brewster, C. E. *et al.* Myofibroblasts and subepithelial fibrosis in bronchial asthma. *Am J Respir Cell Mol Biol* **3**, 507-511, doi:10.1165/ajrcmb/3.5.507 (1990).
324 Pezzulo, A. A. *et al.* The air-liquid interface and use of primary cell cultures are important to recapitulate the transcriptional profile of in vivo airway epithelia. *Am J Physiol Lung Cell Mol Physiol* **300**, L25-31, doi:10.1152/ajplung.00256.2010 (2011).
325 Kicic, A., Sutanto, E. N., Stevens, P. T., Knight, D. A. & Stick, S. M. Intrinsic biochemical and functional differences in bronchial epithelial cells of children with asthma. *American journal of respiratory and critical care medicine* **174**, 1110-1118, doi:10.1164/rccm.200603-392OC (2006).
326 Gras, D. *et al.* An ex vivo model of severe asthma using reconstituted human bronchial epithelium. *J Allergy Clin Immunol* **129**, 1259-1266.e1251, doi:10.1016/j.jaci.2012.01.073 (2012).
327 Hiemstra, P. S., McCray, P. B., Jr. & Bals, R. The innate immune function of airway epithelial cells in inflammatory lung disease. *The European respiratory journal* **45**, 1150-1162, doi:10.1183/09031936.00141514 (2015).
328 Kusko, R. L. *et al.* Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **194**, 948-960, doi:10.1164/rccm.201510-2026OC (2016).
329 Kim, S. Y. *et al.* Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *The Lancet. Respiratory medicine* **3**, 473-482, doi:10.1016/s2213-2600(15)00140-x (2015).
330 Vukmirovic, M. & Kaminski, N. Impact of Transcriptomics on Our Understanding of Pulmonary Fibrosis. *Front Med (Lausanne)* **5**, 87, doi:10.3389/fmed.2018.00087 (2018).
331 Luzina, I. G. *et al.* Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis. *Cellular immunology* **325**, 1-13, doi:10.1016/j.cellimm.2018.01.002 (2018).
332 Sala, M. A. *et al.* Inflammatory pathways are upregulated in the nasal epithelium in patients with idiopathic pulmonary fibrosis. *Respiratory research* **19**, 233, doi:10.1186/s12931-018-0932-7 (2018).

333 Desch, A. N. *et al.* Flow Cytometric Analysis of Mononuclear Phagocytes in Nondiseased Human Lung and Lung-Draining Lymph Nodes. *American journal of respiratory and critical care medicine* **193**, 614-626, doi:10.1164/rccm.201507-1376OC (2016).
334 Bharat, A. *et al.* Flow Cytometry Reveals Similarities Between Lung Macrophages in Humans and Mice. *Am J Respir Cell Mol Biol* **54**, 147-149, doi:10.1165/rcmb.2015-0147LE (2016).
335 Gallego Romero, I., Pai, A. A., Tung, J. & Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol* **12**, 42, doi:10.1186/1741-7007-12-42 (2014).
336 Chen, E. A. *et al.* Effect of RNA integrity on uniquely mapped reads in RNA-Seq. *BMC Res Notes* **7**, 753, doi:10.1186/1756-0500-7-753 (2014).
337 Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3, doi:10.1186/1471-2199-7-3 (2006).
338 Bloom, C. I. *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PloS one* **8**, e70630, doi:10.1371/journal.pone.0070630 (2013).
339 Litosh, V. A. *et al.* Calpain-14 and its association with eosinophilic esophagitis. *J Allergy Clin Immunol* **139**, 1762-1771.e1767, doi:10.1016/j.jaci.2016.09.027 (2017).
340 Zilionis, R. *et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317-1334.e1310, doi:10.1016/j.immuni.2019.03.009 (2019).
341 Pettersson, M., Dannaeus, K., Nilsson, K. & Jönsson, J. I. Isolation of MYADM, a novel hematopoietic-associated marker gene expressed in multipotent progenitor cells and up-regulated during myeloid differentiation. *J Leukoc Biol* **67**, 423-431, doi:10.1002/jlb.67.3.423 (2000).
342 Stein, M., Keshav, S., Harris, N. & Gordon, S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. *J Exp Med* **176**, 287-292, doi:10.1084/jem.176.1.287 (1992).
343 Tarique, A. A. *et al.* Phenotypic, functional, and plasticity features of classical and alternatively activated human macrophages. *Am J Respir Cell Mol Biol* **53**, 676-688, doi:10.1165/rcmb.2015-0012OC (2015).
344 Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-1076, doi:10.1038/nature08975 (2010).
345 Fessler, M. B. A New Frontier in Immunometabolism. Cholesterol in Lung Health and Disease. *Ann Am Thorac Soc* **14**, S399-s405, doi:10.1513/AnnalsATS.201702-136AW (2017).
346 Nakata, K. *et al.* Granulocyte-macrophage colony-stimulating factor promotes the proliferation of human alveolar macrophages in vitro. *J Immunol* **147**, 1266-1272 (1991).
347 Schneider, C. *et al.* Induction of the nuclear receptor PPAR-γ by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages. *Nature immunology* **15**, 1026-1037, doi:10.1038/ni.3005 (2014).
348 Shibata, Y. *et al.* GM-CSF regulates alveolar macrophage differentiation and innate immunity in the lung through PU.1. *Immunity* **15**, 557-567, doi:10.1016/s1074-7613(01)00218-7 (2001).

349 Yu, X. *et al.* The Cytokine TGF-β Promotes the Development and Homeostasis of Alveolar Macrophages. *Immunity* **47**, 903-912.e904, doi:10.1016/j.immuni.2017.10.007 (2017).
350 Branchett, W. J. & Lloyd, C. M. Regulatory cytokine function in the respiratory tract. *Mucosal Immunol* **12**, 589-600, doi:10.1038/s41385-019-0158-0 (2019).
351 Meng, X. M., Nikolic-Paterson, D. J. & Lan, H. Y. TGF-β: the master regulator of fibrosis. *Nat Rev Nephrol* **12**, 325-338, doi:10.1038/nrneph.2016.48 (2016).
352 Mercier, S. *et al.* Mutations in FAM111B cause hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis. *Am J Hum Genet* **93**, 1100-1107, doi:10.1016/j.ajhg.2013.10.013 (2013).
353 Krausgruber, T. *et al.* Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296-302, doi:10.1038/s41586-020-2424-4 (2020).
354 Wang, C. *et al.* The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* **32**, 926-932, doi:10.1038/nbt.3001 (2014).
355 Champliaud, M. F., Burgeson, R. E., Jin, W., Baden, H. P. & Olson, P. F. cDNA cloning and characterization of sciellin, a LIM domain protein of the keratinocyte cornified envelope. *J Biol Chem* **273**, 31547-31554, doi:10.1074/jbc.273.47.31547 (1998).
356 Mendez-Enriquez, E. & García-Zepeda, E. A. The multiple faces of CCL13 in immunity and inflammation. *Inflammopharmacology* **21**, 397-406, doi:10.1007/s10787-013-0177-5 (2013).
357 Lamkhioued, B. *et al.* Monocyte chemoattractant protein (MCP)-4 expression in the airways of patients with asthma. Induction in epithelial cells and mononuclear cells by proinflammatory cytokines. *American journal of respiratory and critical care medicine* **162**, 723-732, doi:10.1164/ajrccm.162.2.9901080 (2000).
358 McDonough, J. E. *et al.* Transcriptional regulatory model of fibrosis progression in the human lung. *JCI insight* **4**, doi:10.1172/jci.insight.131597 (2019).
359 Neighbors, M. *et al.* Prognostic and predictive biomarkers for patients with idiopathic pulmonary fibrosis treated with pirfenidone: post-hoc assessment of the CAPACITY and ASCEND trials. *The Lancet. Respiratory medicine* **6**, 615-626, doi:10.1016/s2213-2600(18)30185-1 (2018).
360 Foster, M. W. *et al.* Quantitative proteomics of bronchoalveolar lavage fluid in idiopathic pulmonary fibrosis. *J Proteome Res* **14**, 1238-1249, doi:10.1021/pr501149m (2015).
361 Schaefer, D. *et al.* Endothelial and epithelial expression of eotaxin-2 (CCL24) in nasal polyps. *Int Arch Allergy Immunol* **140**, 205-214, doi:10.1159/000093206 (2006).
362 Ancuta, P. *et al.* CD16+ monocyte-derived macrophages activate resting T cells for HIV infection by producing CCR3 and CCR4 ligands. *J Immunol* **176**, 5760-5771, doi:10.4049/jimmunol.176.10.5760 (2006).
363 Forssmann, U. *et al.* Eotaxin-2, a novel CC chemokine that is selective for the chemokine receptor CCR3, and acts like eotaxin on human eosinophil and basophil leukocytes. *J Exp Med* **185**, 2171-2176, doi:10.1084/jem.185.12.2171 (1997).
364 Kohan, M., Puxeddu, I., Reich, R., Levi-Schaffer, F. & Berkman, N. Eotaxin-2/CCL24 and eotaxin-3/CCL26 exert differential profibrogenic effects on human lung fibroblasts. *Ann Allergy Asthma Immunol* **104**, 66-72, doi:10.1016/j.anai.2009.11.003 (2010).

365 Beck, L. A. *et al.* Functional analysis of the chemokine receptor CCR3 on airway epithelial cells. *J Immunol* **177**, 3344-3354, doi:10.4049/jimmunol.177.5.3344 (2006).
366 Mor, A. *et al.* Blockade of CCL24 with a monoclonal antibody ameliorates experimental dermal and pulmonary fibrosis. *Ann Rheum Dis* **78**, 1260-1268, doi:10.1136/annrheumdis-2019-215119 (2019).
367 Bdeir, K. *et al.* Platelet-Specific Chemokines Contribute to the Pathogenesis of Acute Lung Injury. *Am J Respir Cell Mol Biol* **56**, 261-270, doi:10.1165/rcmb.2015-0245OC (2017).
368 Simonson, L. *et al.* Keratin 13 deficiency causes white sponge nevus in mice. *Dev Biol* **468**, 146-153, doi:10.1016/j.ydbio.2020.07.016 (2020).
369 Madissoon, E. *et al.* scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol* **21**, 1, doi:10.1186/s13059-019-1906-x (2019).
370 van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods* **14**, 935-936, doi:10.1038/nmeth.4437 (2017).
371 Lee, J. *et al.* Bronchoalveolar lavage (BAL) cells in idiopathic pulmonary fibrosis express a complex pro-inflammatory, pro-repair, angiogenic activation pattern, likely associated with macrophage iron accumulation. *PloS one* **13**, e0194803, doi:10.1371/journal.pone.0194803 (2018).
372 Yogo, Y. *et al.* Macrophage derived chemokine (CCL22), thymus and activation-regulated chemokine (CCL17), and CCR4 in idiopathic pulmonary fibrosis. *Respiratory research* **10**, 80, doi:10.1186/1465-9921-10-80 (2009).
373 Craig, V. J., Zhang, L., Hagood, J. S. & Owen, C. A. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* **53**, 585-600, doi:10.1165/rcmb.2015-0020TR (2015).
374 Maher, T. M. *et al.* An epithelial biomarker signature for idiopathic pulmonary fibrosis: an analysis from the multicentre PROFILE cohort study. *The Lancet. Respiratory medicine* **5**, 946-955, doi:10.1016/s2213-2600(17)30430-7 (2017).
375 Fujishima, S. *et al.* Production and activation of matrix metalloproteinase 7 (matrilysin 1) in the lungs of patients with idiopathic pulmonary fibrosis. *Arch Pathol Lab Med* **134**, 1136-1142, doi:10.1043/2009-0144-oa.1 (2010).
376 McGuire, J. K., Li, Q. & Parks, W. C. Matrilysin (matrix metalloproteinase-7) mediates E-cadherin ectodomain shedding in injured lung epithelium. *The American journal of pathology* **162**, 1831-1843, doi:10.1016/s0002-9440(10)64318-0 (2003).
377 García-Alvarez, J. *et al.* Tissue inhibitor of metalloproteinase-3 is up-regulated by transforming growth factor-beta1 in vitro and expressed in fibroblastic foci in vivo in idiopathic pulmonary fibrosis. *Exp Lung Res* **32**, 201-214, doi:10.1080/01902140600817481 (2006).
378 Becher, B., Tugues, S. & Greter, M. GM-CSF: From Growth Factor to Central Mediator of Tissue Inflammation. *Immunity* **45**, 963-973, doi:10.1016/j.immuni.2016.10.026 (2016).
379 Trapnell, B. C. *et al.* Pulmonary alveolar proteinosis. *Nat Rev Dis Primers* **5**, 16, doi:10.1038/s41572-019-0066-3 (2019).

380 Trapnell, B. C. & Whitsett, J. A. Gm-CSF regulates pulmonary surfactant homeostasis and alveolar macrophage-mediated innate host defense. *Annu Rev Physiol* **64**, 775-802, doi:10.1146/annurev.physiol.64.090601.113847 (2002).
381 Joshi, N. *et al.* A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signalling in monocyte-derived alveolar macrophages. *The European respiratory journal* **55**, doi:10.1183/13993003.00646-2019 (2020).
382 Xue, J. *et al.* Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* **40**, 274-288, doi:10.1016/j.immuni.2014.01.006 (2014).
383 Kiemle-Kallee, J. *et al.* Alveolar macrophages in idiopathic pulmonary fibrosis display a more monocyte-like immunophenotype and an increased release of free oxygen radicals. *The European respiratory journal* **4**, 400-406 (1991).
384 Engler, A. E. *et al.* Airway-Associated Macrophages in Homeostasis and Repair. *Cell Rep* **33**, 108553, doi:10.1016/j.celrep.2020.108553 (2020).
385 Neumark, N., Cosme, C., Jr., Rose, K. A. & Kaminski, N. The Idiopathic Pulmonary Fibrosis Cell Atlas. *Am J Physiol Lung Cell Mol Physiol* **319**, L887-l893, doi:10.1152/ajplung.00451.2020 (2020).
386 Zeisberg, M. & Neilson, E. G. Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest* **119**, 1429-1437, doi:10.1172/jci36183 (2009).
387 Danopoulos, S. *et al.* Human lung branching morphogenesis is orchestrated by the spatiotemporal distribution of ACTA2, SOX2, and SOX9. *Am J Physiol Lung Cell Mol Physiol* **314**, L144-l149, doi:10.1152/ajplung.00379.2017 (2018).
388 Schafer, M. J. *et al.* Cellular senescence mediates fibrotic pulmonary disease. *Nature communications* **8**, 14532, doi:10.1038/ncomms14532 (2017).
389 Duclos, G. E. *et al.* Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Science Advances* **5**, eaaw3413, doi:10.1126/sciadv.aaw3413 (2019).
390 Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* **20**, 63, doi:10.1186/s13059-019-1662-y (2019).
391 Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545-554, doi:10.1093/bioinformatics/btu674 (2015).
392 Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, doi:10.1093/gigascience/giy083 (2018).
393 Revinski, D. R. *et al.* CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. *Nature communications* **9**, 4668, doi:10.1038/s41467-018-06768-z (2018).
394 Kumar, B. V. *et al.* Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep* **20**, 2921-2934, doi:10.1016/j.celrep.2017.08.078 (2017).
395 Baßler, K. *et al.* Alterations of multiple alveolar macrophage states in chronic obstructive pulmonary disease. *bioRxiv*, 2020.2005.2028.121541, doi:10.1101/2020.05.28.121541 (2020).
396 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).

397 Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* **106**, 12771-12775, doi:10.1073/pnas.0906850106 (2009).
398 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049, doi:10.1038/ncomms14049 (2017).
399 Sui, P. *et al.* Pulmonary neuroendocrine cells amplify allergic asthma responses. *Science* **360**, doi:10.1126/science.aan8546 (2018).
400 Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* **9**, 75, doi:10.1186/s13073-017-0467-4 (2017).
401 Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* **144**, 3625-3632, doi:10.1242/dev.151142 (2017).
402 Zuo, W. L. *et al.* Ontogeny and Biology of Human Small Airway Epithelial Club Cells. *American journal of respiratory and critical care medicine* **198**, 1375-1388, doi:10.1164/rccm.201710-2107OC (2018).
403 Pardo-Saganta, A. *et al.* Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell stem cell* **16**, 184-197, doi:10.1016/j.stem.2015.01.002 (2015).
404 Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649-654, doi:10.1038/s41586-018-0449-8 (2018).
405 Burel, J. G. *et al.* The Challenge of Distinguishing Cell-Cell Complexes from Singlet Cells in Non-Imaging Flow Cytometry and Single-Cell Sorting. *Cytometry A* **97**, 1127-1135, doi:10.1002/cyto.a.24027 (2020).
406 Giladi, A. *et al.* Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat Biotechnol* **38**, 629-637, doi:10.1038/s41587-020-0442-2 (2020).
407 Habiel, D. M. *et al.* Characterization of CD28(null) T cells in idiopathic pulmonary fibrosis. *Mucosal Immunol* **12**, 212-222, doi:10.1038/s41385-018-0082-8 (2019).
408 Celada, L. J. *et al.* PD-1 up-regulation on CD4(+) T cells promotes pulmonary fibrosis through STAT3-mediated IL-17A and TGF-β1 production. *Science translational medicine* **10**, doi:10.1126/scitranslmed.aar8356 (2018).
409 Mould, K. J., Jackson, N. D., Henson, P. M., Seibold, M. & Janssen, W. J. Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. *JCI insight* **4**, doi:10.1172/jci.insight.126556 (2019).
410 Pardo, A. *et al.* Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. *PLoS Med* **2**, e251, doi:10.1371/journal.pmed.0020251 (2005).
411 Burgstaller, G. *et al.* The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. *The European respiratory journal* **50**, doi:10.1183/13993003.01805-2016 (2017).
412 Rennard, S. I. & Crystal, R. G. Fibronectin in human bronchopulmonary lavage fluid. Elevation in patients with interstitial lung disease. *J Clin Invest* **69**, 113-122, doi:10.1172/jci110421 (1982).
413 Adachi, K. *et al.* Evaluation of fibronectin gene expression by in situ hybridization. Differential expression of the fibronectin gene among populations of human alveolar macrophages. *The American journal of pathology* **133**, 193-203 (1988).

414 Rennard, S. I., Hunninghake, G. W., Bitterman, P. B. & Crystal, R. G. Production of fibronectin by the human alveolar macrophage: mechanism for the recruitment of fibroblasts to sites of tissue injury in interstitial lung diseases. *Proc Natl Acad Sci U S A* **78**, 7147-7151, doi:10.1073/pnas.78.11.7147 (1981).
415 Arebro, J. *et al.* Antigen-presenting epithelial cells can play a pivotal role in airway allergy. *J Allergy Clin Immunol* **137**, 957-960.e957, doi:10.1016/j.jaci.2015.08.053 (2016).
416 Rossi, G. A. *et al.* Human ciliated bronchial epithelial cells: expression of the HLA-DR antigens and of the HLA-DR alpha gene, modulation of the HLA-DR antigens by gamma-interferon and antigen-presenting function in the mixed leukocyte reaction. *Am J Respir Cell Mol Biol* **3**, 431-439, doi:10.1165/ajrcmb/3.5.431 (1990).
417 Papi, A. *et al.* Rhinovirus infection induces major histocompatibility complex class I and costimulatory molecule upregulation on respiratory epithelial cells. *J Infect Dis* **181**, 1780-1784, doi:10.1086/315463 (2000).
418 Burkhardt, A. M. *et al.* CXCL17 is a mucosal chemokine elevated in idiopathic pulmonary fibrosis that exhibits broad antimicrobial activity. *J Immunol* **188**, 6399-6406, doi:10.4049/jimmunol.1102903 (2012).
419 Ernst, S. *et al.* An annexin 1 N-terminal peptide activates leukocytes by triggering different members of the formyl peptide receptor family. *J Immunol* **172**, 7669-7676, doi:10.4049/jimmunol.172.12.7669 (2004).
420 Reynolds, S. D. *et al.* CCSP regulates cross talk between secretory cells and both ciliated cells and macrophages of the conducting airway. *Am J Physiol Lung Cell Mol Physiol* **293**, L114-123, doi:10.1152/ajplung.00014.2007 (2007).
421 Bin, L. H., Nielson, L. D., Liu, X., Mason, R. J. & Shu, H. B. Identification of uteroglobin-related protein 1 and macrophage scavenger receptor with collagenous structure as a lung-specific ligand-receptor pair. *J Immunol* **171**, 924-930, doi:10.4049/jimmunol.171.2.924 (2003).
422 Jaspars, L. H., Vos, W., Aust, G., Van Lier, R. A. & Hamann, J. Tissue distribution of the human CD97 EGF-TM7 receptor. *Tissue Antigens* **57**, 325-331, doi:10.1034/j.1399-0039.2001.057004325.x (2001).
423 Hamann, J. *et al.* CD97 in leukocyte trafficking. *Adv Exp Med Biol* **706**, 128-137, doi:10.1007/978-1-4419-7913-1_11 (2010).
424 Calandra, T. & Roger, T. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat Rev Immunol* **3**, 791-800, doi:10.1038/nri1200 (2003).
425 Leng, L. *et al.* MIF signal transduction initiated by binding to CD74. *J Exp Med* **197**, 1467-1476, doi:10.1084/jem.20030286 (2003).
426 Sheppard, D. Functions of pulmonary epithelial integrins: from development to disease. *Physiol Rev* **83**, 673-686, doi:10.1152/physrev.00033.2002 (2003).
427 Prasse, A. *et al.* Serum CC-chemokine ligand 18 concentration predicts outcome in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **179**, 717-723, doi:10.1164/rccm.200808-1201OC (2009).
428 Prasse, A. *et al.* A vicious circle of alveolar macrophages and fibroblasts perpetuates pulmonary fibrosis via CCL18. *American journal of respiratory and critical care medicine* **173**, 781-792, doi:10.1164/rccm.200509-1518OC (2006).
429 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

430 Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, doi:10.1126/science.aba3066 (2020).
431 Jonsdottir, H. R. *et al.* Basal cells of the human airways acquire mesenchymal traits in idiopathic pulmonary fibrosis and in culture. *Lab Invest* **95**, 1418-1428, doi:10.1038/labinvest.2015.114 (2015).
432 Cassandras, M. *et al.* Gli1(+) mesenchymal stromal cells form a pathological niche to promote airway progenitor metaplasia in the fibrotic lung. *Nat Cell Biol* **22**, 1295-1306, doi:10.1038/s41556-020-00591-9 (2020).
433 Zhou, Y. *et al.* Extracellular matrix in lung development, homeostasis and disease. *Matrix Biol* **73**, 77-104, doi:10.1016/j.matbio.2018.03.005 (2018).
434 Herrera, J., Henke, C. A. & Bitterman, P. B. Extracellular matrix as a driver of progressive fibrosis. *J Clin Invest* **128**, 45-53, doi:10.1172/jci93557 (2018).
435 Liu, F. *et al.* Mechanosignaling through YAP and TAZ drives fibroblast activation and fibrosis. *Am J Physiol Lung Cell Mol Physiol* **308**, L344-357, doi:10.1152/ajplung.00300.2014 (2015).
436 Chen, H. *et al.* Mechanosensing by the α6-integrin confers an invasive fibroblast phenotype and mediates lung fibrosis. *Nature communications* **7**, 12564, doi:10.1038/ncomms12564 (2016).
437 Yamada, K. M. & Sixt, M. Mechanisms of 3D cell migration. *Nat Rev Mol Cell Biol* **20**, 738-752, doi:10.1038/s41580-019-0172-9 (2019).
438 Coraux, C., Roux, J., Jolly, T. & Birembaut, P. Epithelial cell-extracellular matrix interactions and stem cells in airway epithelial regeneration. *Proc Am Thorac Soc* **5**, 689-694, doi:10.1513/pats.200801-010AW (2008).
439 Mostaço-Guidolin, L. B. *et al.* Collagen morphology and texture analysis: from statistics to classification. *Sci Rep* **3**, 2190, doi:10.1038/srep02190 (2013).
440 Chen, X., Nadiarynkh, O., Plotnikov, S. & Campagnola, P. J. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nat Protoc* **7**, 654-669, doi:10.1038/nprot.2012.009 (2012).
441 van Helvert, S., Storm, C. & Friedl, P. Mechanoreciprocity in cell migration. *Nat Cell Biol* **20**, 8-20, doi:10.1038/s41556-017-0012-0 (2018).
442 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
443 Sureshbabu, A. *et al.* IGFBP5 induces cell adhesion, increases cell survival and inhibits cell migration in MCF-7 human breast cancer cells. *J Cell Sci* **125**, 1693-1705, doi:10.1242/jcs.092882 (2012).
444 Fridell, Y. W., Villa, J., Jr., Attar, E. C. & Liu, E. T. GAS6 induces Axl-mediated chemotaxis of vascular smooth muscle cells. *J Biol Chem* **273**, 7123-7126, doi:10.1074/jbc.273.12.7123 (1998).
445 Hung, J. Y. *et al.* Secreted protein acidic and rich in cysteine (SPARC) induces cell migration and epithelial mesenchymal transition through WNK1/snail in non-small cell lung cancer. *Oncotarget* **8**, 63691-63702, doi:10.18632/oncotarget.19475 (2017).
446 Chlenski, A. *et al.* SPARC expression is associated with impaired tumor growth, inhibited angiogenesis and changes in the extracellular matrix. *Int J Cancer* **118**, 310-316, doi:10.1002/ijc.21357 (2006).

447 Jung, J. *et al.* Gas6 Prevents Epithelial-Mesenchymal Transition in Alveolar Epithelial Cells via Production of PGE(2), PGD(2) and Their Receptors. *Cells* **8**, doi:10.3390/cells8070643 (2019).
448 Hérard, A. L. *et al.* Fibronectin and its alpha 5 beta 1-integrin receptor are involved in the wound-repair process of airway epithelium. *Am J Physiol* **271**, L726-733, doi:10.1152/ajplung.1996.271.5.L726 (1996).
449 James, D. S. *et al.* Probing ECM remodeling in idiopathic pulmonary fibrosis via second harmonic generation microscopy analysis of macro/supramolecular collagen structure. *J Biomed Opt* **25**, 1-13, doi:10.1117/1.Jbo.25.1.014505 (2019).
450 Tjin, G. *et al.* Lysyl oxidases regulate fibrillar collagen remodelling in idiopathic pulmonary fibrosis. *Dis Model Mech* **10**, 1301-1312, doi:10.1242/dmm.030114 (2017).
451 Tilbury, K. *et al.* Second harmonic generation microscopy analysis of extracellular matrix changes in human idiopathic pulmonary fibrosis. *J Biomed Opt* **19**, 086014, doi:10.1117/1.Jbo.19.8.086014 (2014).
452 Saldiva, P. H., Delmonte, V. C., de Carvalho, C. R., Kairalla, R. A. & Auler Júnior, J. O. Histochemical evaluation of lung collagen content in acute and chronic interstitial diseases. *Chest* **95**, 953-957, doi:10.1378/chest.95.5.953 (1989).
453 Lacomb, R., Nadiarnykh, O., Townsend, S. S. & Campagnola, P. J. Phase Matching considerations in Second Harmonic Generation from tissues: Effects on emission directionality, conversion efficiency and observed morphology. *Opt Commun* **281**, 1823-1832, doi:10.1016/j.optcom.2007.10.040 (2008).
454 Kottmann, R. M. *et al.* Second harmonic generation microscopy reveals altered collagen microstructure in usual interstitial pneumonia versus healthy lung. *Respiratory research* **16**, 61, doi:10.1186/s12931-015-0220-8 (2015).
455 Mostaço-Guidolin, L. B. *et al.* Defective Fibrillar Collagen Organization by Fibroblasts Contributes to Airway Remodeling in Asthma. *American journal of respiratory and critical care medicine* **200**, 431-443, doi:10.1164/rccm.201810-1855OC (2019).
456 Saglani, S. *et al.* Ultrastructure of the reticular basement membrane in asthmatic adults, children and infants. *The European respiratory journal* **28**, 505-512, doi:10.1183/09031936.06.00056405 (2006).
457 Jones, M. G. *et al.* Nanoscale dysregulation of collagen structure-function disrupts mechano-homeostasis and mediates pulmonary fibrosis. *Elife* **7**, doi:10.7554/eLife.36354 (2018).
458 Kuhn, C. & Mason, R. J. Immunolocalization of SPARC, tenascin, and thrombospondin in pulmonary fibrosis. *The American journal of pathology* **147**, 1759-1769 (1995).
459 Aufderheide, E. & Ekblom, P. Tenascin during gut development: appearance in the mesenchyme, shift in molecular forms, and dependence on epithelial-mesenchymal interactions. *J Cell Biol* **107**, 2341-2349, doi:10.1083/jcb.107.6.2341 (1988).
460 Tisler, M. *et al.* Analysis of fibroblast migration dynamics in idiopathic pulmonary fibrosis using image-based scaffolds of the lung extracellular matrix. *Am J Physiol Lung Cell Mol Physiol* **318**, L276-l286, doi:10.1152/ajplung.00087.2019 (2020).
461 Rickard, K. A., Taylor, J., Spurzem, J. R. & Rennard, S. I. Extracellular matrix and bronchial epithelial cell migration. *Chest* **101**, 17s-18s, doi:10.1378/chest.101.3_supplement.17s (1992).
462 Zahm, J. M., Chevillard, M. & Puchelle, E. Wound repair of human surface respiratory epithelium. *Am J Respir Cell Mol Biol* **5**, 242-248, doi:10.1165/ajrcmb/5.3.242 (1991).

463 Pilewski, J. M., Latoche, J. D., Arcasoy, S. M. & Albelda, S. M. Expression of integrin cell adhesion receptors during human airway epithelial repair in vivo. *Am J Physiol* **273**, L256-263, doi:10.1152/ajplung.1997.273.1.L256 (1997).
464 White, S. R., Dorscheid, D. R., Rabe, K. F., Wojcik, K. R. & Hamann, K. J. Role of very late adhesion integrins in mediating repair of human airway epithelial cell monolayers after mechanical injury. *Am J Respir Cell Mol Biol* **20**, 787-796, doi:10.1165/ajrcmb.20.4.3318 (1999).
465 Gokey, J. J. *et al.* MEG3 is increased in idiopathic pulmonary fibrosis and regulates epithelial cell differentiation. *JCI insight* **3**, doi:10.1172/jci.insight.122490 (2018).
466 Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J Cell Sci* **123**, 4195-4200, doi:10.1242/jcs.023820 (2010).
467 Kaukonen, R. *et al.* Normal stroma suppresses cancer cell proliferation via mechanosensitive regulation of JMJD1a-mediated transcription. *Nature communications* **7**, 12237, doi:10.1038/ncomms12237 (2016).
468 Ahlfors, J. E. & Billiar, K. L. Biomechanical and biochemical characteristics of a human fibroblast-produced and remodeled matrix. *Biomaterials* **28**, 2183-2191, doi:10.1016/j.biomaterials.2006.12.030 (2007).
469 Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-196, doi:10.1038/nrm3758 (2014).
470 Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871-890, doi:10.1016/j.cell.2009.11.007 (2009).
471 Parsons, J. T., Horwitz, A. R. & Schwartz, M. A. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nat Rev Mol Cell Biol* **11**, 633-643, doi:10.1038/nrm2957 (2010).
472 Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. Taking cell-matrix adhesions to the third dimension. *Science* **294**, 1708-1712, doi:10.1126/science.1064829 (2001).
473 Doyle, A. D., Carvajal, N., Jin, A., Matsumoto, K. & Yamada, K. M. Local 3D matrix microenvironment regulates cell migration through spatiotemporal dynamics of contractility-dependent adhesions. *Nature communications* **6**, 8720, doi:10.1038/ncomms9720 (2015).
474 Roach, K. M. *et al.* A model of human lung fibrogenesis for the assessment of anti-fibrotic strategies in idiopathic pulmonary fibrosis. *Sci Rep* **8**, 342, doi:10.1038/s41598-017-18555-9 (2018).
475 Alsafadi, H. N. *et al.* An ex vivo model to induce early fibrosis-like changes in human precision-cut lung slices. *Am J Physiol Lung Cell Mol Physiol* **312**, L896-l902, doi:10.1152/ajplung.00084.2017 (2017).
476 Providence, K. M. *et al.* SERPINE1 (PAI-1) is deposited into keratinocyte migration "trails" and required for optimal monolayer wound repair. *Arch Dermatol Res* **300**, 303-310, doi:10.1007/s00403-008-0845-2 (2008).
477 Legrand, C. *et al.* Airway epithelial cell migration dynamics. MMP-9 role in cell-extracellular matrix remodeling. *J Cell Biol* **146**, 517-529, doi:10.1083/jcb.146.2.517 (1999).

478 Wolf, K. *et al.* Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. *Nat Cell Biol* **9**, 893-904, doi:10.1038/ncb1616 (2007).
479 Kechagia, J. Z., Ivaska, J. & Roca-Cusachs, P. Integrins as biomechanical sensors of the microenvironment. *Nat Rev Mol Cell Biol* **20**, 457-473, doi:10.1038/s41580-019-0134-2 (2019).
480 Sheppard, D. Airway epithelial integrins: why so many? *Am J Respir Cell Mol Biol* **19**, 349-351, doi:10.1165/ajrcmb.19.3.f144 (1998).
481 Teoh, C. M., Tan, S. S. & Tran, T. Integrins as Therapeutic Targets for Respiratory Diseases. *Curr Mol Med* **15**, 714-734, doi:10.2174/1566524015666150921105339 (2015).
482 Pilewski, J. M., Liu, L., Henry, A. C., Knauer, A. V. & Feghali-Bostwick, C. A. Insulin-like growth factor binding proteins 3 and 5 are overexpressed in idiopathic pulmonary fibrosis and contribute to extracellular matrix deposition. *The American journal of pathology* **166**, 399-407, doi:10.1016/s0002-9440(10)62263-8 (2005).
483 Zhang, Y. *et al.* GDF15 is an epithelial-derived biomarker of idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* **317**, L510-l521, doi:10.1152/ajplung.00062.2019 (2019).
484 Rydell-Törmänen, K. *et al.* Aberrant nonfibrotic parenchyma in idiopathic pulmonary fibrosis is correlated with decreased β-catenin inhibition and increased Wnt5a/b interaction. *Physiol Rep* **4**, doi:10.14814/phy2.12727 (2016).
485 Espindola, M. S. *et al.* Targeting of TAM Receptors Ameliorates Fibrotic Mechanisms in Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **197**, 1443-1456, doi:10.1164/rccm.201707-1519OC (2018).
486 Chang, W. *et al.* SPARC suppresses apoptosis of idiopathic pulmonary fibrosis fibroblasts through constitutive activation of beta-catenin. *J Biol Chem* **285**, 8196-8206, doi:10.1074/jbc.M109.025684 (2010).
487 Yasuoka, H., Yamaguchi, Y. & Feghali-Bostwick, C. A. The pro-fibrotic factor IGFBP-5 induces lung fibroblast and mononuclear cell migration. *Am J Respir Cell Mol Biol* **41**, 179-188, doi:10.1165/rcmb.2008-0211OC (2009).
488 Tsou, W. I. *et al.* Receptor tyrosine kinases, TYRO3, AXL, and MER, demonstrate distinct patterns and complex regulation of ligand-induced activation. *J Biol Chem* **289**, 25750-25763, doi:10.1074/jbc.M114.569020 (2014).
489 Fujino, N. *et al.* Sensing of apoptotic cells through Axl causes lung basal cell proliferation in inflammatory diseases. *J Exp Med* **216**, 2184-2201, doi:10.1084/jem.20171978 (2019).
490 Bellan, M. *et al.* Gas6/TAM System: A Key Modulator of the Interplay between Inflammation and Fibrosis. *Int J Mol Sci* **20**, doi:10.3390/ijms20205070 (2019).
491 Fujino, N., Kubo, H. & Maciewicz, R. A. Phenotypic screening identifies Axl kinase as a negative regulator of an alveolar epithelial cell phenotype. *Lab Invest* **97**, 1047-1062, doi:10.1038/labinvest.2017.52 (2017).
492 Sottile, J. & Hocking, D. C. Fibronectin polymerization regulates the composition and stability of extracellular matrix fibrils and cell-matrix adhesions. *Mol Biol Cell* **13**, 3546-3559, doi:10.1091/mbc.e02-01-0048 (2002).

493 Torr, E. E. *et al.* Myofibroblasts exhibit enhanced fibronectin assembly that is intrinsic to their contractile phenotype. *J Biol Chem* **290**, 6951-6961, doi:10.1074/jbc.M114.606186 (2015).
494 Hocking, D. C. & Chang, C. H. Fibronectin matrix polymerization regulates small airway epithelial cell migration. *Am J Physiol Lung Cell Mol Physiol* **285**, L169-179, doi:10.1152/ajplung.00371.2002 (2003).
495 Åhrman, E. *et al.* Quantitative proteomic characterization of the lung extracellular matrix in chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *J Proteomics* **189**, 23-33, doi:10.1016/j.jprot.2018.02.027 (2018).
496 Duerr, J. *et al.* Conditional deletion of Nedd4-2 in lung epithelial cells causes progressive pulmonary fibrosis in adult mice. *Nature communications* **11**, 2012, doi:10.1038/s41467-020-15743-6 (2020).
497 Tian, Y. *et al.* Quantitative proteomic characterization of lung tissue in idiopathic pulmonary fibrosis. *Clin Proteomics* **16**, 6, doi:10.1186/s12014-019-9226-4 (2019).
498 Calle, E. A. *et al.* Targeted proteomics effectively quantifies differences between native lung and detergent-decellularized lung extracellular matrices. *Acta Biomater* **46**, 91-100, doi:10.1016/j.actbio.2016.09.043 (2016).
499 Kutys, M. L., Doyle, A. D. & Yamada, K. M. Regulation of cell adhesion and migration by cell-derived matrices. *Exp Cell Res* **319**, 2434-2439, doi:10.1016/j.yexcr.2013.05.030 (2013).
500 Petrie, R. J., Gavara, N., Chadwick, R. S. & Yamada, K. M. Nonpolarized signaling reveals two distinct modes of 3D cell migration. *J Cell Biol* **197**, 439-455, doi:10.1083/jcb.201201124 (2012).
501 Chambers, R. C. & Mercer, P. F. Mechanisms of alveolar epithelial injury, repair, and fibrosis. *Ann Am Thorac Soc* **12 Suppl 1**, S16-20, doi:10.1513/AnnalsATS.201410-448MG (2015).
502 Spira, A. *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* **13**, 361-366, doi:10.1038/nm1556 (2007).
503 Bauer, R. N., Müller, L., Brighton, L. E., Duncan, K. E. & Jaspers, I. Interaction with epithelial cells modifies airway macrophage response to ozone. *Am J Respir Cell Mol Biol* **52**, 285-294, doi:10.1165/rcmb.2014-0035OC (2015).
504 Hung, L. Y. *et al.* Macrophages promote epithelial proliferation following infectious and non-infectious lung injury through a Trefoil factor 2-dependent mechanism. *Mucosal Immunol* **12**, 64-76, doi:10.1038/s41385-018-0096-2 (2019).
505 Lang, F. M., Lee, K. M., Teijaro, J. R., Becher, B. & Hamilton, J. A. GM-CSF-based treatments in COVID-19: reconciling opposing therapeutic approaches. *Nat Rev Immunol* **20**, 507-514, doi:10.1038/s41577-020-0357-7 (2020).
506 Boland, S. *et al.* Mechanisms of GM-CSF increase by diesel exhaust particles in human airway epithelial cells. *Am J Physiol Lung Cell Mol Physiol* **278**, L25-32, doi:10.1152/ajplung.2000.278.1.L25 (2000).
507 Ramachandran, P. *et al.* Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512-518, doi:10.1038/s41586-019-1631-3 (2019).
508 Conway, B. R. *et al.* Kidney Single-Cell Atlas Reveals Myeloid Heterogeneity in Progression and Regression of Kidney Disease. *J Am Soc Nephrol* **31**, 2833-2854, doi:10.1681/asn.2020060806 (2020).

509 Bain, C. C. *et al.* Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nature immunology* **15**, 929-937, doi:10.1038/ni.2967 (2014).
510 Yang, Y. *et al.* Cross-talk between the gut microbiota and monocyte-like macrophages mediates an inflammatory response to promote colitis-associated tumourigenesis. *Gut*, doi:10.1136/gutjnl-2020-320777 (2020).
511 Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619-625, doi:10.1038/s41586-020-2922-4 (2020).
512 Muus, C. *et al.* Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat Med* **27**, 546-559, doi:10.1038/s41591-020-01227-z (2021).
513 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).
514 Hynes, R. O. The extracellular matrix: not just pretty fibrils. *Science* **326**, 1216-1219, doi:10.1126/science.1176009 (2009).
515 Tomos, I. P. *et al.* Extracellular matrix remodeling in idiopathic pulmonary fibrosis. It is the 'bed' that counts and not 'the sleepers'. *Expert Rev Respir Med* **11**, 299-309, doi:10.1080/17476348.2017.1300533 (2017).
516 Charras, G. & Sahai, E. Physical influences of the extracellular environment on cell migration. *Nat Rev Mol Cell Biol* **15**, 813-824, doi:10.1038/nrm3897 (2014).
517 Qu, F. *et al.* Maturation State and Matrix Microstructure Regulate Interstitial Cell Migration in Dense Connective Tissues. *Sci Rep* **8**, 3295, doi:10.1038/s41598-018-21212-4 (2018).
518 Wang, J. C., Sonnylal, S., Arnett, F. C., De Crombrugghe, B. & Zhou, X. Attenuation of expression of extracellular matrix genes with siRNAs to Sparc and Ctgf in skin fibroblasts of CTGF transgenic mice. *Int J Immunopathol Pharmacol* **24**, 595-601, doi:10.1177/039463201102400305 (2011).
519 Butler, C. R. *et al.* Rapid Expansion of Human Epithelial Stem Cells Suitable for Airway Tissue Engineering. *American journal of respiratory and critical care medicine* **194**, 156-168, doi:10.1164/rccm.201507-1414OC (2016).

Appendix

Figure 7.1 Standard curve generated for each analyte in the LEGENDplex™ assay. (A) IL-1β (B) IFN-α2 (C) TNF-α (D) MCP-1 (E) IL-6 (F) IL-8 (G) IL-18 (H) IL-33

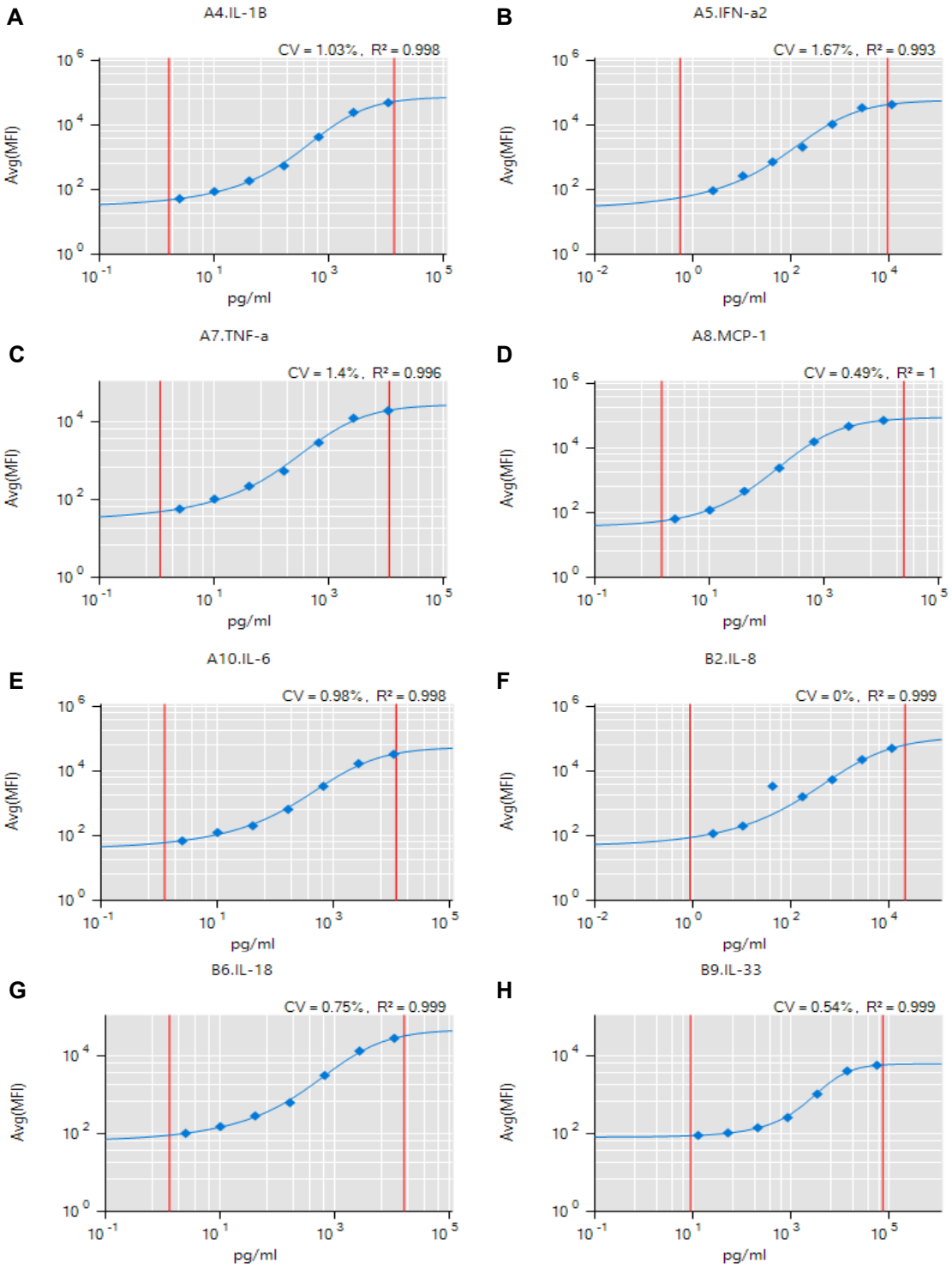


Table 7.1 Wound healing genes assayed by QIAGEN RT^2 Profiler PCR Array

Symbol	Description
ACTA2	Actin, alpha 2, smooth muscle, aorta
ACTC1	Actin, alpha, cardiac muscle 1
ANGPT1	Angiopoietin 1
CCL2	Chemokine (C-C motif) ligand 2
CCL7	Chemokine (C-C motif) ligand 7
CD40LG	CD40 ligand
CDH1	Cadherin 1, type 1, E-cadherin (epithelial)
COL14A1	Collagen, type XIV, alpha 1
COL1A1	Collagen, type I, alpha 1
COL1A2	Collagen, type I, alpha 2
COL3A1	Collagen, type III, alpha 1
COL4A1	Collagen, type IV, alpha 1
COL4A3	Collagen, type IV, alpha 3 (Goodpasture antigen)
COL5A1	Collagen, type V, alpha 1
COL5A2	Collagen, type V, alpha 2
COL5A3	Collagen, type V, alpha 3
CSF2	Colony stimulating factor 2 (granulocyte-macrophage)
CSF3	Colony stimulating factor 3 (granulocyte)
CTGF	Connective tissue growth factor
CTNNB1	Catenin (cadherin-associated protein), beta 1, 88kDa
CTSG	Cathepsin G
CTSK	Cathepsin K
CTSV	Cathepsin L2
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
CXCL11	Chemokine (C-X-C motif) ligand 11
CXCL2	Chemokine (C-X-C motif) ligand 2
CXCL5	Chemokine (C-X-C motif) ligand 5
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
F13A1	Coagulation factor XIII, A1 polypeptide
F3	Coagulation factor III (thromboplastin, tissue factor)
FGA	Fibrinogen alpha chain
FGF10	Fibroblast growth factor 10
FGF2	Fibroblast growth factor 2 (basic)
FGF7	Fibroblast growth factor 7
HBEGF	Heparin-binding EGF-like growth factor
HGF	Hepatocyte growth factor (hepapoietin A; scatter factor)
IFNG	Interferon, gamma
IGF1	Insulin-like growth factor 1 (somatomedin C)
IL10	Interleukin 10
IL1B	Interleukin 1, beta

IL2	Interleukin 2
IL4	Interleukin 4
IL6	Interleukin 6 (interferon, beta 2)
IL6ST	Interleukin 6 signal transducer (gp130, oncostatin M receptor)
ITGA1	Integrin, alpha 1
ITGA2	Integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)
ITGA3	Integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)
ITGA4	Integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)
ITGA5	Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
ITGA6	Integrin, alpha 6
ITGAV	Integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)
ITGB1	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
ITGB3	Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
ITGB5	Integrin, beta 5
ITGB6	Integrin, beta 6
MAPK1	Mitogen-activated protein kinase 1
MAPK3	Mitogen-activated protein kinase 3
MIF	Macrophage migration inhibitory factor (glycosylation-inhibiting factor)
MMP1	Matrix metallopeptidase 1 (interstitial collagenase)
MMP2	Matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)
MMP7	Matrix metallopeptidase 7 (matrilysin, uterine)
MMP9	Matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)
PDGFA	Platelet-derived growth factor alpha polypeptide
PLAT	Plasminogen activator, tissue
PLAU	Plasminogen activator, urokinase
PLAUR	Plasminogen activator, urokinase receptor
PLG	Plasminogen
PTEN	Phosphatase and tensin homolog
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
RAC1	Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
RHOA	Ras homolog gene family, member A
SERPINE1	Serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1
STAT3	Signal transducer and activator of transcription 3 (acute-phase response factor)
TAGLN	Transgelin
TGFA	Transforming growth factor, alpha
TGFB1	Transforming growth factor, beta 1
TGFBR3	Transforming growth factor, beta receptor III
TIMP1	TIMP metallopeptidase inhibitor 1
TNF	Tumor necrosis factor
VEGFA	Vascular endothelial growth factor A
VTN	Vitronectin

WISP1	WNT1 inducible signaling pathway protein 1
WNT5A	Wingless-type MMTV integration site family, member 5A

Table 7.2 Lung tissue samples used in Section 6.4.1

Sample ID	Diagnosis	Histology	Age (yrs)	Gender
NORM19	Control	NSCLC; adenocarcinoma	66	F
NORM20	Control	NSCLC; pleomorphic carcinoma	74	M
NORM21	Control	NSCLC; pleomorphic carcinoma	68	M
NORM22	Control	NSCLC; pleomorphic carcinoma	66	F
NORM23	Control	NSCLC; adenocarcinoma	64	M
IPF02	IPF	Usual interstitial pneumonia	64	F
IPF15	IPF	Usual interstitial pneumonia	63	M
IPF21	IPF	Usual interstitial pneumonia	64	M
IPF24	IPF	Usual interstitial pneumonia	64	M
IPF07	IPF	Usual interstitial pneumonia	59	M
IPF26	IPF	Usual interstitial pneumonia	57	F
IPF27	IPF	Usual interstitial pneumonia	69	M
IPF28	IPF	Usual interstitial pneumonia	77	M

Table 7.3 PBECs used in Section 6.4.3

Sample ID	Diagnosis	Age (yrs)	Gender	Smoking status
1018	Healthy control	67	M	NA
1028	Healthy control	65	M	NA
1036	Healthy control	63	M	Ex
1037	Healthy control	55	F	Never
1042	Healthy control	63	M	Never
5224	IPF	53	M	Current
5233	IPF	67	M	Current
5371	IPF	78	M	Ex
5274	IPF	76	F	Ex
5380	IPF	68	M	Ex

Table 7.4 Primary cells used in Section 6.4.4

Sample ID	Cell type	Diagnosis	Age (yrs)	Gender	Smoking status
1055	PBEC KRT5+ basal	Healthy control	47	M	Ex
1056	PBEC KRT5+ basal	Healthy control	26	F	Ex
1057	PBEC KRT5+ basal	Healthy control	51	M	Ex
1059	PBEC KRT5+ basal	Healthy control	50	F	Never
5741	PBEC KRT5+ basal	IPF	80	F	Never
5756	PBEC KRT5+ basal	IPF	73	M	Never
5759	PBEC KRT5+ basal	IPF	80	M	Never
5767	PBEC KRT5+ basal	IPF	72	M	Ex
ER35	HLF	IPF	67	F	NA

Table 7.5 PBECs used in Section 6.4.5

Sample ID	Cell type	Diagnosis	Age (yrs)	Gender	Smoking status
1054	PBEC KRT5+ basal	Healthy control	67	M	Ex
1055	PBEC KRT5+ basal	Healthy control	47	M	Ex
1056	PBEC KRT5+ basal	Healthy control	26	F	Ex
5737	PBEC KRT5+ basal	IPF	52	M	Ex
5759	PBEC KRT5+ basal	IPF	80	M	Never
5767	PBEC KRT5+ basal	IPF	72	M	Ex

Table 7.6 Primary cells used in Section 6.4.6

Sample ID	Cell type	Diagnosis	Age (yrs)	Gender	Smoking status
ER35	HLF	IPF	67	F	NA
ER55	HLF	IPF	62	F	NA
5190	HLF	IPF	51	M	Never
5340	HLF	Normal tissue, carcinoid	65	F	Never
5139	HLF	Normal tissue, adenocarcinoma	64	M	Never
5135	HLF	Normal tissue, adenocarcinoma	66	M	Ex
1054	PBEC KRT5+ basal	Healthy	67	M	Ex

Conference abstracts & presentations arising from this thesis

1) Poster presentation at "Niche-Epithelial Stem Cell Interactions in Lung Health and Disease", Lisbon, Portugal 2019

AIRWAY BASAL CELL MIGRATION INFLUENCED BY EXTRACELLULAR MATRIX CUES IN FIBROTIC LUNG DISEASE

RJ Hewitt[1], F Puttur[1], D Gaboriau[2], P Molyneaux[1,4], F Fercoq[3], L Carlin[3], AJ Byrne[1], TM Maher[1,4], CM Lloyd[1]

[1]Inflammation, Repair and Development Section, National Heart & Lung Institute, Imperial College London, UK; [2]Facility for Imaging by Light Microscopy, Imperial College London, UK; [3]Leukocyte Dynamics, Beatson Institute, Glasgow, UK; [4]Interstitial Lung Disease Unit, Royal Brompton Hospital, London, UK

Airway basal cells (BCs) expressing cytokeratin 5 (KRT5) have been identified in fibrotic distal lung in idiopathic pulmonary fibrosis (IPF), a devastating scarring lung disease. Migration of KRT5+ airway progenitor cells into injured alveolar regions was observed in an influenza mouse model. Whether human airway KRT5+ BCs migrate distally into areas of fibrotic lung in IPF is unknown. The extracellular matrix (ECM) is significantly altered in IPF so we sought to understand how this influences KRT5+ BC migration.

To confirm the presence of KRT5+ BCs in distal regions of fibrotic lung, immunofluorescence staining of lung tissue was performed. Human airway epithelial cells from IPF patients (n=5) and healthy controls (n=5) were obtained by bronchial brushings, expanded in culture and phenotyped by flow cytometry to express BC markers KRT5 and p63. BCs were seeded on culture plates coated with collagen I, III and IV, versican and fibronectin, and imaged over 12-hours using JuLI Stage Real-Time Cell History Recorder. Cells were tracked with Nikon NIS-Elements software. Phalloidin was used to stain F-actin in fixed BCs for analysis of cell morphology.

Immunofluorescence staining of lung biopsies identified KRT5+ BCs in fibrotic parenchymal regions in IPF. Tracking KRT5+ BCs in a 2D culture model demonstrated differences in migration on collagen I and III compared to collagen IV and versican. KRT5+ BCs from IPF patients exhibited reduced track length and speed on Collagen I and III compared with controls.

Overall, we demonstrate that human BC migration is influenced by the ECM microenvironment. The migratory properties of IPF BCs differed from healthy controls on contact with collagen I and III. Changes to IPF BC migratory dynamics may influence distal lung repair and remodelling. Further studies will use matrix derived from healthy and IPF human lung fibroblasts to recapitulate the distal lung microenvironment in health and disease.

2) Poster presentation at American Thoracic Society International Meeting 2020:

A TRANSCRIPTOMIC PROFILE OF THE PROXIMAL AIRWAY EPITHELIAL-IMMUNE NICHE IN IDIOPATHIC PULMONARY FIBROSIS

RJ Hewitt[1], C Graham[2], J Perez-Lloret[2], P Ghai[1], PP Ogger[1], SV Kemp[3], PL Molyneaux[1,3], F Puttur[1], AJ Byrne[1], TM Maher[1,3], A O'Garra[1,2], CM Lloyd[1]

[1]National Heart & Lung Institute, Imperial College London, UK; [2]Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London, UK; [3]NIHR Respiratory Biomedical Research Unit, Royal Brompton Hospital, London, UK

RATIONALE:

Airway epithelial cells and resident macrophages are crucially positioned at a lung-environment interface and play a pivotal role in immune defence of the lower airspace. In idiopathic pulmonary fibrosis (IPF), a progressive scarring lung disease, alveolar epithelial injury drives dysregulated fibroblast activity with excessive extracellular matrix deposition and fibrotic remodeling of the interstitium. Transcriptomic analysis of lung tissue samples has provided insight into cell populations driving disease in the distal lung, but the role of proximal airway epithelial – macrophage interactions in IPF is unknown. We used bulk RNA-sequencing and single-cell RNA-seq (scRNA-seq) to investigate the proximal airway epithelial niche in patients with IPF and healthy controls.

METHODS:

To study shared pathways relevant to disease pathogenesis, paired proximal airway brushings and bronchoalveolar lavage (BAL) were taken from IPF patients and healthy controls. RNA was extracted from brushes and airway macrophages (AMs) sorted from BAL by fluorescence activated (FACS) or magnetic-activated (MACS) cell sorting using CD206, an airway-resident macrophage marker (Control n=10, IPF n=8). High-quality RNA was used to prepare cDNA libraries using an Illumina TruSeq Stranded mRNA kit and sequenced on an Illumina HiSeq4000 sequencer. The Strand NGS RNA-seq workflow was used for downstream analysis. For scRNA-seq, a single-cell suspension was generated from bronchial brushes by collagenase D/ DNase I treatment. Viability was assessed using an EVE automatic cell counter. A 10X Chromium Chip G was loaded with a total of 10,000 cells and a 3' mRNA-Seq kit (v3) used to prepare libraries. Gene expression was analysed using the 10X Cell Ranger pipeline.

RESULTS:

RNA-seq of purified CD206 expressing AMs revealed co-expression of genes associated with a mature macrophage phenotype (*MARCO, PPARG*), but also of monocytes (*FCN1, IL1R2).* Differential expression analysis using DESeq2 revealed

genes related to cell adhesion (*ITGB3*), motility (*HMMR)*, cell- matrix interactions (*IBSP, SDC2, DDR1*), matrix remodelling (*MMP2, MMP7, MMP10, TIMP3*) and collagen production (*COL22A1, COL23A1, COL5A3*) were significantly upregulated in IPF AMs. In proximal airway brushings, *CAPN14, ADAM9* and *SERPINB8,* known to modulate integrin-cytoskeleton interactions, cell-matrix interactions and cell-cell adhesion respectively, were over-abundant in IPF compared to controls.

CONCLUSIONS:

CD206 expressing airway macrophages in IPF co-express gene markers of recruited monocytes and mature macrophages. They herald a distinct transcriptomic profile reflecting cell migration, direct interactions with the extracellular matrix and collagen production. Additionally, genes regulating epithelial cell barrier integrity and extracellular matrix organisation were upregulated in the epithelial brush samples obtained from the proximal airway.

Am J Respir Crit Care Med 2020; 201:A2243