The UIP honeycomb airway cells are the site of mucin biogenesis with deranged cilia

Jeremy A. Herrera[1,4†], Lewis A. Dingle[3,4], M. Angeles Montero[5], Rajamiyer V. Venkateswaran[4,5], John F. Blaikley[4,5], Felice Granato[5], Stella Pearson[1,2,4], Craig Lawless[1,4], David J. Thornton[1,2,4]

[1] The Wellcome Centre for Cell-Matrix Research, [2] Lydia Becker Institute of Immunology and Inflammation, [3] Blond McIndoe Laboratories, [4] Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, Great Manchester, United Kingdom.

[5] Manchester University NHS Foundation Trust, Manchester, Greater Manchester, United Kingdom.

†Corresponding author: Jeremy.Herrera@manchester.ac.uk , (+44) 0161 275 5072, University of Manchester, 4.019 AV Hill, Oxford Road, Manchester, Greater Manchester, M13 9PT, United Kingdom (orcid.org/0000-0003-4845-8494)

Conflict of Interests: The authors have declared that no conflict of interests exists.

Abstract:

Honeycombing (HC) is a histological pattern consistent with Usual Interstitial Pneumonia (UIP). HC refers to cystic airways (HC airways) located at sites of dense fibrosis with marked mucus accumulation. Utilizing laser capture microdissection coupled mass spectrometry (LCM-MS), we interrogated the fibrotic HC airway cells and fibrotic uninvolved airway cells (distant from sites of UIP and morphologically intact) in 10 UIP specimens; 6 non-fibrotic airway cell specimens served as controls. Furthermore, we performed LCM-MS on the mucus plugs found in 6 UIP and 6 mucinous adenocarcinoma (MA) specimens. The mass spectrometry data were subject to both qualitative and quantitative analysis and validated by immunohistochemistry. Surprisingly, fibrotic uninvolved airway cells share a similar protein profile to HC airway cells, showing deregulation of SLITs and ROBO pathway as the strongest category. We find that BPIFB1 is the most significantly increased secretome-associated protein in UIP, whereas MUC5AC is the most significantly increased in MA. We conclude that spatial proteomics demonstrates that the fibrotic uninvolved airway cells are abnormal. In addition, fibrotic HC airway cells are enriched in mucin biogenesis proteins with a marked derangement in proteins essential for ciliogenesis. This unbiased spatial proteomic approach will generate novel and testable hypotheses to decipher fibrosis progression.

Introduction:

Usual Interstitial Pneumonia (UIP) is a fibrotic disease that is associated with a variety of fibrotic entities (Idiopathic Pulmonary Fibrosis – IPF, sarcoidosis, non-specific interstitial pneumonia - NSIP, connective tissue disease - CTD, and hypersensitivity pneumonitis - HP) [1-3]. The UIP histological pattern is patchy with regions of relatively normal-appearing lung adjacent to dense fibrosis and honeycombing (HC). HC refers to the clustering of airspaces within dense fibrotic tissue and is associated with the thickening of airway walls. Accumulation and plugging of mucus and other airway debris within the HC airways likely negatively impacts lung function.

Our current understanding of fibrotic airway pathogenesis has been improved with the advancement of structural, genetic, and molecular analyses. Structurally, UIP/IPF lung have reduced numbers of terminal bronchioles in both regions of minimal and established fibrosis [4-6]. Genetically, sequence changes affecting alveolar cells (*MUC5B*, *SFTPC*, and *SFTPA2*) have been reported [7-9]. Cells comprising the HC airways present as either multi-layer or as a single layer [1], with cellular subtypes including; basal, ciliated, columnar, pseudostratified and secretory epithelium; while there are variable reports on the presence of alveolar type II (ATII) cells [10-12]. Functionally, single-cell RNA sequencing of IPF epithelial cells identify marked cellular heterogeneity as compared to control [13]. Collectively, these factors are believed to lead to airway homeostasis impairment and facilitate disease progression.

An important function of the airway is to produce mucus. Not only does mucus serve as a physical barrier but mucus has antimicrobial properties to protect distant airways [14]. During the fibrotic process, mucus fills and plugs the HC airways, which affects pathogen clearance and blood-oxygen exchange. The secreted mucus hydrogel is underpinned by two gel-forming mucins, of which, mucin 5B (MUC5B) is the most abundant in health, whereas MUC5AC is also detected but at a lower level [10]. A gain-of-function *MUC5B* polymorphism is amongst one of the highest risk factors associated with lung fibrosis [10, 15-18]. However, knowledge on the molecular composition of the mucus plug in UIP is incomplete.

We have recently created a tissue atlas of the fibrotic front of UIP/IPF utilizing laser capture microdissection coupled mass spectrometry (LCM-MS) [19, 20]. In this study, we used the same approach to define the composition and provide mechanistic themes of the fibrotic HC airway cells with the aim to identify key targets and pathways to intercept fibrosis progression. In addition, we identify the composition of mucus in HC airways in lung fibrosis (UIP) and compare this to lung cancer (mucinous adenocarcinoma [MA]) to determine if mucus heterogeneity exists in varying disease states where mucus plugs are found in the airways.

Results:

Laser capture microdissection of fibrotic and non-fibrotic airway cells

Figure 1 shows our approach to laser capture microdissection (LCM). Using alcian blue/ periodic acid Schiff's (AB/PAS) stain, mucus is visualized as purple in color within the fibrotic HC airway (**Figure 1A, upper row**); note how the AB/PAS stain lines the airway cells (red arrows) in a manner that suggests mucin is being secreted centrally into the airway lumen. We show that we precisely captured the mucus in a fibrotic specimen, including its cellular infiltrates (**Figure 1A, middle row**). In addition, we captured the mucin-rich epithelial lining of HC airways (**Figure 1A, lowest row**). We also captured fibrotic uninvolved airway cells defined as being in distant regions demonstrating minimal fibrosis (**Figure 1B**). Our LCM capabilities allow us to precisely isolate this region while leaving behind mucus associated in uninvolved airways, denoted with a red asterisk. To serve as a control, we performed LCM on airway cells from non-fibrotic specimens (**Figure 1C**). In total, we performed LCM on 10 fibrotic specimens (n = 10 fibrotic HC airway cells and n = 10 fibrotic uninvolved airway cells) and on 6 non-fibrotic airway cell controls (a total of 26 samples).

The fibrotic HC and uninvolved airway cells are similar in protein composition

We prepared our samples for mass spectrometry (MS) following our established protocol [19, 20] and performed a qualitative analysis to determine which proteins are present per group: non-fibrotic, fibrotic uninvolved, and fibrotic HC airway cells. We define a protein present if it is detected in 3 of the 6 non-fibrotic airway cell samples or 5 of the 10 fibrotic airway cell samples. We detected 2,668 proteins in human lung airway cells (**Figure 2A**) and provide a complete list of these proteins (**Supplemental File 1**). We found that more

109

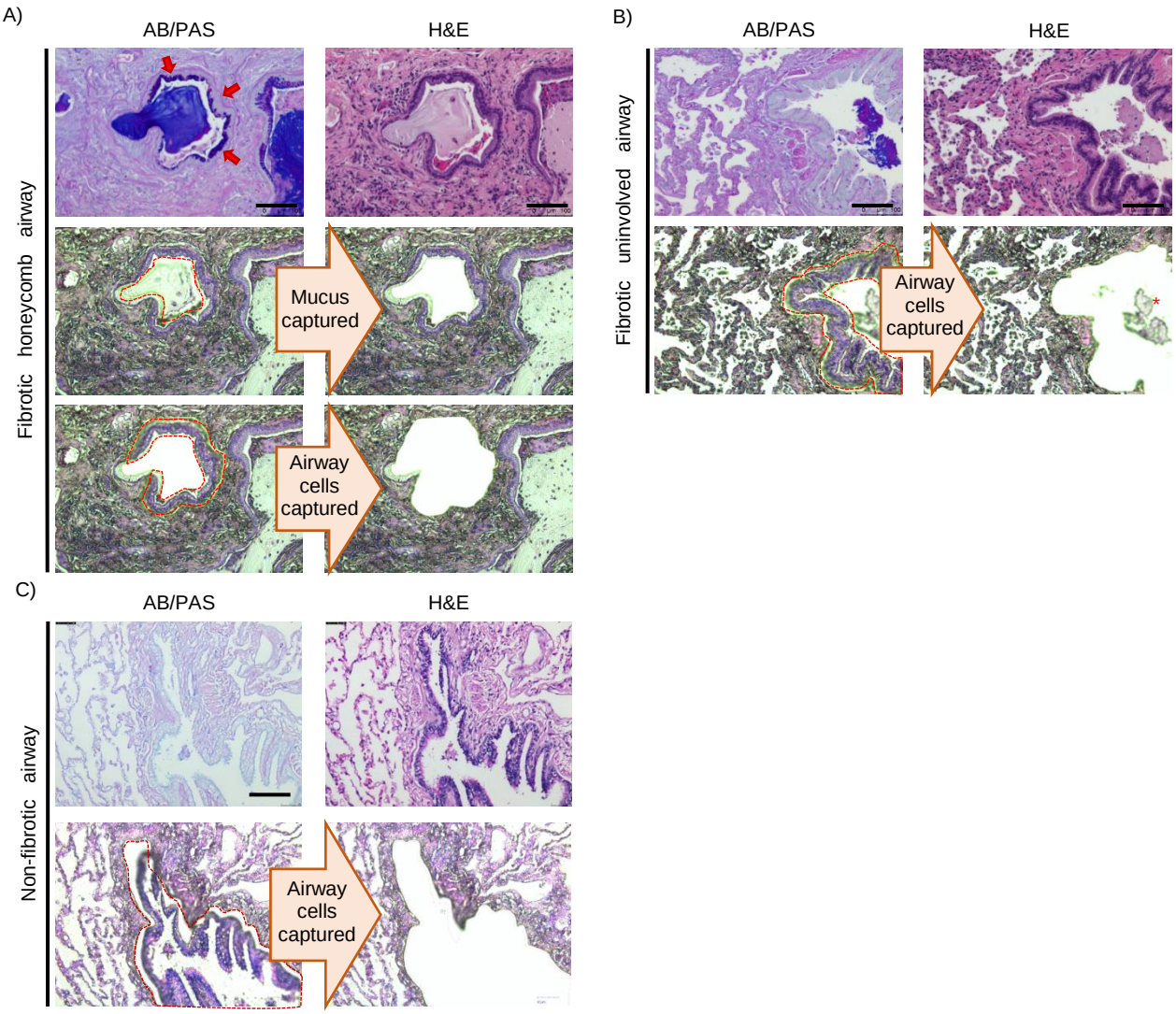


Figure 1: Laser capture microdissection of the mucus, fibrotic honeycomb, fibrotic uninvolved and non-fibrotic airway cell controls. Formalin-fixed paraffin-embedded specimens were serially sectioned at 5 microns and stained with alcian blue/periodic acid Schiff's (AB/PAS) stain or Hemetoxylin & Eosin (H&E). (**A**) A representation of laser capture microdissection in a fibrotic specimen. AB/PAS (Top left) or H&E (the other 5 panels). Mucus (purple) was visualized with AB/PAS stain; notice how the mucin lines the inner airway consistent with these cells producing mucin centrally into the airspace [red arrows]. We individually captured the mucus and fibrotic honeycomb airway cells for mass spectrometry preparation. (**B**) In the same fibrotic patient, we found uninvolved airways in the morphologically intact regions of the fibrotic lung and captured the airway cells for mass spectrometry preparation (notice how the mucus plug is left behind– depicted with red asterisk). (**C**) A representation of non-fibrotic airway cells captured for mass spectrometry preparation. Scale bar represents 100 microns.

proteins are detected in fibrotic HC airway cells, which may be attributed to the metabolically demanding process of mucin production [21].

We next performed a quantitative analysis, which compares the relative abundance of the detected proteins, to create a 3-dimensional (3-D) principal component analysis (PCA) (**Figure 2B**). Firstly, we showed that non-fibrotic airway cells (red dots) separate from both the fibrotic HC (green dots) and fibrotic uninvolved airway cells (yellow dots). Surprisingly, we found that both the fibrotic HC airway cells and fibrotic uninvolved airway cells closely cluster with some deviation. This analysis suggests that the fibrotic uninvolved airway cells (found in morphologically intact lung) display an abnormal protein profile as the mucin-rich HC airway cells (found within the densely fibrotic region of the lung).

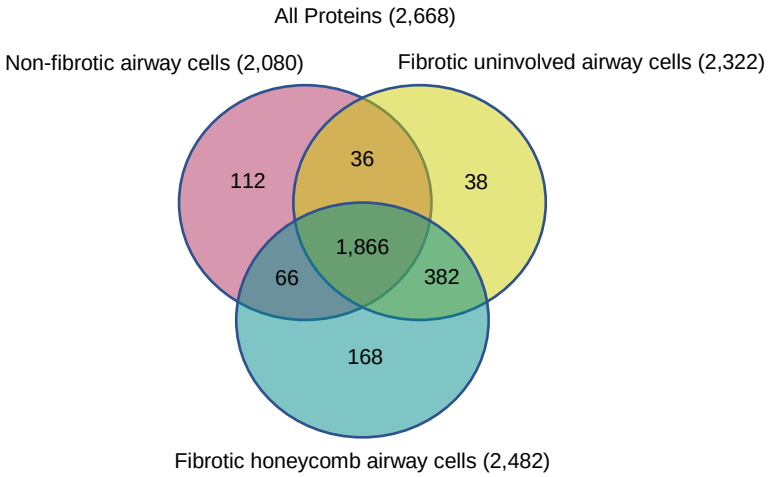
Honeycomb airway cells are enriched in proteins involved in mucus biogenesis and have decreased cilia-associated proteins

Although the fibrotic HC and fibrotic uninvolved airway cells cluster by PCA, we sought to further compare these groups. Of the 2,957 proteins detected, we found that there are 101 proteins significantly increased in fibrotic HC airway cells while 18 are statistically increased in the fibrotic uninvolved airway cells (**Figure 3A,** a full list in **Supplemental File 2**). A list of the highest and lowest proteins is provided in **Table 1**. Consistent with our approach of capturing mucin-rich HC airway cells, we found that MUC5B is

132

133

A)



B)

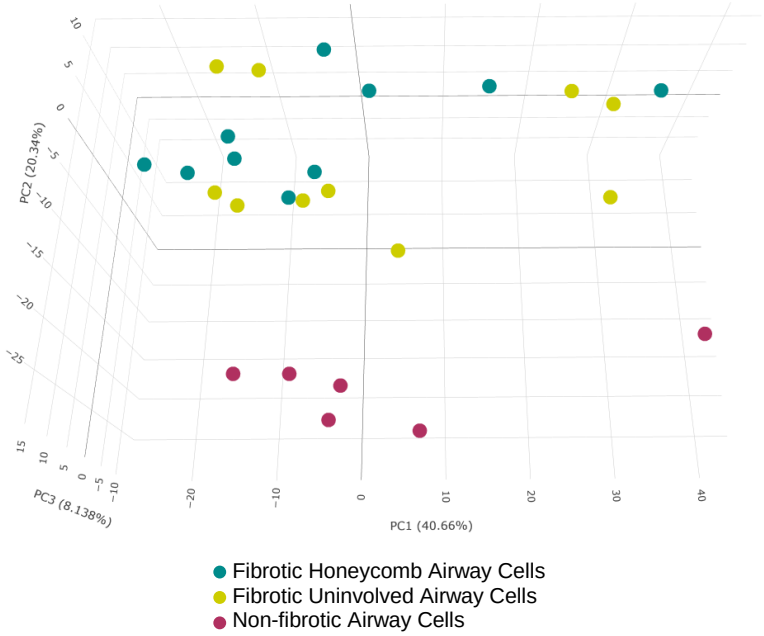


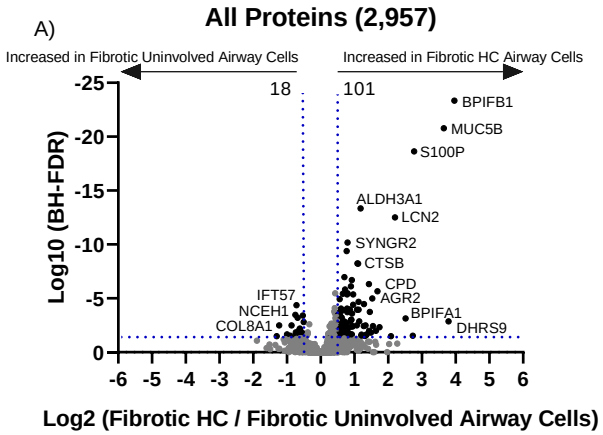
Figure 2. Spatial proteomic analysis of the fibrotic airway cells. Fibrotic and non-fibrotic specimens were subjected to laser capture microdissection coupled mass spectrometry (LCM-MS) to collect non-fibrotic airway cells (n = 6 specimens), fibrotic uninvolved airway cells (n = 10 specimens), and fibrotic honeycomb (HC) airway cells (n = 10 specimens). (**A**) Venn diagram showing the number of proteins found in each airway cell type. (**B**) 3-D Principal component analysis showing that the non-fibrotic airway cells (red dots) cluster away from the fibrotic airway cells (the other dots). Surprisingly, fibrotic uninvolved airway cells (yellow dots) and HC airway cells (green dots) cluster together.

significantly increased in the fibrotic HC airway cells and not in the fibrotic uninvolved airway cells.

Strikingly, many of the proteins increased in the fibrotic HC airway cells are involved in mucin biogenesis and/or regulation. For instance, bactericidal/permeability-increasing (BPI) fold-containing family B member 1 (*BPIFB1*) is a negative regulator of *MUC5B* expression [22] and is at the top of the list. Similarly, secretory leukocyte protease inhibitor (*SLPI*) reduces mucin expression *in vitro* and is enriched in the fibrotic HC airway cells [23]. This suggests that negative regulators of mucin expression in lung fibrosis are insufficient to stop mucin production. Retinoic acid signalling induces mucin gene expression and secretion [24, 25]; Dehydrogenase reductase SDR family member 9 (*DHRS9*) and cellular retinoic acid-binding protein 2 (*CRABP2*) both modulate retinoic acid synthesis and are increased in the fibrotic HC airway cells. Recently, CRABP2 was shown to be increased in IPF airway cells [26]. In accord with mucin biogenesis, anterior gradient homolog 2 (*AGR2*) has been shown to be essential for MUC2 production and FK506-binding protein 11 (*FKBP11*) has been demonstrated to have a mucin secretory function [27]. Both AGR2 and FKBP11 are increased in the fibrotic HC airway cells. Some unique proteins to the fibrotic HC airway cells are GALNT12, GALNT3, ST6GAL1, and GALNT6, which are associated with O-linked glycosylation of mucins. In accord with these findings, Reactome pathway analysis demonstrated that a variety of pathways pertaining to mucin production are increased, such as ‘Post-translational protein modification’, ‘Transport to the Golgi and subsequent modification’, and ‘ER-phagosome pathway’ (**Figure 3B**); unfortunately, mucin biogenesis is not an established Reactome

157

A)



B) Increased in Fibrotic HC Airway Cells

Reactome ID/Pathway	FDR	Gene Count
R-HSA-168249/ Innate Immune System	1.37E-05	51
R-HSA-597592/ Post-translational protein modification	1.29E-04	45
R-HSA-6798695/ Neutrophil degranulation	1.29E-04	36
R-HSA-446203/ Asparagine N-linked glycosylation	1.29E-04	23
R-HSA-948021/ Transport to the Golgi and subsequent modification	6.96E-03	15
R-HSA-1236974/ ER-Phagosome pathway	6.96E-03	12
R-HSA-199977/ ER to Golgi Anterograde Transport	6.96E-03	14
R-HSA-1236975/ Antigen processing-Cross presentation	6.96E-03	12

C) Decreased in Fibrotic HC Airway Cells

Reactome ID/Pathway	FDR	Gene Count
R-HSA-6790901/ rRNA modification in the nucleus and cytosol	3.46E-03	3
R-HSA-1474244/ Extracellular matrix organization	3.46E-03	6
R-HSA-1566948/ Elastic fibre formation	3.46E-03	3
R-HSA-2129379/ Molecules associated with elastic fibres	3.46E-03	3

Figure 3: The fibrotic honeycomb airway cells have a pro-mucin protein signature. (**A**) Volcano plot comparing the fibrotic honeycomb (HC) airway to fibrotic uninvolved airway cells showing the negative natural log of the false discovery values (FDR) values plotted against the base 2 log (fold change) for each protein. Reactome pathway showing the most (**B**) increased or (**C**) decreased for the fibrotic HC airway cells compared to fibrotic uninvolved airway cells.

Table 1: Highest and lowest 15 significantly changed proteins in the fibrotic HC airway cells compared to fibrotic uninvolved airway cells.

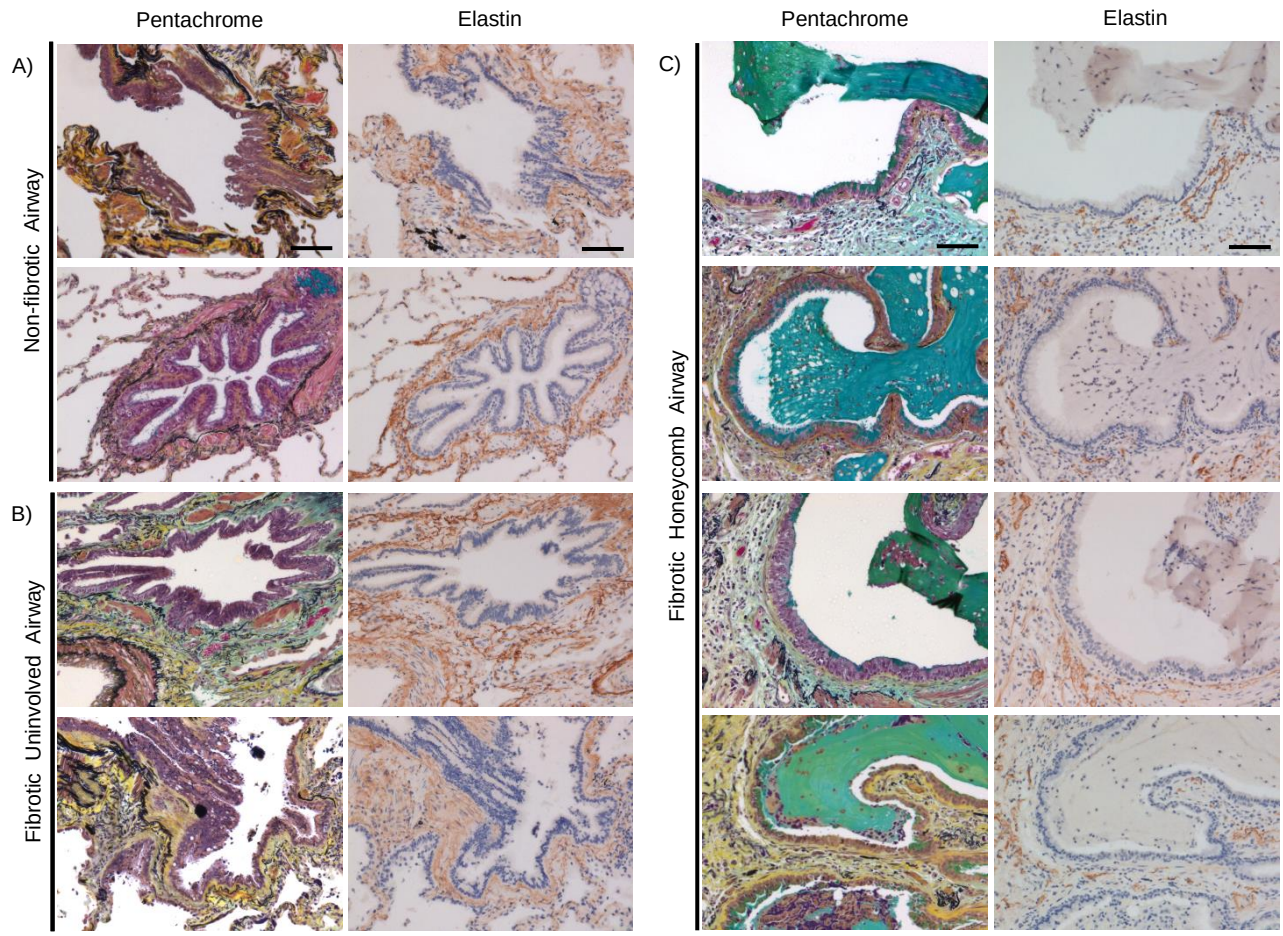
***Increased* in fibrotic HC airway cells**			***Decreased* in fibrotic HC airway cells**		
Protein	***Log_2***	***FDR***	***Protein***	***Log_2***	***FDR***
BPIFB1	3.97	4.60E-24	CEP135	-1.30	3.18E-02
DHRS9	3.79	1.40E-03	COL8A1	-1.23	3.28E-03
MUC5B	3.65	1.70E-21	FBLN2	-1.00	2.16E-02
S100P	2.77	2.40E-19	LRRC45	-0.90	2.88E-02
FAM3D	2.73	2.90E-02	TTC21B	-0.86	3.29E-03
BPIFA1	2.52	7.40E-04	LAMB2	-0.75	3.40E-04
LCN2	2.20	3.13E-13	APCS	-0.74	4.45E-02
CRABP2	2.08	3.22E-02	CGN	-0.74	1.45E-02
DMBT1	1.74	4.82E-03	FBN1	-0.72	2.24E-02
CPD	1.68	2.30E-06	IFT57	-0.72	4.30E-05
ST6GAL1	1.64	1.02E-02	NCEH1	-0.69	6.30E-04
IGJ	1.55	3.763E-03	PLG	-0.63	8.00E-03
AGR2	1.53	1.00E-05	H1F0	-0.62	6.25E-03
FKBP11	1.49	1.53E-02	CYP51A1	-0.60	2.90E-02
SLPI	1.46	1.80E-04	CERS2	-0.59	2.59E-02

158

159

pathway. Reactome pathway analysis also demonstrated that 'extracellular matrix organization' and 'elastic fibre formation' are decreased in the fibrotic HC airway cells [**Figure 3C**]. We confirm that there are disorganized elastic fibres in the HC airways, which is in accord with the loss of airway structure and increased fibrosis in this region (**Supplemental Figure 1**). Thus, our spatial proteomic approach identifies a pro-mucin protein signature associated with the fibrotic HC airway cells as compared to fibrotic uninvolved airway cells.

Cilia are conserved organelles that function to clear airway mucus and associated debris. Herein, we demonstrated that multiple proteins associated with ciliogenesis are decreased in the fibrotic HC airway cells. For instance, Centrosomal protein 135 (CEP135) is the most decreased protein and is required for ciliogenesis initiation [28]. Ciliogenesis relies on a variety of proteins, including intraflagellar transport (IFT) proteins [29]. Intraflagellar transport protein 57 (IFT57) is required for cilia maintenance and is decreased in the fibrotic HC airway cells; other intraflagellar transport proteins (IFT81, IFT46) are not expressed in the fibrotic HC airway cells but expressed in the non-fibrotic or fibrotic uninvolved airway cells. Leucine-rich repeat protein 21B (*LRRC45*) and tetratricopeptide repeat protein 21B (*TTC21B*) are critical for ciliogenesis and are also decreased in the fibrotic HC airway cells [30, 31]. A variety of proteins that are not expressed in the fibrotic HC airway cells include proteins associated with cilia function, such as CEP131, CCP110, KIF3A, CYB5D1, DYNLRBR2, RPGR, and WRD66 [32-38]. These data suggest that the loss of proteins associated with ciliogenesis within the fibrotic HC airways may be part of the mechanism of fibrosis progression.



Supplemental Figure 1. Elastin disorganization in the fibrotic honeycomb airway. 2 Non-fibrotic and 4 fibrotic specimens were stained for pentachrome or immunostained for elastin. Shown are representative images for (**A**) non-fibrotic airway, (**B**) fibrotic uninvolved airways, and (**C**) fibrotic honeycomb airways. Note that elastin fibres (black in color in the pentachrome) surround airways in the non-fibrotic airway and fibrotic uninvolved airways, but is disorganized in the honeycomb airways.

A previous study demonstrated that ciliary microtubules are morphologically disorganized in UIP/IPF, which will have profound effects on cilia structure and function [39]. To demonstrate that abnormal ciliogenesis is a potential mechanistic theme in the fibrotic HC airway cells, we immunostained for tubulin alpha 4a (TUBA4*A*; a marker of cilia) in 4 UIP specimens and 2 controls. We found that the cilia marker is widely expressed in cells lining the airway cells of non-fibrotic and fibrotic uninvolved airways (**Figure 4A – 4B**). In contrast, the mucin-rich regions of the HC airway (red arrows) are largely devoid of cilia (black arrows) (**Figure 4C**).

The fibrotic uninvolved airway cells have an abnormal protein signature

We next compared both the fibrotic airway cells to the non-fibrotic airway cell controls. We showed that 333 proteins are significantly increased in the fibrotic HC airway cells, whereas 157 proteins are significantly increased in the non-fibrotic airway cell controls (**Figure 5A**; a full list in **Supplemental File 2**). Reactome pathway analysis demonstrated that 'regulation of expression of *SLIT*s and *ROBO*s' and 'Signaling by *ROBO* receptors' are the strongest categories increased in the fibrotic HC airway cells as compared to non-fibrotic controls (**Figure 5B**). The Slit/Robo pathway is implicated in a variety of cellular processes, including neurogenesis, cell proliferation, and migration [40]. Slit/Robo has been shown to be involved in airway development [41, 42], and can improve alveolar regeneration in lung-injury models [43].

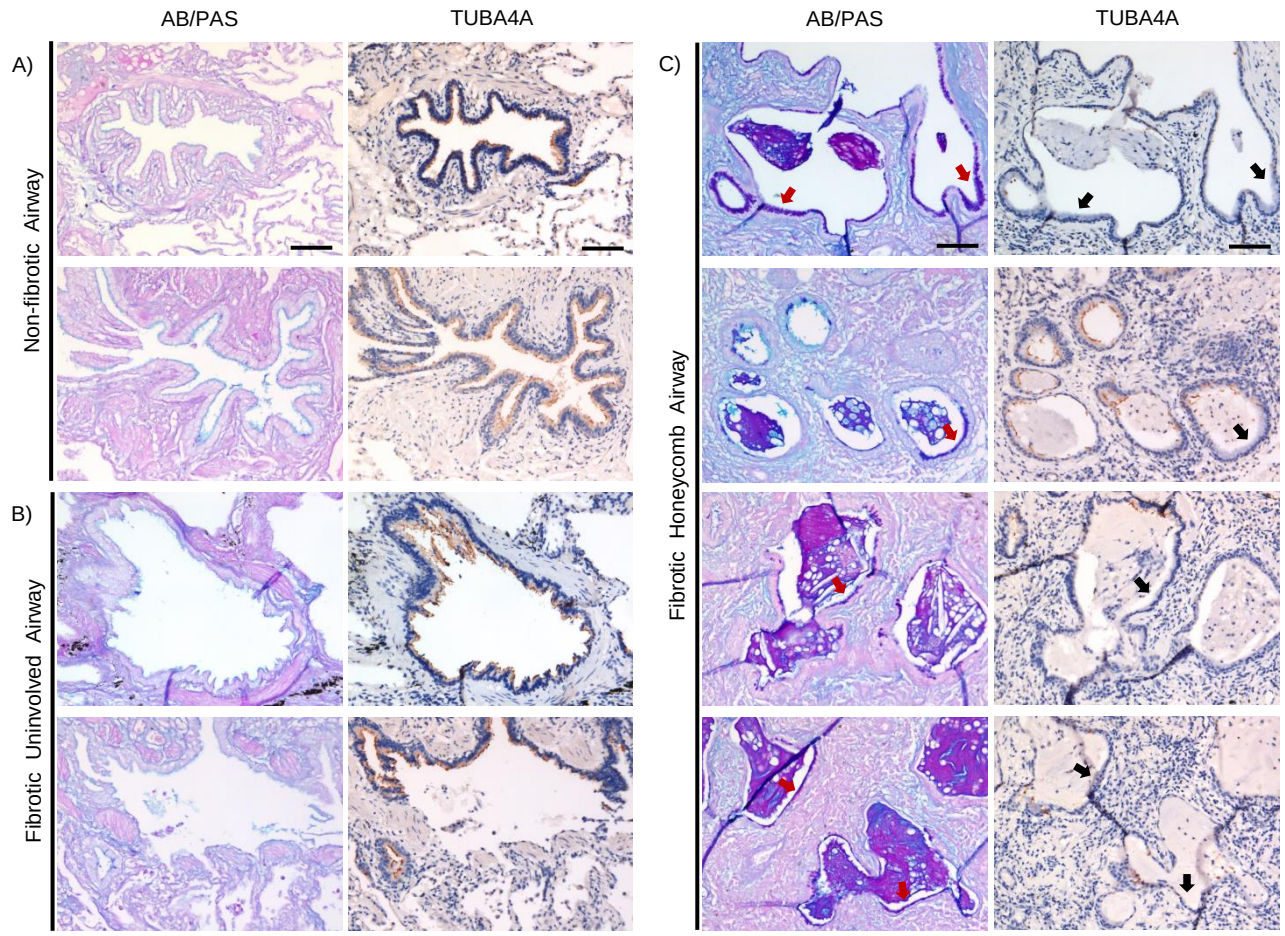
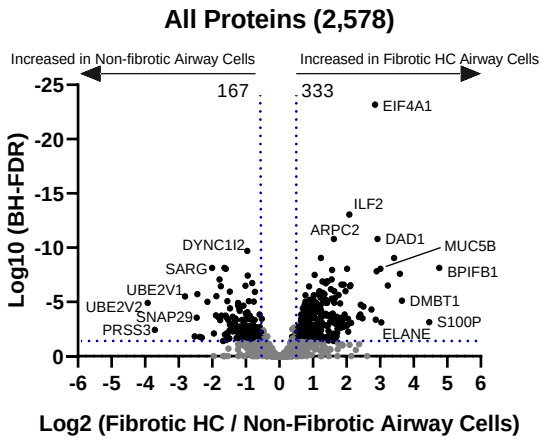


Figure 4. Cilia expression in the fibrotic honeycomb airway cells. Two Non-fibrotic and 4 fibrotic specimens were stained for alcian blue/periodic acid Schiff's (AB/PAS) or immunostained for TUBA4A (a marker of cilia). Shown are representative images for (**A**) non-fibrotic airway, (**B**) fibrotic uninvolved airways, and (**C**) fibrotic honeycomb (HC) airways. Note that regions of mucin positivity (red arrows) are absent of cilia (black arrows) in the fibrotic HC airway cells. Scale bar represents 100 microns.

209 A)

210

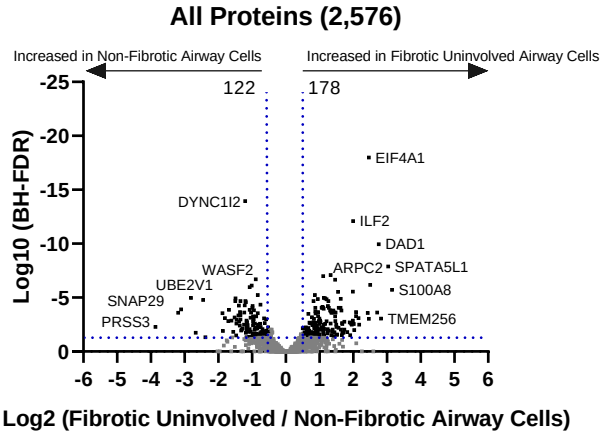


B)

Up-regulated in Fibrotic HC Airway Cells

Reactome ID/ Pathway	FDR	Gene Count
R-HSA-9010553/ Regulation of expression of SLITs and ROBOs	3.85E-05	42
R-HSA-376176/ Signaling by ROBO receptors	7.95E-05	43
R-HSA-71291/ Metabolism of amino acids and derivatives	5.54E-04	50
R-HSA-422475/ Axon guidance	5.54E-04	56
R-HSA-4086400/ PCP/CE pathway	5.54E-04	21
R-HSA-168249/ Innate Immune System	5.54E-04	84
R-HSA-9675108/ Nervous system development	5.54E-04	56
R-HSA-5663205/ Infectious disease	5.54E-04	73

C)



D)

Up-regulated in Fibrotic Uninvolved Airway Cells

Reactome ID/Pathway	FDR	Gene Count
R-HSA-9010553/ Regulation of expression of SLITs and ROBOs	2.38E-06	30
R-HSA-376176/ Signaling by ROBO receptors	1.03E-05	30
R-HSA-422475/ Axon guidance	1.49E-05	39
R-HSA-9675108/ Nervous system development	1.72E-05	39
R-HSA-1266738/ Developmental Biology	1.72E-05	43
R-HSA-156842/ Eukaryotic Translation Elongation	7.23E-05	20
R-HSA-71291/ Metabolism of amino acids and derivatives	1.86E-04	32
R-HSA-72766/ Translation	2.55E-04	28

Figure 5: The fibrotic uninvolved airway cells have an abnormal protein signature. (**A**) Volcano plot comparing the fibrotic honeycomb airway cells to non-fibrotic airway cells showing the negative natural log of the false discovery values (FDR) values plotted against the base 2 log (fold change) for each protein. (**B**) Reactome pathway showing the most increased pathway for the fibrotic honeycomb airway cells compared to non-fibrotic airway cells. (**C**) Volcano plot comparing the fibrotic uninvolved airway cells to non-fibrotic airway cells showing the negative natural log of the FDR values plotted against the base 2 log (fold change) for each protein. (**D**) Reactome pathway showing the most increased pathway for the fibrotic uninvolved airway cells compared to non-fibrotic airway cell controls.

We next compared the fibrotic uninvolved airway cells to the non-fibrotic airway cell controls. We detected 178 proteins significantly increased in fibrotic uninvolved airway cells, whereas we found 202 proteins significantly increased in non-fibrotic airway cell controls (**Figure 5C**; a full list in **Supplemental file 2**). Surprisingly, the 15 highest proteins increased in the fibrotic uninvolved airway cells are also increased in the fibrotic HC airway cells. Reactome pathway analysis again show that 'Regulation of expression of *SLIT*s and *ROBO*s' and 'Signaling by *ROBO* receptors' are the strongest categories in the fibrotic uninvolved airway cells as compared to non-fibrotic controls (**Figure 5D**); Reactome pathway analysis did not find any significantly decreased pathway enrichment in either fibrotic airway groups. Given that similar pathways and proteins are increased in the fibrotic uninvolved airway cells as the fibrotic HC airway cells, suggests that the fibrotic uninvolved airway cells are abnormal.

A heatmap of the 568 significantly changed proteins across the groups: fibrotic HC, fibrotic uninvolved, and non-fibrotic airway cell controls is shown in **Figure 6A** (a full list in **Supplemental File 3**). The fibrotic HC airway cells share similarities to the fibrotic uninvolved airway cells, but with substantial deviations. We next show the 25 highest and lowest changed proteins (**Figure 6B**). Ras-related protein 3D (RAB3D) was the most increased in the fibrotic HC airway cells and is involved in the biogenesis of secretory granules [44]. MUC5B and MUC5AC are both packaged in the secretory granules of airway cells [45]. Similarly, prolyl endopeptidase (PREP) is found within exosomes in airway cells and released upon LPS stimulation [46]. Cyclase associated protein 1 (CAP1) is associated with lung cancer and post-translational modification promotes proliferation

234

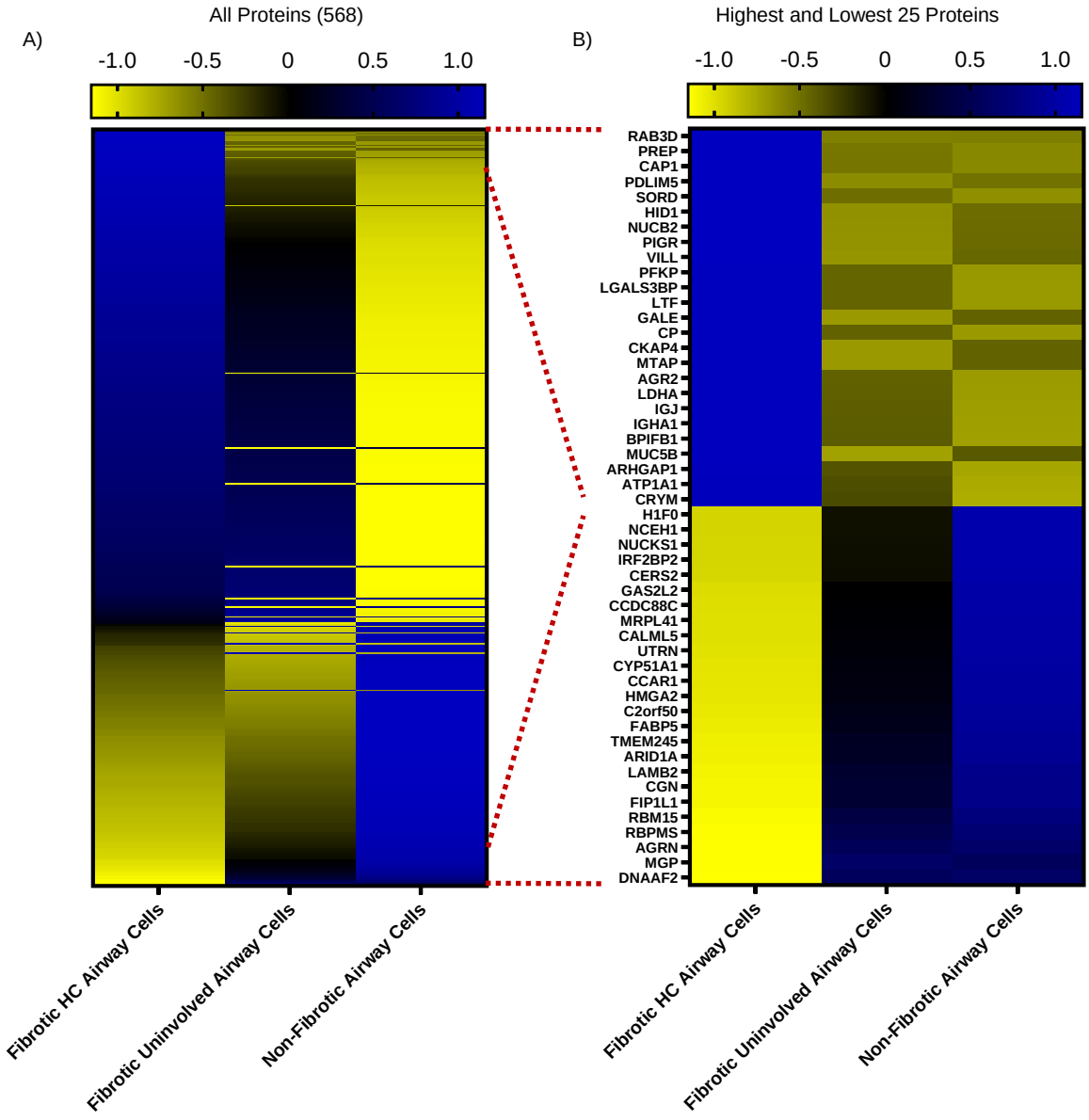


Figure 6: Fibrotic airway cells differ than controls. Shown are heatmaps of (**A**) all 568 statistically changed proteins or (**B**) the highest and lowest 25 proteins in the fibrotic honeycomb, fibrotic uninvolved, and non-fibrotic airway cells. Proteins are arranged by increasing abundance with reference to the honeycomb airway cells.

and migration [47]. At the bottom of the list are dynein axonemal assembly factor 2 (DNAAF2), matrix gla protein (MGP), and agrin (AGRN) which are only decreased in the fibrotic HC airway cells (detected in the fibrotic uninvolved and non-fibrotic airway cell controls). *DNAAF2* is involved in cilia homeostasis and mutations in *DNAAF2* lead to cilia defects [48]. *MGP* is considered an inhibitor of calcification based on the extensive cardiovascular calcification observed in *MGP*-null mice [49]; calcification occurs in UIP/IPF patients, and is associated within regions of honeycombing [50]. *AGRN* is a proteoglycan that serves a variety of biological functions, including the promotion of regeneration [51]. Further work interrogating the collective roles of these changed proteins will help decipher the mechanism of fibrosis progression.

The composition of fibrotic lung mucus

We previously demonstrated our capacity to microdissect the mucus plugs in fibrotic specimens (**Figure 1A**, middle panels). Utilizing our MS approach, we detected 650 proteins in the fibrotic/UIP mucus plugs (detected in 3 or more of the 6 samples; **Supplemental File 4**). Using intensity Based Absolute Quantification (iBAQ; a measure of protein abundance) [52], we provide a list of the most abundant proteins in UIP mucus (**Supplemental File 5**). We found that the mucus is enriched with immunoglobulins (Ig) which is in accord with increased protein expression of polymeric Ig receptor (PIGR) in the fibrotic honeycomb airway cells and mucus. In epithelial cells, PIGR mediates the transcytosis of Igs into the airway, which serves as a mucosal defence mechanism [53]. Given that fibrotic mucus is enriched with cellular infiltrates, we next focused our list using the 'secretome' (secreted proteins) dataset [54] and show that BPIFB1 was the most

abundant secretome-associated protein found in fibrotic mucus whereas MUC5B is the ninth most abundant (**Figure 7A**; a full list in **Supplemental File 6**). This is consistent with BPIFB1 and MUC5B being amongst the most significantly expressed proteins in the fibrotic HC airway cells (**Figure 3A**). To validate some the most abundant protein hits, we show immunoreactivity for MUC5B, BPIFB1, PIGR, and TF within the UIP mucus plugs (**Figure 7B**). We also included the other gel-forming mucin, MUC5AC (60th on the abundance list), which showed a patchy/incomplete staining pattern.

Using the entire fibrotic mucus proteome, Reactome pathway enrichment analysis demonstrates that the mucus plug is defined by 'neutrophil degranulation' as the strongest category (**Figure 7C**). Neutrophil degranulation pathway is also implicated in SARS-CoV-2 lung infection models [55] and in chronic obstructive pulmonary disease (COPD) [56]; in IPF bronchoalveolar lavage fluid (BALF), proteins associated with neutrophil granules are amongst the most abundant [57].

Fibrotic-derived lung mucus is distinct from cancer-derived lung mucus

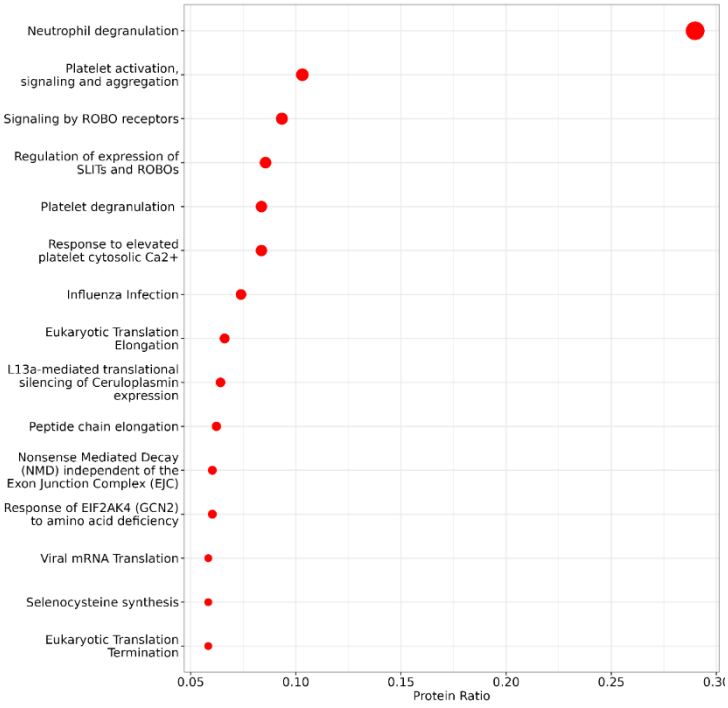
To further understand mucus in the context of lung disease, we additionally performed LCM-MS on 6 mucinous adenocarcinoma (MA) specimens (**Supplemental Figure 2A**). MA is a lung cancer with pronounced mucus accumulation within the alveolar space [58, 59]. We detected a total of 535 proteins in MA mucus (found in 3 or more of the 6 samples; a full list in **Supplemental File 4**). Consistent with *MUC5AC* being the most abundantly expressed transcript in MA [60], MUC5AC protein is the second most abundant

280

A)

The 10 most abundant secretome-associated proteins in the fibrotic mucus. (172 proteins total)	
Gene, protein name	Median iBAQ (x10⁴)
BPIFB1, BPI fold-containing family B member 1	9920
DEFA3/1, Neutrophil defensin 3/1	9324
PIGR, Polymeric immunoglobulin receptor	7718
SFTPB, pulmonary surfactant-associated protein B	6418
SFTPA1/2, Pulmonary surfactant-associated protein A1/2	3667
S100A9, Protein S100-A9	3361
DCD, Dermcidin	3111
APOA1, Apolipoprotein A-I	2833
MUC5B, Mucin-5B	2771
TF, Serotransferrin	2586

C)



B)

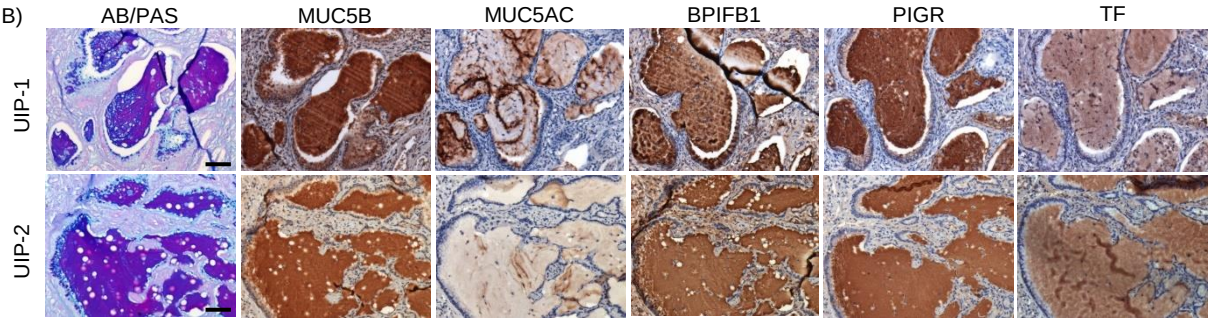
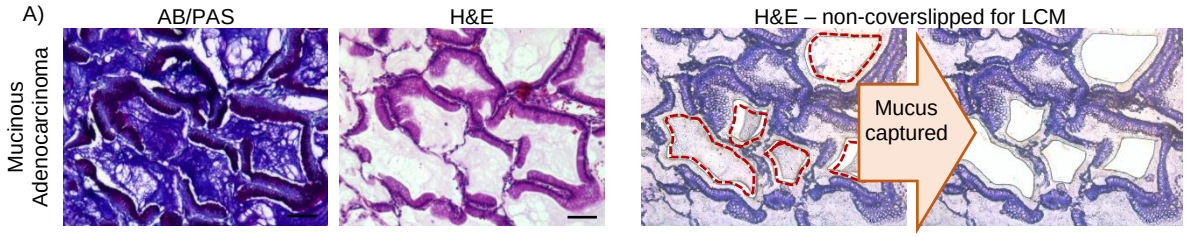


Figure 7: The composition of fibrotic mucus. Laser capture microdissection coupled mass spectrometry was performed on the mucus plugs of 6 Usual Interstitial Pneumonia (UIP) specimens. (**A**) A list of the most abundant secretome-associated proteins found in the fibrotic mucus shown as intensity Based Absolute Quantification (iBAQ). (**B**) Reactome pathway enrichment of UIP mucus represented as a dotplot. (**C**) Serial sections of UIP specimens stained for alcian blue/periodic acid Schiff's (AB/PAS) or immunostained for MUC5B, MUC5AC, BPIFB1, PIGR, and TF (N = 4 UIP specimens with 2 representatives shown). Scale bar represents 100 microns.

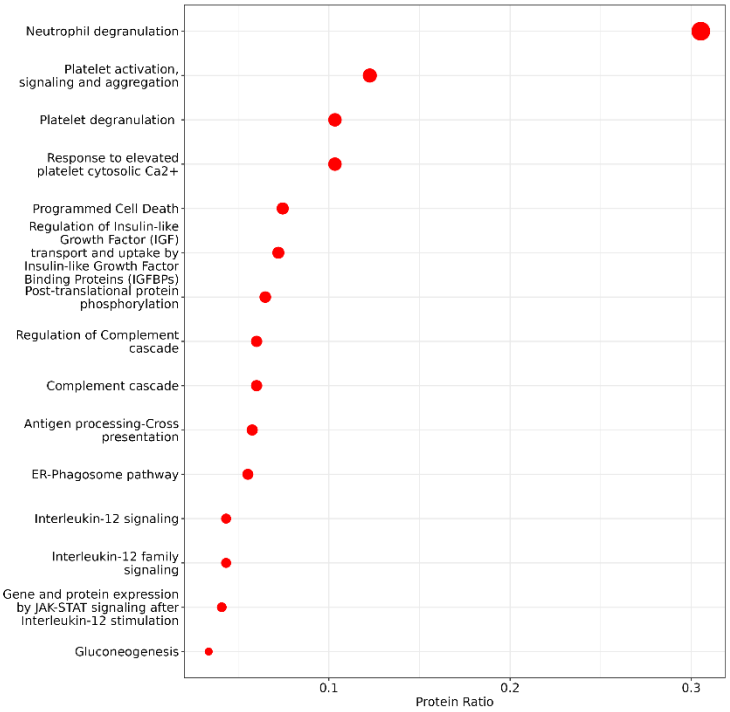
A)



B)

The 10 most abundant secretome-associated proteins in mucinous adenocarcinoma mucus. (171 proteins total)	
Gene, protein name	Median iBAQ ($x10^4$)
SFTPA1/2, Pulmonary surfactant-associated protein A1/2	3899
MUC5AC, Mucin-5AC	3897
APOA1, Apolipoprotein A-I	1976
DCD, Dermcidin	1896
TF, Serotransferrin	1731
DEFA3/1, Neutrophil defensin 3/1	1696
SERPINA1, Alpha-1-antitrypsin	1301
SFTPB, pulmonary surfactant-associated protein B	1296
MUC5B, Mucin-5B	902
FABP5, Fatty acid-binding protein	835

C)



Supplemental Figure 2: The mucus in mucinous adenocarcinoma (MA). (**A**) A MA specimen was serially sectioned at 5 microns and stained with alcian blue/periodic acid Schiff's (AB/PAS) stain or Hemetoxylin & Eosin (H&E). Mucus was laser microdissected for mass spectrometry analysis. Scale bar represented 100 microns. (**B**) A list showing the most abundant secretome-associated proteins found in the mucus of MA. (**C**) Dotplot visualization of the MA mucus using Reactome pathway enrichment.

283 secretome-associated protein found in the mucus of MA (**Supplemental Figure 2B,** a full
284 list in **Supplemental File 5**). Reactome pathway analysis demonstrates that MA mucus
285 is defined by 'Neutrophil degranulation', like fibrotic mucus, as the strongest category
286 (**Supplemental Figure 1C**).

287

288 We next compared the mucus of MA to UIP. In total, we detected 707 lung mucus proteins
289 (**Figure 8A**), with UIP having the most proteins detected (a full list in **Supplemental File**
290 **4**). A 3-D PCA analysis showed that UIP mucus samples largely cluster together, whereas
291 only one MA sample overlaps with UIP (**Figure 8B**). Quantitative analysis of our data
292 show that 9 proteins are significantly enriched in fibrotic mucus whereas 3 are significantly
293 enriched in cancer (MA) mucus (**Figure 8C,** a full list in **Supplemental File 7**). To validate
294 this result, we performed immunofluorescence on both MA (n = 3) and UIP specimens (n
295 = 4) for MUC5B, MUC5AC, and BPIFB1 (**Figure 8D**). We found that UIP mucus has
296 variable expression of MUC5AC (white arrows mark the absence of MUC5AC where
297 MUC5B/BPIFB1 is present) in comparison to MUC5B. Note that one mucus plug was
298 positive for MUC5AC, which resembles the chromogenic patchy/incomplete stain in
299 **Figure 7B**; MUC5AC has been previously reported to have variable staining [10].
300 Inversely, MA mucus was positive for MUC5AC, whereas MUC5B and BPIFB1 are largely
301 negative at the immunofluorescence level (**Figure 8E**, white arrows point where
302 MUC5B/BPIFB1 are absent). Thus, a distinguishing factor for fibrotic/UIP mucus is the
303 high abundance of MUC5B and BPIFB1, whereas MUC5AC is predominant in MA mucus.

304

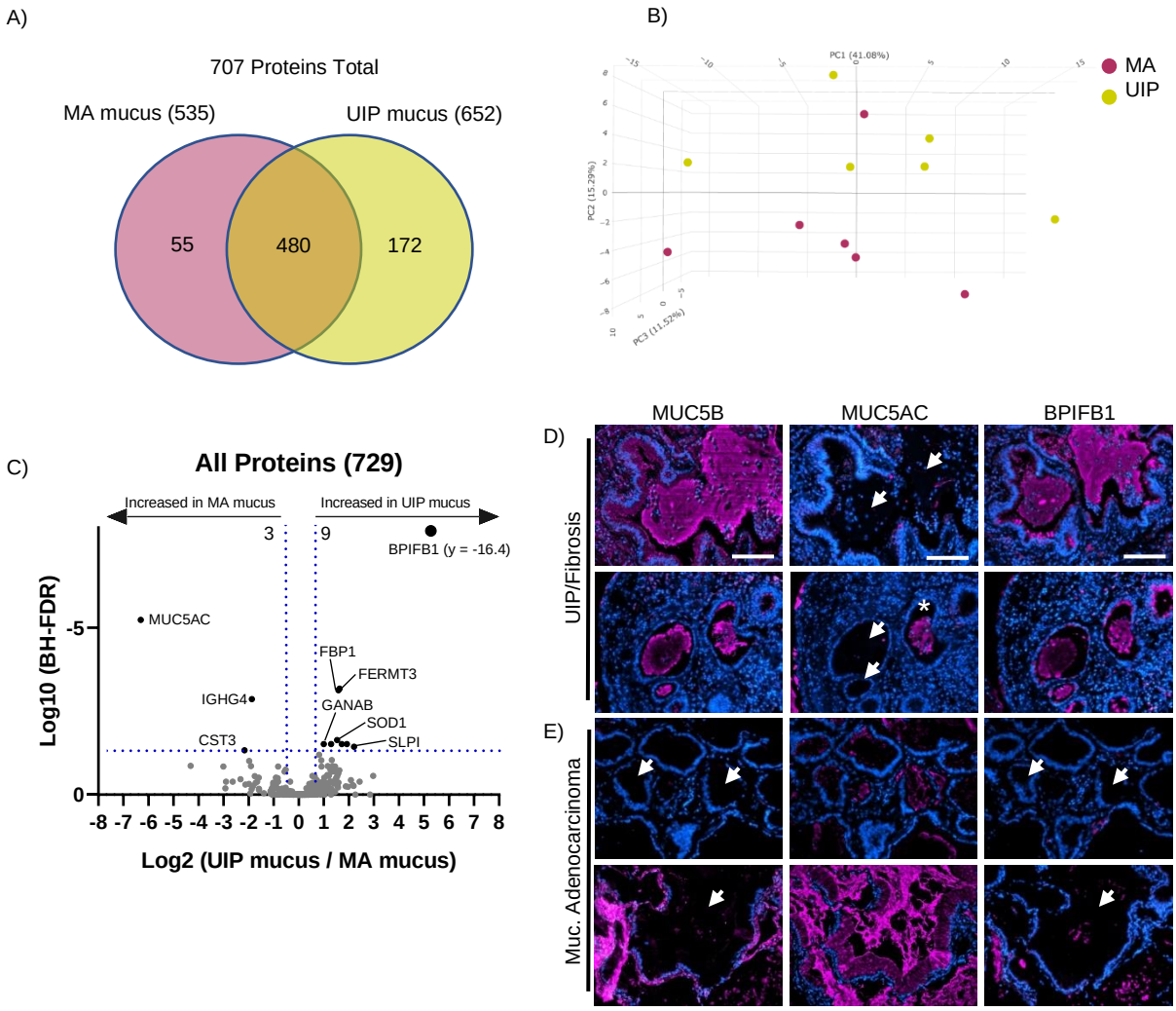


Figure 8: The mucus in usual interstitial pneumonia has distinct features from mucinous adenocarcinoma. (**A**) Venn diagram showing the proteins detected in the mucus of mucinous adenocarcinoma (MA) or usual interstitial pneumonia (UIP). (**B**) 3-dimensional principal component analysis for each mucus type. (**C**) Volcano plot comparing the UIP mucus to MA mucus showing the negative natural log of the false discovery values (FDR) values plotted against the base 2 log (fold change) for each protein. (**D**) Immunofluorescence for MUC5B, MUC5AC, and BPIFB1 in UIP mucus (n = 4 specimens) and MA mucus (n = 3 specimens) with representative images shown for each disease type. White arrows points to regions of mucus accumulation and asterisk shows positivity of MUC5AC within UIP mucus. Scale bar represents 100 microns.

Discussion

In this work, we produced an unbiased spatial proteomic profile of the non-fibrotic, fibrotic uninvolved and honeycomb airway cells to create a tissue map that defines pathway changes along the progression of lung fibrosis (**Figure 9**). We showed that the structurally intact low-in-mucus airway cells in uninvolved regions of the fibrotic lung have an abnormal protein signature with increased Slits/Robo pathway as the strongest category; Slits/Robo pathway is also increased in the fibrotic HC airway cells. We confirmed that the fibrotic honeycomb airways are the site of mucin biogenesis with other categories related to protein modification and transport increased. Importantly, many proteins associated with ciliogenesis are decreased or absent from the fibrotic HC airways. In addition, honeycombing is associated with decreased extracellular matrix organization and elastic fibre formation. Lastly, the fibrotic mucus is enriched with immune defence proteins, including BPIFB1 and MUC5B, and is enriched with neutrophil degranulation pathway.

Other groups support our results that demonstrate that the fibrotic uninvolved airway cells are abnormal at the structural and genetic level. Lung fibrosis is associated with a variety of genetic risk factors affecting epithelial cells, which may abnormally prime lung airway cells to fibrosis initiation [61]. Structurally, regions without microscopic fibrosis are shown to have reduced numbers of terminal airways and have an increase of airway wall areas [4-6], suggesting that early lung airway perturbations precede fibrotic extracellular matrix remodelling. Thus, it is plausible that airway cell dysfunction is an early event in the fibrotic process. Further LCM-MS studies with precise distance registration and patient

330

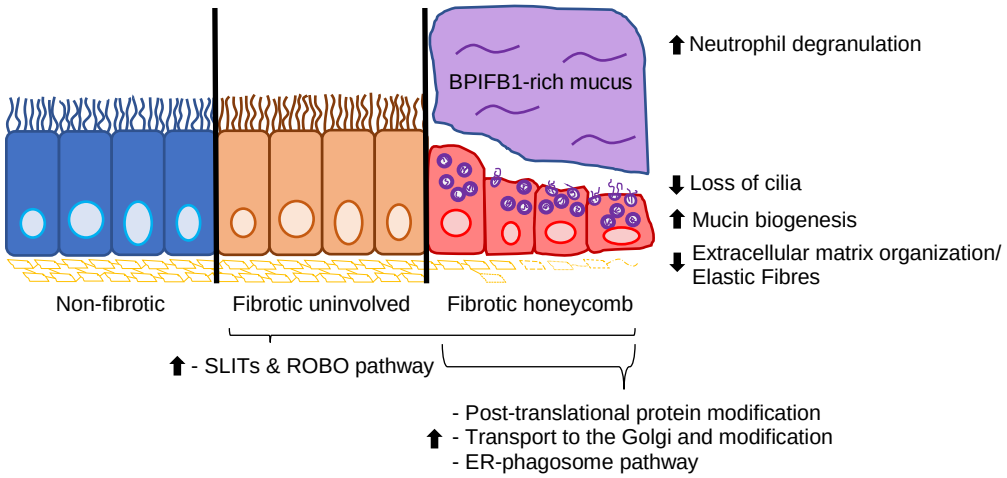


Figure 9: The fibrotic honeycomb airway. Spatial proteomics reveal that the fibrotic uninvolved airway cells (found in regions of structurally intact lung) have an abnormal protein signature. The fibrotic uninvolved airway cells, like the honeycomb (HC) airway cells (a pathological feature of lung fibrosis), are over-represented in proteins associated with SLITs and ROBO pathway. The fibrotic HC airway cells are further defined by increased pathways associated with mucin biogenesis, and a loss of both cilia and extracellular matrix organization/elastic fibres. We find that the mucus proteome is enriched with neutrophil degranulation pathway, with a marked increase of BPIFB1 protein.

331 genotyping will inform whether there exists a transition zone where a normal airway
332 proteome is present, or perhaps it may be that the entire airway proteome is abnormal.

333

334 Aberrant ciliogenesis has previously been described in UIP/IPF. Whole transcriptomic
335 studies demonstrate an elevation of cilium gene expression [62]. In contrast, our results
336 demonstrate a reduced cilia-associated protein profile within the fibrotic HC airway cells.
337 This likely reflects the advancement of our spatial proteomic capability or that previously
338 observed increases in ciliary gene expression represent a compensatory mechanism
339 occurring in response to loss of ciliated epithelial cells in fibrosis. Structurally,
340 transmission electron microscopy demonstrates that the UIP/IPF distal airways display
341 defects in microtubule organization, which will have detrimental effect on cilia function
342 [39]. In the context in cystic fibrosis, it is reported that airway epithelial also have
343 decreased ciliated cells with enhanced mucin expression [63]. Thus, future studies
344 determining the mechanism of deranged ciliogenesis is warranted.

345

346 Current literature suggests that basal airways cells differentiate into either mucin
347 producing cells or cilia-containing cells [64]. Our spatial proteomic data fits the notion that
348 the HC airway microenvironment directs the differentiation of basal airway cells into mucin
349 producing cells whereas the uninvolved airway microenvironment favors ciliated cells.
350 Given that extracellular matrix governs cell differentiation and function [65], we speculate
351 that changes to ECM properties (mechanical, composition, and topography) may play a
352 role in airway cell differentiation and function. Prior work utilizing decellularized COPD

353 airway tissue as a scaffold for cell-matrix interactions (as compared to donor) show that
354 COPD matrix dramatically affects cilia gene expression in epithelial cells [66]. Other
355 studies using decellularized UIP/IPF tissue confirm that fibrotic matrix is a driver of fibrosis
356 progression [67]. Therefore, the changes in mucin and cilia-associated proteins may be
357 reflective, or a consequence of the changes in airway ECM properties.

358

359 Our spatial proteomic data characterizing fibrotic HC airway cells (MUC5B-positive) are
360 in agreement with sc-RNAseq data characterizing *MUC5B*-positive secretory cells in
361 human lung. In one study, secretory airway cells have increased RNA expression of
362 *MUC5B*, *LCN2*, *BPIFB1*, *SERPINB3*, *S100P*, *RARRES1*, *TSPAN8*, *CP*, and *FAM3D* [68],
363 which are also increased or uniquely expressed at the protein level in fibrotic HC airway
364 cells. Other mRNAs increased in *MUC5B*-positive secretory cells include *TSPAN1*,
365 *AKR1C1*, *ZG16B*, *GSTA1*, and *SCGB1A1,* which are unchanged at the protein level in
366 the fibrotic HC airway cells. A separate study showed that *MUC5B*-positive cells by sc-
367 RNAseq have increased mRNA expression of *SCGB1A1*, *SCGB3A1*, *SLPI*, *BPIFB1*,
368 *LCN2*, and *WFDC2* [69]. At the protein level, SLPI, BPIFB1, LCN2, and WFDC2 are
369 increased or uniquely expressed in the fibrotic HC airway cells (SCGB1A1 and SCGB3A1
370 are unchanged at the protein level). Thus, the fibrotic HC airway cells represent a
371 secretory cell phenotype. Future work integrating spatial multi-omic analysis (RNA and
372 protein) will further our understanding of lung fibrosis.

373

To our knowledge, we are the first to determine the composition of UIP mucus plugs by using an LCM-MS approach. This approach allows precise capture of the entirety of mucus plugs without the introduction of contaminants (salivary and upper airway proteins) as seen by traditional BALF. Proteomic analysis of BALF (an unfixed or stained sample) from lung fibrosis patients show agreement with our findings. Several reports utilizing mass spectrometry approaches show increases of immunoglobulins, complement C3, transferrin, Apolipoprotein A1, plastin-2, annexin A2, and CCL18 in fibrotic lung BALF (summarized in [70]); all of which are detected in our LCM-MS dataset. In accord with our findings, Foster et al. demonstrated that MUC5B is an abundant protein in IPF BALF [57]. S100A9, detected by LCM-MS, is a potential BALF biomarker in IPF [71]. In addition, IPF patients with acute exacerbations show increased PIGR, LRG1, and SERPINA1 in BALF, which are also detected in our LCM-MS dataset [72]. Our LCM-MS approach is therefore a useful tool to determine the protein composition of mucus in archived FFPE specimens.

Our results demonstrate that the mucus found in lung cancer (mucinous adenocarcinoma; MA) has elevated levels of MUC5AC as compared to UIP mucus. A likely explanation is that the mucin-secreting cells comprising the UIP/IPF HC airway differ than the mucin-secreting cells in MA and/or that the environmental/immune signals controlling mucin production differ. For instance, reports show that there are 5X more MUC5B-positive cells versus MUC5AC-positive cells in the honeycomb airways of UIP/IPF, suggesting marked cell type heterogeneity [73]. In contrast, the morphology of MA cells are distinct and composed of goblet and/or columnar cells [74]. Another explanation is that *MUC5AC* gene expression is differentially regulated as compared to *MUC5B* [75]. For instance,

MUC5AC gene expression is increased by IL-13. In other disease settings, *MUC5AC* mRNA is increased in asthma, whereas *MUC5B* levels are decreased [76]. Further studies determining the functional consequence of varying MUC5AC to MUC5B protein ratios on fibrosis progression are needed.

Increases of BPIFB1 in both the UIP mucus and HC airway cells is of interest. BPIFB1 is a secretory protein that is implicated in immune regulatory functions and shown to have anti-tumor effects (reviewed in [77]). In other lung disorders, BPIFB1 is increased in cystic fibrosis, COPD, asthma, and IPF [78]. It is decreased in nasopharyngeal carcinoma, gastic cancer, and lung cancer, which agrees with our findings that mucinous adenocarcinoma mucus has low expression of BPIFB1. Understanding of its function in lung fibrosis is currently incomplete.

Conclusion:

Spatial proteomics has allowed us to create an unbiased protein tissue map of the fibrotic/UIP lung airway cells. We show that the fibrotic honeycomb airway cells are the active site of mucin biogenesis affiliated with a loss of cilia. Importantly, we show that the fibrotic uninvolved airway cells have an abnormal protein signature. Therapeutic intervention of the fibrotic uninvolved airway cells may therefore slow fibrosis progression.

Materials and Methods:

418 ***Histological staining:*** Five micron sections of formalin-fixed and paraffin-embedded
419 (FFPE) specimens were H&E-stained by using an automated stainer (Leica XL) at the
420 University of Manchester Histology Core Facility as previously described [19]. Importantly,
421 slides were stored at 4^{o}C for up to one week while laser capture microdissection (LCM)
422 was being performed. Captured material was stored at -80^{o}C until all samples were ready
423 for mass spectrometry processing. Alcian Blue/Periodic Acid Schiff (AB-PAS) was
424 performed as follows. De-paraffinized slide sections were incubated for 5 minutes in 1%
425 alcian blue 8GX (Sigma; A5268), 3% acetic acid. Slides were then washed in tap water
426 followed by a 5-minute incubation in 1% periodic acid (Sigma; 375810). Finally, slides
427 were washed in tap water and incubated in Schiff's reagent (Sigma – 3952016) for 15
428 minutes. After extensive washing in tap water, slides were coverslipped without
429 counterstain. For pentachrome, we followed a protocol as previously described [19].

430

431 For immunohistochemistry (IHC), we utilized the Novolink Polymer Detection Systems
432 (Leica, RE7200-CE) as previously described in detail [79]. We used the following
433 antibodies anti-BPIFB1 (Abcam; ab219098, titre 1:60,000), anti-elastin (Proteintech;
434 15257-1-AP; titre 1:16,000) anti-PIGR (Abcam; ab224086, titre1:8,000), anti-
435 serotransferrin (Abcam; ab268117, titre 1: 30,000). Anti-MUC5B (titre 1:10,000) and anti-
436 MUC5AC (titre 1:12,000) was previously purified and used here [80]. For all samples, we
437 used antigen heat retrieval using citrate buffer pH 6.0 (Sigma, C9999), with the exception
438 of EDTA pH 9.0 antigen heat retrieval for serotransferrin and elastin. Slides were
439 hematoxylin counterstained and coverslipped using permount (ThermoScientific, SP15).

440

For MUC5B, immunostains followed a modified protocol. After citrate buffer pH 6.0 antigen heat retrieval, the sections underwent reduction and alkylation. Sections were reduced by incubation at 37ºC for 30 minutes in 10 mM DTT, 0.1 M Tris/HCl pH 8.0. Sections are washed in water and then incubated in 25 mM Iodoacetamide, and 0.1 M Tris/HCl pH 8.0 for 30 minutes at room temperature (kept in the dark). Lastly, sections were washed in water followed by blocking and primary antibody incubation.

For immunofluorescence, dewaxed slides were subjected to citrate buffer pH 6.0 antigen heat retrieval and probed overnight with anti-MUC5AC (titre 1:100), anti-MUC5B (post reduction/alkylation; titre 1:100), or BPIFB1 (Abcam; ab219098, titre 1:100). Sections were then incubated with secondary anti-mouse fluorophore 680 (Invitrogen, A21058, 1:500) or anti-rabbit fluorophore 680 (Invitrogen; A21109; 1:500) for 1 hour. Sections were coverslipped using ProLong antifade with DAPI (Invitrogen; P36931).

Laser Capture Microdissection: The MMI CellCut Laser Microdissection System (Molecular Machines & Industries) was used to capture regions of interest on MMI membrane slides (MMI, 50102) as previously described [19, 20]. For this set of experiments, we collected a volume 0.03 mm^3 of tissue per sample.

Histological Imaging: For fluorescence microscopy, all stains were performed at the same time. In addition, images were taken at the same intensity utilizing EVOS FL

462 imaging system (ThermoScientific). For light microscopy, we used a DMC2900 Leica
463 instrument with Leica Application Suite X software.

464

465 **Data Availability:** We have deposited the raw mass spectrometry data files to
466 ProteomeXchange under the identifier of PD036465.

467

468 ***Mass spectrometry sample preparation:*** Samples were prepared as described [19, 20].
469 In short, samples underwent a series of steps to maximize protein yield, including high
470 detergent treatment, heating, and physical disruption.

471

472 ***Liquid chromatography coupled tandem mass spectrometry:*** The separation was
473 performed on a Thermo RSLC system (ThermoFisher) consisting of a NCP3200RS nano
474 pump, WPS3000TPS autosampler and TCC3000RS column oven configured with buffer
475 A as 0.1% formic acid in water and buffer B as 0.1% formic acid in acetonitrile. An injection
476 volume of 4 µl was loaded into the end of a 5 µl loop and reversed flushed on to the
477 analytical column (Waters nanoEase M/Z Peptide CSH C18 Column, 130Å, 1.7 µm, 75
478 µm X 250 mm) kept at 35 °C at a flow rate of 300 ηl/min with an initial pulse of 500 ηl/min
479 for 0.1 minute to rapidly re-pressurize the column. The separation consisted of a
480 multistage gradient of 1% B to 6% B over 2 minutes, 6% B to 18% B over 44 minutes,
481 18% B to 29% B over 7 minutes and 29% B to 65% B over 1 minute before washing for
482 4 minutes at 65% B and dropping down to 2% B in 1 minute. The complete method time
483 was 85 minutes.

The analytical column was connected to a Thermo Exploris 480 mass spectrometry system via a Thermo nanospray Flex Ion source via a 20 µm ID fused silica capillary. The capillary was connected to a fused silica spray tip with an outer diameter of 360 µm, an inner diameter of 20 µm, a tip orifice of 10 µm and a length of 63.5 mm (New Objective Silica Tip FS360-20-10-N-20-6.35CT) via a butt-to-butt connection in a steel union using a custom-made gold frit (Agar Scientific AGG2440A) to provide the electrical connection. The nanospray voltage was set at 1900 V and the ion transfer tube temperature set to 275 °C.

Data was acquired in a data dependent manner using a fixed cycle time of 1.5 sec, an expected peak width of 15 sec and a default charge state of 2. Full MS data was acquired in positive mode over a scan range of 300 to 1750 Th, with a resolution of 120,000, a normalized AGC target of 300% and a max fill time of 25 mS for a single microscan. Fragmentation data was obtained from signals with a charge state of +2 or +3 and an intensity over 5,000 and they were dynamically excluded from further analysis for a period of 15 sec after a single acquisition within a 10-ppm window. Fragmentation spectra were acquired with a resolution of 15,000 with a normalized collision energy of 30%, a normalized AGC target of 300%, first mass of 110 Th and a max fill time of 25 mS for a single microscan. All data was collected in profile mode.

505 ***Mass spectrometry data analysis and statistics:*** Raw data for regional samples were
506 processed using MaxQuant [81] version 1.6.17.0 against the human proteome obtained
507 from uniprot (May 2021) [82]. Raw data for UIP and MA mucus samples were processed
508 using MaxQuant [81] version 2.0.3.0 against the human proteome obtained from uniprot
509 (May 2022) [82]. All Maxquant processing were performed with a fixed modification of
510 carbamidomethylation of cysteine, with variable modifications of methionine oxidation and
511 protein N-terminal acetylation. Precursor tolerance was set at 20ppm and 4.5pm for the
512 first and main searches, with MS/MS tolerance set at 20ppm. A false discovery rate (FDR)
513 of 0.01 was set for PSM and protein level, up to two missed cleavages were permitted
514 and “match-between-runs” was selected.

515

516 Stastical analysis was carried out in R (v4.1.2) [83] using the MSqRob package (v0.7.7)
517 [84]. Significantly changing proteins were taken at a 5% false discovery rate (FDR).
518 Pathway analysis utilising Reactome Pathways was performed on significantly changing
519 proteins using the R package ReactomePA (1.38.0) [85]

520

521

522 ***Study Approval:*** For this study, we utilized a variety of Research Ethics Committee
523 (REC) protocols to obtain patient-consented lung tissue: REC#14/NW/0260 (provided by
524 JFB and RVV, Manchester, United Kingdom) for transplanted fibrotic lung;
525 REC#20/NW/0302 (provided by MAM and FG, Manchester, United Kingdom) for non-
526 fibrotic lung specimens; REC#18/NW/0092 (provided by Manchester Cancer Research

Centre Biobank, Manchester, United Kingdom) for mucinous lung adenocarcinoma. Usual Interstitial Pneumonia (UIP) specimens were defined by current guidelines [1, 86]. Non-fibrotic controls were collected from morphologically normal lung tissue distal to tumor during resection (fibrotic and control patient demographics may be found in **Supplemental Figure 3**). Mucinous adenocarcinoma (MA) was defined by current guidelines (MA patient demographics may be found in **Supplemental Figure 4**) [59]. In this study, we utilized 10 UIP, 6 non-fibrotic, and 6 MA specimens.

Author Contributions: J.A.H. designed and conducted all LCM-MS experiments and C.L. performed the associated analyses. L.D. and S.P. performed the immunohistochemistry/immunofluorescence and imaging. M.A.M. assisted in characterizing the histological stains and identification of clinical morphologies associated with Usual Interstitial Pneumonia and Mucinous Adenocarcinoma. R.V.V., J.F.B., M.A.M. and F.G. contributed to reagents. J.A.H. wrote the manuscript with all author inputs. J.A.H. and D.J.T conceived and supervised the project.

Funding: This was work was supported by the Wellcome Centre for Cell-Matrix Research's directors discretional funds (WCCMR; 203128/Z/16/Z) to JAH, Medical Research Council transition support (MR/T032529/1) to JFB, and Medical Research Council (MR/R002800/1) to DJT.

549 **Acknowledgements:** The authors would like to thank the Histology, BioMS, and
550 BioImaging Facilities at University of Manchester for making this work possible. The
551 authors would like to acknowledge the Manchester Allergy, Respiratory and Thoracic
552 Surgery (ManARTS) Biobank; Manchester Cancer Research Centre (MCRC) Biobank;
553 Transplant Unit staff at University Hospital of South Manchester NHS Foundation Trust;
554 and the North West Lung Centre Charity (NWLC) for supporting this project. The views
555 expressed in this publication are those of the authors and not necessarily those of
556 ManARTs, MCRC, NHS, or NWLC. In addition, we thank the study participants for their
557 contribution.

558

559 **References:**

560 1. Raghu, G., et al., *Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical*
561 *Practice Guideline.* Am J Respir Crit Care Med, 2018. **198**(5): p. e44-e68.
562 2. Lynch, D.A., et al., *Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White*
563 *Paper.* Lancet Respir Med, 2018. **6**(2): p. 138-153.
564 3. Collins, B.F., et al., *Sarcoidosis and IPF in the same patient-a coincidence, an association or a*
565 *phenotype?* Respir Med, 2018. **144S**: p. S20-S27.
566 4. Verleden, S.E., et al., *Small airways pathology in idiopathic pulmonary fibrosis: a retrospective*
567 *cohort study.* Lancet Respir Med, 2020. **8**(6): p. 573-584.
568 5. Xu, F., et al., *The transition from normal lung anatomy to minimal and established fibrosis in*
569 *idiopathic pulmonary fibrosis (IPF).* EBioMedicine, 2021. **66**: p. 103325.
570 6. Ikezoe, K., et al., *Small Airway Reduction and Fibrosis is an Early Pathologic Feature of Idiopathic*
571 *Pulmonary Fibrosis.* Am J Respir Crit Care Med, 2021.
572 7. Fingerlin, T.E., et al., *Genome-wide association study identifies multiple susceptibility loci for*
573 *pulmonary fibrosis.* Nat Genet, 2013. **45**(6): p. 613-20.
574 8. Lawson, W.E., et al., *Genetic mutations in surfactant protein C are a rare cause of sporadic cases*
575 *of IPF.* Thorax, 2004. **59**(11): p. 977-80.
576 9. Wang, Y., et al., *Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis*
577 *and lung cancer.* Am J Hum Genet, 2009. **84**(1): p. 52-9.
578 10. Seibold, M.A., et al., *The idiopathic pulmonary fibrosis honeycomb cyst contains a mucocilary*
579 *pseudostratified epithelium.* PLoS One, 2013. **8**(3): p. e58658.
580 11. Chilosi, M., et al., *Abnormal re-epithelialization and lung remodeling in idiopathic pulmonary*
581 *fibrosis: the role of deltaN-p63.* Lab Invest, 2002. **82**(10): p. 1335-45.

12. Schruf, E., et al., *Recapitulating idiopathic pulmonary fibrosis related alveolar epithelial dysfunction in a human iPSC-derived air-liquid interface model.* FASEB J, 2020. **34**(6): p. 7825-7846.
13. Carraro, G., et al., *Single-Cell Reconstruction of Human Basal Cell Diversity in Normal and Idiopathic Pulmonary Fibrosis Lungs.* Am J Respir Crit Care Med, 2020. **202**(11): p. 1540-1550.
14. Thornton, D.J., K. Rousseau, and M.A. McGuckin, *Structure and function of the polymeric mucins in airways mucus.* Annu Rev Physiol, 2008. **70**: p. 459-86.
15. Seibold, M.A., et al., *A common MUC5B promoter polymorphism and pulmonary fibrosis.* N Engl J Med, 2011. **364**(16): p. 1503-12.
16. Araki, T., et al., *Development and Progression of Interstitial Lung Abnormalities in the Framingham Heart Study.* Am J Respir Crit Care Med, 2016. **194**(12): p. 1514-1522.
17. Raghu, G., et al., *Incidence and prevalence of idiopathic pulmonary fibrosis.* Am J Respir Crit Care Med, 2006. **174**(7): p. 810-6.
18. Fahy, J.V. and B.F. Dickey, *Airway mucus function and dysfunction.* N Engl J Med, 2010. **363**(23): p. 2233-47.
19. Herrera, J.A., et al., *Laser capture microdissection coupled mass spectrometry (LCM-MS) for spatially resolved analysis of formalin-fixed and stained human lung tissues.* Clin Proteomics, 2020. **17**: p. 24.
20. Herrera, J., Dingle LA, Montero Fernandez MA, Venkateswaran RV, Blaikley JF, Schwartz MA, *The UIP/IPF fibroblastic focus is a collagen biosynthesis factory embedded in a distinct extracellular matrix.* JCI Insight, 2022(e156115).
21. McShane, A., et al., *Mucus.* Curr Biol, 2021. **31**(15): p. R938-R945.
22. Donoghue, L.J., et al., *Identification of trans Protein QTL for Secreted Airway Mucins in Mice and a Causal Role for Bpifb1.* Genetics, 2017. **207**(2): p. 801-812.
23. Griffin, S., et al., *Effect of pro-inflammatory stimuli on mucin expression and inhibition by secretory leucoprotease inhibitor.* Cell Microbiol, 2007. **9**(3): p. 670-9.
24. Koo, J.S., et al., *Restoration of the mucous phenotype by retinoic acid in retinoid-deficient human bronchial cell cultures: changes in mucin gene expression.* Am J Respir Cell Mol Biol, 1999. **20**(1): p. 43-52.
25. Kim, S.W., et al., *Regulation of mucin gene expression by CREB via a nonclassical retinoic acid signaling pathway.* Mol Cell Biol, 2007. **27**(19): p. 6933-47.
26. Ghandikota, S., et al., *Consensus Gene Co-Expression Network Analysis Identifies Novel Genes Associated with Severity of Fibrotic Lung Disease.* Int J Mol Sci, 2022. **23**(10).
27. Sajuthi, S.P., et al., *Nasal airway transcriptome-wide association study of asthma reveals genetically driven mucus pathobiology.* Nat Commun, 2022. **13**(1): p. 1632.
28. Kurtulmus, B., et al., *WDR8 is a centriolar satellite and centriole-associated protein that promotes ciliary vesicle docking during ciliogenesis.* J Cell Sci, 2016. **129**(3): p. 621-36.
29. Taschner, M. and E. Lorentzen, *The Intraflagellar Transport Machinery.* Cold Spring Harb Perspect Biol, 2016. **8**(10).
30. Kurtulmus, B., et al., *LRRC45 contributes to early steps of axoneme extension.* J Cell Sci, 2018. **131**(18).
31. Niwa, S., *The nephronophthisis-related gene ift-139 is required for ciliogenesis in Caenorhabditis elegans.* Sci Rep, 2016. **6**: p. 31544.
32. Ma, L. and A.P. Jarman, *Dilatory is a Drosophila protein related to AZI1 (CEP131) that is located at the ciliary base and required for cilium formation.* J Cell Sci, 2011. **124**(Pt 15): p. 2622-30.
33. Spektor, A., et al., *Cep97 and CP110 suppress a cilia assembly program.* Cell, 2007. **130**(4): p. 678-90.
34. Marszalek, J.R., et al., *Situs inversus and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II.* Proc Natl Acad Sci U S A, 1999. **96**(9): p. 5043-8.

35. Zhao, L., et al., *Heme-binding protein CYB5D1 is a radial spoke component required for coordinated ciliary beating.* Proc Natl Acad Sci U S A, 2021. **118**(17).
36. Tsurumi, Y., et al., *Interactions of the dynein-2 intermediate chain WDR34 with the light chains are required for ciliary retrograde protein trafficking.* Mol Biol Cell, 2019. **30**(5): p. 658-670.
37. Patnaik, S.R., et al., *The Role of RPGR and Its Interacting Proteins in Ciliopathies.* J Ophthalmol, 2015. **2015**: p. 414781.
38. McClintock, T.S., et al., *Tissue expression patterns identify mouse cilia genes.* Physiol Genomics, 2008. **32**(2): p. 198-206.
39. Kim, E., et al., *Aberrant Multiciliogenesis in Idiopathic Pulmonary Fibrosis.* Am J Respir Cell Mol Biol, 2022.
40. Tong, M., et al., *The Role of the Slit/Robo Signaling Pathway.* J Cancer, 2019. **10**(12): p. 2694-2705.
41. Greenberg, J.M., et al., *Slit and robo expression in the developing mouse lung.* Dev Dyn, 2004. **230**(2): p. 350-60.
42. Anselmo, M.A., et al., *Slit and robo: expression patterns in lung development.* Gene Expr Patterns, 2003. **3**(1): p. 13-9.
43. Park, J.S., et al., *Effect of Slit/Robo signaling on regeneration in lung emphysema.* Exp Mol Med, 2021. **53**(5): p. 986-992.
44. Valentijn, J.A., et al., *Novel localization of Rab3D in rat intestinal goblet cells and Brunner's gland acinar cells suggests a role in early Golgi trafficking.* Am J Physiol Gastrointest Liver Physiol, 2007. **293**(1): p. G165-77.
45. Hoang, O.N., et al., *Mucins MUC5AC and MUC5B Are Variably Packaged in the Same and in Separate Secretory Granules.* Am J Respir Crit Care Med, 2022.
46. Szul, T., et al., *Toll-Like Receptor 4 Engagement Mediates Prolyl Endopeptidase Release from Airway Epithelia via Exosomes.* Am J Respir Cell Mol Biol, 2016. **54**(3): p. 359-69.
47. Zeng, J., et al., *Phosphorylation of CAP1 regulates lung cancer proliferation, migration, and invasion.* J Cancer Res Clin Oncol, 2022. **148**(1): p. 137-153.
48. Cheong, A., et al., *A null allele of Dnaaf2 displays embryonic lethality and mimics human ciliary dyskinesia.* Hum Mol Genet, 2019. **28**(16): p. 2775-2784.
49. Murshed, M., et al., *Extracellular matrix mineralization is regulated locally; different roles of two gla-containing proteins.* J Cell Biol, 2004. **165**(5): p. 625-30.
50. Egashira, R., et al., *Diffuse Pulmonary Ossification in Fibrosing Interstitial Lung Diseases: Prevalence and Associations.* Radiology, 2017. **284**(1): p. 255-263.
51. Bassat, E., et al., *The extracellular matrix protein agrin promotes heart regeneration in mice.* Nature, 2017. **547**(7662): p. 179-184.
52. Schwanhausser, B., et al., *Corrigendum: Global quantification of mammalian gene expression control.* Nature, 2013. **495**(7439): p. 126-7.
53. Pilette, C., et al., *Lung mucosal immunity: immunoglobulin-A revisited.* Eur Respir J, 2001. **18**(3): p. 571-88.
54. Uhlen, M., et al., *The human secretome.* Sci Signal, 2019. **12**(609).
55. Rosa, B.A., et al., *IFN signaling and neutrophil degranulation transcriptional signatures are induced during SARS-CoV-2 infection.* Commun Biol, 2021. **4**(1): p. 290.
56. Hoenderdos, K. and A. Condliffe, *The neutrophil in chronic obstructive pulmonary disease.* Am J Respir Cell Mol Biol, 2013. **48**(5): p. 531-9.
57. Foster, M.W., et al., *Quantitative proteomics of bronchoalveolar lavage fluid in idiopathic pulmonary fibrosis.* J Proteome Res, 2015. **14**(2): p. 1238-49.
58. Travis, W.D., et al., *International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma.* J Thorac Oncol, 2011. **6**(2): p. 244-85.

59. Travis, W.D., et al., *The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification.* J Thorac Oncol, 2015. **10**(9): p. 1243-1260.
60. Guo, M., et al., *Gene signature driving invasive mucinous adenocarcinoma of the lung.* EMBO Mol Med, 2017. **9**(4): p. 462-481.
61. Wolters, P.J., et al., *Time for a change: is idiopathic pulmonary fibrosis still idiopathic and only fibrotic?* Lancet Respir Med, 2018. **6**(2): p. 154-160.
62. Yang, I.V., et al., *Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis.* Thorax, 2013. **68**(12): p. 1114-21.
63. Collin, A.M., et al., *Loss of ciliated cells and altered airway epithelial integrity in cystic fibrosis.* J Cyst Fibros, 2021. **20**(6): p. e129-e139.
64. Tilley, A.E., et al., *Cilia dysfunction in lung disease.* Annu Rev Physiol, 2015. **77**: p. 379-406.
65. Herrera, J., C.A. Henke, and P.B. Bitterman, *Extracellular matrix as a driver of progressive fibrosis.* J Clin Invest, 2018. **128**(1): p. 45-53.
66. Hedstrom, U., et al., *Impaired Differentiation of Chronic Obstructive Pulmonary Disease Bronchial Epithelial Cells Grown on Bronchial Scaffolds.* Am J Respir Cell Mol Biol, 2021. **65**(2): p. 201-213.
67. Herrera, J., et al., *Dicer1 Deficiency in the Idiopathic Pulmonary Fibrosis Fibroblastic Focus Promotes Fibrosis by Suppressing MicroRNA Biogenesis.* Am J Respir Crit Care Med, 2018. **198**(4): p. 486-496.
68. Basil, M.C., et al., *Human distal airways contain a multipotent secretory cell that can regenerate alveoli.* Nature, 2022. **604**(7904): p. 120-126.
69. Kadur Lakshminarasimha Murthy, P., et al., *Human distal lung maps and lineage hierarchies reveal a bipotent progenitor.* Nature, 2022. **604**(7904): p. 111-119.
70. Khan, T., et al., *Proteomics in idiopathic pulmonary fibrosis: the quest for biomarkers.* Mol Omics, 2021. **17**(1): p. 43-58.
71. Hara, A., et al., *S100A9 in BALF is a candidate biomarker of idiopathic pulmonary fibrosis.* Respir Med, 2012. **106**(4): p. 571-80.
72. Carleo, A., et al., *Proteomic characterization of idiopathic pulmonary fibrosis patients: stable versus acute exacerbation.* Monaldi Arch Chest Dis, 2020. **90**(2).
73. Conti, C., et al., *Mucins MUC5B and MUC5AC in Distal Airways and Honeycomb Spaces: Comparison among Idiopathic Pulmonary Fibrosis/Usual Interstitial Pneumonia, Fibrotic Nonspecific Interstitial Pneumonitis, and Control Lungs.* Am J Respir Crit Care Med, 2016. **193**(4): p. 462-4.
74. Dacic, S., *Pros: the present classification of mucinous adenocarcinomas of the lung.* Transl Lung Cancer Res, 2017. **6**(2): p. 230-233.
75. Bonser, L.R. and D.J. Erle, *Airway Mucus and Asthma: The Role of MUC5AC and MUC5B.* J Clin Med, 2017. **6**(12).
76. Woodruff, P.G., et al., *T-helper type 2-driven inflammation defines major subphenotypes of asthma.* Am J Respir Crit Care Med, 2009. **180**(5): p. 388-95.
77. Li, J., et al., *Molecular biology of BPIFB1 and its advances in disease.* Ann Transl Med, 2020. **8**(10): p. 651.
78. Bingle, C.D., et al., *What is top of the charts? BPIFB1/LPLUNC1 localises to the bronchiolised epithelium in the honeycomb cysts in UIP.* Thorax, 2013. **68**(12): p. 1167-8.
79. Herrera, J., et al., *Registration of the extracellular matrix components constituting the fibroblastic focus in idiopathic pulmonary fibrosis.* JCI Insight, 2019. **4**(1).
80. Kirkham, S., et al., *Heterogeneity of airways mucus: variations in the amounts and glycoforms of the major oligomeric mucins MUC5AC and MUC5B.* Biochem J, 2002. **361**(Pt 3): p. 537-46.

81. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.
82. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
83. *R Core Team. The R Project for Statistical Computing. https://www.R-project.org/. Accessed July 12, 2022.*
84. Goeminne, L.J.E., K. Gevaert, and L. Clement, *Experimental design and data-analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob.* J Proteomics, 2018. **171**: p. 23-36.
85. Yu, G. and Q.Y. He, *ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization.* Mol Biosyst, 2016. **12**(2): p. 477-9.
86. Raghu, G., et al., *Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline.* Am J Respir Crit Care Med, 2022. **205**(9): p. e18-e47.