JCI insight

Transcriptional regulatory model of fibrosis progression in the human lung

John E. McDonough, … , Wim A. Wuyts, Naftali Kaminski

JCI Insight. 2019. https://doi.org/10.1172/jci.insight.131597.

Research In-Press Preview Pulmonology

To develop a systems biology model of fibrosis progression within the human lung we performed RNAseq and microRNA analysis on 95 samples obtained from 10 idiopathic pulmonary fibrosis (IPF) and 6 control lungs. Extent of fibrosis in each sample was assessed by microCT measured alveolar surface density (ASD) and confirmed by histology. Regulatory gene expression networks were identified using linear mixed-effect models and dynamic regulatory events miner (DREM). Differential gene expression analysis identified a core set of genes increased or decreased before fibrosis was histologically evident that continued to change with advanced fibrosis. DREM generated a systems biology model of fibrosis progression (available at http: www.sb.cs.cmu.edu/IPFReg) that identified progressively divergent gene expression tracks with microRNAs and transcription factors that specifically regulate early or advanced fibrosis. We confirmed model predictions by demonstrating that expression of POU2AF1, previously unassociated with lung disease but proposed by the model as regulator, is increased in B-lymphocytes in IPF lungs and that POU2AF1 knockout mice were protected from bleomycin induced lung fibrosis. Our results reveal distinct regulation of gene expression changes in IPF tissue that remained structurally normal compared with moderate or advanced fibrosis and suggest distinct regulatory mechanisms for each stage.

Transcriptional regulatory model of fibrosis progression in the human lung

John E. McDonough[1*], Farida Ahangari[1*], Qin Li[1*], Siddhartha Jain[3], Stijn E. Verleden[2], Jose Herazo-Maya[1], Milica Vukmirovic[1], Giuseppe DeIuliis[1], Argyrios Tzouvelekis[4], Naoya Tanabe[5], Fanny Chu[5], Xiting Yan[1], Johny Verschakelen[2], Robert J. Homer[6,7], Dimitris V. Manatakis[8], Junke Zhang[8], Jun Ding[3], Karen Maes[2], Laurens De Sadeleer[2], Robin Vos[2], Arne Neyrinck[2], Panayiotis V. Benos[8], Ziv Bar-Joseph[3], Dean Tantin[9], James C. Hogg[5], Bart M. Vanaudenaerde[2], Wim A. Wuyts[2], Naftali Kaminski[1]

1. Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine, New Haven USA
2. Department of Chronic Diseases, Metabolism, and Ageing, KU Leuven, Leuven Belgium
3. Carnegie Mellon University of Computing Science, Pittsburgh USA
4. Division of Immunology, Biomedical Sciences Research Center "Alexander Fleming", Athens Greece
5. Centre for Heart Lung Innovation, University of British Columbia, Vancouver Canada
6. Department of Pathology, Yale University School of Medicine, New Haven USA
7. Pathology and Laboratory Medicine Service, VA CT HealthCare System, West Haven USA
8. Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh USA
9. Department of Pathology, University of Utah School of Medicine, Salt Lake City USA

* these authors contributed equally.

Corresponding author:

Dr. Naftali Kaminski
Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine
300 Cedar Street, TAC-S441D
New Haven, CT 06510, USA
Email: naftali.kaminski@yale.edu
Phone Number: +1 (203) 737-4612

Conflict of interest statement: Dr. Kaminski reports personal fees from Biogen Idec, Boehringer Ingelheim, Third Rock, Miragen, Pliant, Samumed, NuMedii, Indaloo, Theravance, LifeMax, Optikira, Three Lake Partners, outside the submitted work; in addition, Dr. Kaminski has patents on New Therapies in Pulmonary Fibrosis Peripheral Blood Gene Expression in IPF. All other authors declare no competing interests.

Key Words: microCT**,** disease progression, gene regulatory network

Abstract

To develop a systems biology model of fibrosis progression within the human lung we performed RNAseq and microRNA analysis on 95 samples obtained from 10 idiopathic pulmonary fibrosis (IPF) and 6 control lungs. Extent of fibrosis in each sample was assessed by microCT measured alveolar surface density (ASD) and confirmed by histology. Regulatory gene expression networks were identified using linear mixed-effect models and dynamic regulatory events miner (DREM). Differential gene expression analysis identified a core set of genes increased or decreased before fibrosis was histologically evident that continued to change with advanced fibrosis. DREM generated a systems biology model (www.sb.cs.cmu.edu/IPFReg) that identified progressively divergent gene expression tracks with microRNAs and transcription factors that specifically regulate mild or advanced fibrosis. We confirmed model predictions by demonstrating that expression of POU2AF1, previously unassociated with lung disease but proposed by the model as regulator, is increased in B-lymphocytes in IPF lungs and that POU2AF1 knockout mice were protected from bleomycin induced lung fibrosis. Our results reveal distinct regulation of gene expression changes in IPF tissue that remained structurally normal compared with moderate or advanced fibrosis and suggest distinct regulatory mechanisms for each stage.

Brief summary

Regulatory gene transcription model of lung fibrosis from tissue samples with varying disease of disease severity in the IPF lung

Introduction

Idiopathic Pulmonary Fibrosis (IPF) is the most common interstitial lung disease and the most lethal with mortality of 50% of patients 3-5 years after diagnosis (1, 2). It is a chronic, progressive fibrosing disease that occurs more commonly in older male subjects. Diagnosis is defined based on a combination of radiological or histological patterns of usual interstitial pneumonia (UIP) in the absence of known causes for these changes (3). IPF was initially considered a chronic inflammatory disease due to association of inflammatory infiltrates and the fibrosis present within these lungs (4, 5) and treated with immunosuppression. More recently, IPF has become thought of as being a disease of repetitive injury of the alveolar epithelial cells with aberrant wound healing resulting in fibrosis (2). This view has been further supported by convincing evidence that immunosuppressive therapies have a detrimental effect on the patient and that recently developed antifibrotic drugs are effective in slowing the progression of the disease (6, 7).

While the diagnosis of IPF relies on the integration of histologic, radiologic, and clinical criteria, the underlying pathologic changes are rooted in lung morphology, described as usual interstitial pneumonia (UIP) (8). UIP is characterized by marked fibrosis with or without honeycombing in a predominantly subpleural and paraseptal location with central areas being relatively spared. The fibrosis must be heterogeneous, with normal lung adjacent to established fibrosis. At the boundary between these regions are myofibroblast foci, defined by accumulation of immature hyaluronic acid rich matrix underneath epithelial cells undergoing injury and apoptosis. UIP is therefore characterized by temporal heterogeneity — the presence of acute or active disease (fibroblastic foci with or without epithelization) along with progressive disease (mature fibrotic scar) and spatial heterogeneity — the presence of fibrotic lung adjacent to histologically normal lung. In most cases there is no necrosis or substantial inflammatory cell infiltration, although some mild inflammatory response can be seen, especially in areas of collagen deposition. These patterns are not replicated in animal models of pulmonary fibrosis which is the reason

that despite the significant progress in our understanding of the molecular events underlying pulmonary fibrosis in animals (9) we have no information about what regulates the progression of IPF in the human lung. Understanding the molecular changes that characterize the progression of IPF from histologically normal regions into end-stage disease within the human lung would allow development of therapeutic strategies in all stages of disease.

To develop a model of the gene regulatory changes associated with progressive fibrosis in the human lung, and given the fact that it is impossible to directly observe the progression of fibrosis by sampling the same lung multiple time, we hypothesized that analysis of differentially affected regions within the IPF lung could provide an approximation of the trajectory of pathological changes suitable for a systems biology approach. We sampled multiple regions in human IPF or control lungs. We determined the extent of fibrosis using quantitative microCT imaging and tissue histology and classified the samples as controls, IPF with no fibrosis, IPF with moderate fibrosis and IPF with advanced fibrosis. We then performed RNA-seq and microRNA profiling to determine gene expression at these selected stages, applied a linear mixed-effects model to identify differentially expressed genes, and discovered a core set of gene expression changes that preceded the occurance of fibrosis in the IPF lung. We used computational modelling to reconstruct a dynamic gene regulatory model of IPF which uncovered expression changes distinct to different levels of fibrosis severity and identified well known, as well as novel regulators of disease progression. We confirmed model predictions that POU2AF1, a lymphocyte transcription factor previously unassociated with lung disease, is a regulator of fibrosis by immunohistochemistry and an animal model of lung fibrosis. The manuscript is accompanied by a supplemental web resource (10) that allows interactive and detailed mining of the model.

Results

Patient demographics. All subjects used in this study were males and matched for age. The use of only male subjects was due to limited numbers of female transplant recipients for IPF with only one lung being available for research over a period of 5 years. Patient demographics and lung function are described in table 1.

MicroCT quantitation and histological analysis of differentially affected samples in the IPF lungs. In total, 59 samples were available from IPF lungs and 36 samples from controls for microCT imaging (Figure 1 A-C). Samples were histologically assessed by a pathologist and 11 samples with emphysema, large airway, or large vessels were excluded from the analysis. Alveolar surface density (ASD) computed from microCT images as previously described (11) was significantly negatively correlated ($r = -0.81$, $p = 1.74 \times 10^{-12}$) with the Ashcroft histology fibrosis score (12), validating the use of ASD as an unbiased quantitative measure of fibrosis (Figure 1D). For our analysis, we clustered samples based on principal component of surface density and collagen 1 staining into control (n=35), samples obtained from IPF patients with ASD similar to control — IPF1 (n = 19), samples with lower surface density that reflects moderate fibrosis — IPF2 (n = 16), and samples with the lowest surface density that reflects advanced fibrosis — IPF3 (n = 14) (Figure 1E) and referred to them as distinct stages of fibrosis progression. While ASD was decreased in most IPF samples, IPF1 showed a large overlap with controls reflecting conserved normal lung structure (Figure 1F). This was consistent with the pathological report of tissue samples which found 14/19 of IPF1 samples had normal lung histology and the remaining 5/19 samples with minimal levels of fibrosis. As expected, total collagen staining was not increased in IPF1, but was significantly increased in later stages of IPF, starting with IPF2 (Figure 1G). Staining for collagen 1 revealed a trend towards increased staining with stage that reached significance in IPF3 (Figure 1H). The staining for collagen 3 did not reveal any changes potentially reflecting reduced sensitivity (Figure 1I).

Differential gene expression identified a core set of IPF genes that are changed before fibrosis is histologically evident. Using a linear mixed-effects model we identified 3672 genes that were differentially expressed in at least one IPF group with 2141 genes increased and 1531 genes decreased. In IPF1, 591 genes were increased and 990 were decreased, potentially reflecting the gene expression changes that precede overt histological changes. In IPF2, 1299 genes were increased and 890 were decreased and in IPF3, 1753 genes were increased and 932 were decreased. Of these, 908 genes were common to all groups of IPF suggesting they represented the core IPF genes (Figure 2A), which are changed even before fibrosis is histologically evident and continue to change as disease progresses.

The 100 genes with the greatest increase or decrease in expression at any stage (Figure 2B and supplemental table S1) were significantly overrepresented amongst these core genes (67/100, $p < 2.2x10^{-16}$). The 67 genes highest differentially expressed core genes included several collagens and extracellular matrix proteins showing overexpression (COL1A1, COL1A2, COL2A1, COL3A1, COL5A1, COL10A1, COL15A1, COMP) with COL1A1 being the highest expressed gene in every stage of IPF (Figure 2C). Despite COL1A1 being highly expressed, we did not find significant increase in the degree of staining of collagen 1 on tissue sections until IPF3, potentially reflecting RNA changes that precede changes in protein, or more likely limited sensitivity of the staining (Figure 1H). Some chemokines (CCL13, CCL18, CXCL6, CXCl13, CXCL14) were also among the most highly expressed genes in our dataset. CXCL14, a chemokine expressed by fibroblasts and associated with monocyte recruitment and inhibition of angiogenesis (13), and CCL18, a chemokine secreted by alveolar macrophages, have both been proposed as IPF biomarkers (14, 15). Among the highly overexpressed genes were the fibroblast genes CTHRC1, FAP, ISLR and THY1, and the bronchial epithelial cell markers KRT5, KRT14, KRT15, and KRT17. The metalloproteinase MMP7, previously observed in the alveolar epithelium and whose blood levels have been consistently associated with IPF disease progression and survival (16, 17), was overexpressed in all groups of IPF, as was MMP11 (Figure 1C). Genes related to B-cell function were also highly expressed and included CD79A, MZB1, previously

identified in IPF from a proteomic screen (18), as well as IGJ, the immunoglobulin J-chain linking IgM and IgA antibodies, and IGLL5, the immunoglobulin lambda light chain.

Among the most highly decreased genes in IPF were AGER, an alveolar type I marker previously shown to be decreased in IPF (19), innate immunity related genes including the S100 family genes S100A9, S100A9, and S100A12, and the neutrophil defensins DEFA1, DEFA3, and DEFA4. The solute carrier SLCO4A1 was also significantly underexpressed in all stages of IPF. This gene is involved in the transport of thyroid hormones recently implicated as potential therapies in IPF (20) (Figure 2B).

Globally, results were consistent with previous transcriptomic studies. Half (34/69) of the genes recently highlighted in a review of transcriptomic studies of IPF (21) were differentially expressed in at least one group of IPF in our dataset (supplemental table S2) and 16 were amongst our core set of IPF genes. This included 4 genes (CTHRC1, GREM1, FHL2, and MMP7) previously associated with disease severity as measured by percent predicted diffusing capacity of carbon monoxide ($\%DL_{CO}$) (22). GREM1, FHL2, and MMP7 showed a progressive increase related to disease severity in IPF, whereas CTHRC1 was increased in IPF1, but did not increase any further in IPF 2 or 3 (supplemental Figure S1). In comparing our core list of genes to differentially expressed genes from the Lung Tissue Research Consortium dataset of IPF lung tissues (GSE47460), as recently analyzed by us (23), we found 535 genes differentially expressed in both datasets with the vast majority of genes (527/535 genes, $p < 2.2x10^{-16}$) showing a similar direction of expression (supplemental Figure S2).

While many gene changes relevant to IPF disease were present from the earliest stages of IPF, other genes that have been associated with fibrosis or IPF pathogenesis were increased only in the severe stages. Among these genes are MUC5B, known to be increased and genetically associated with IPF (24–26). MUC5B was not changed in IPF1 but was increased in IPF2 and IPF3 (MUC5B: IPF1 p = 0.235, fold-change (fc) = 2.1; IPF2 p = 0.013, fc = 3.1; IPF3 p = 0.015, fc = 3.7, Figure 2C). FKBP10 has been shown to be increased in IPF and involved in regulating collagen production (27). In our data FKBP10 was significantly

increased only in IPF3 compared to controls (FKBP10: IPF1 p = 0.1, fc = 1.9; IPF2 p = 0.1, fc = 1.7; IPF3 p = 0.027, fc = 2.0). WNT signaling has also been associated with IPF disease pathogenesis, in particular involving epithelial cell developmental pathways (28). Similar to what has been shown previously (29), we found decreased WNT3A (WNT3A: IPF1 p = 0.089, fc = 0.61; IPF2 p = 0.035, fc = 0.53; IPF3 p = 0.016, fc = 0.51) and increased WNT10A (WNT10B: IPF1 p = 0.122, fc = 1.4; IPF2 p = 0.023, fc = 1.6; IPF3 p = 0.023, fc = 1.7) in IPF compared to controls but this was only significant in the severe stages of disease. Overall, these data indicate that many of the transcriptional changes in IPF are already present in regions with relatively normal lung structure and minimal change in cellular infiltration (IPF1) while other gene expression changes only occur in more severely affected regions with an obvious UIP pattern (IPF2 or IPF3) potentially reflecting cellular changes associated with lung remodeling such as honeycomb formation.

MicroRNAs differentially expressed in IPF are regulators of many of the differentially expressed genes but are stage specific. Of the 800 measured microRNAs, we identified 86 as differentially expressed between IPF and controls in at least one stage of disease (supplemental table S3). Examining the microRNA targets using the validated target dataset (miRTarBase) in the multi-miR database (30), we found 50% of differentially expressed genes (1860/3672 genes, $p = 9.97 \times 10^{-8}$) were predicted to be regulated by these 86 microRNAs. In focusing our observations to significant ($p < 0.05$) microRNAs and target genes with at least a 2 fold change in IPF to controls, we found 4 microRNAs were significant with greater than 2-fold increased expression (miR-127-3p, miR-21-5p, miR-382-5p, miR-495) that were known to downregulate 20 genes from the core gene set (supplementary table S4) with miR-21 being a regulator for all 20 of these genes. MiR-21 has previously been shown to have an important role in fibrogenesis (31). Pathway enrichment for these 20 genes reveals that they are related to HIF-1 (KEGG:04066, HIF-1 signaling pathway, $p = 5.65 \times 10^{-3}$) and thyroid hormone signaling (KEGG:04919, Thyroid hormone signaling pathway, $p = 8.79 \times 10^{-3}$).

Notably, microRNA changes seemed to be specific to distinct IPF groups (supplemental table S4 for detail). In IPF1, 6 microRNAs were decreased and associated with 16 increased target genes that included several collagens (COL1A1, COL1A2, COL3A1, COL5A2, COL10A1, COL15A1), all known to be regulated by miR-29c-3p, a well-established antifibrotic microRNA (32, 33), as well as MMP7 and TGFBI (Table S4). Five microRNAs that were increased in IPF1 (miR-21-5p , miR-127-3p, miR-382-5p, miR-451a, miR-495) targeted 43 decreased genes. These genes were enriched for the hypoxia pathway (KEGG:04066, HIF-1 signaling pathway, $p = 2.83x10^{-7}$) and included CLOCK, MAPK1, and VEGFA. In IPF2, 11 microRNAs, including let-7d-5p known to be decreased in IPF (34), targeted 18 increased genes that included collagens (COL3A1, COL5A2, and COL16A1), regulators of ECM such as MMP1 and HAS2, and the myofibroblast transcription factor TWIST1. Ten microRNAs were increased that targeted 114 downregulated genes with miR-155-5p, previously implicated in IPF (35), accounting for 48 of these genes and miR-21 accounting for 39. For IPF3, 8 microRNAs, including 3 from the let-7 microRNA family, were decreased. These microRNAs targeted 31 increased genes including ECM proteins (COL1A2, COL3A1, COL16A1, FN1), ECM regulators (LOXL2, HAS2, MMP2, MMP14) and transcription factors (SOX9, TWIST1). Twelve microRNAs were increased in this stage that targeted 130 decreased genes with miR-34a-5p (targeting 33 genes), joining previously mentioned miR-155-5p (43 genes) and miR-21-5p (35) genes. As also observed more formally in the DREM analysis, decreased microRNAs exhibited a shifting regulatory effect on gene expression with disease severity, in which the miR-29 family dominated in mild disease and the let-7 in severe disease, whereas increased microRNAs exhibit an additive effect with miR-21 being the main regulator in IPF1 with miR-155 joining in IPF2 and miR-34a in IPF3.

Dynamic regulatory events miner (DREM) analysis identified distinctly regulated gene expression tracks related to extent of fibrosis in the lung. To develop a systems biology model of the distinct regulatory events active in the differentially affected IPF microenvironments, we applied the DREM algorithm on the differentially expressed genes and microRNAs. The model identified 13 gene expression tracks regulated

at 4 nodes (nodes 1-4, Figure 3A) that were established in the minimally or non-affected regions of IPF (i.e. from control to IPF1) that differed in their patterns of gene expression (increased, nodes 1-3 and decreased, node 4), as well as regulation, and biological function (see below). At IPF2, there was divergence with each track being regulated by a distinct set of regulators (nodes 5-17, Figure 3A) indicating a shift in gene expression regulation between more normal regions and advanced stages of the disease. At this stage, 8 tracks were characterized by increased gene expression (tracks A-H, Figure 3A) and 5 decreased tracks (I-M, Figure 3A). Only one decreased advanced track (M, Figure 3A) diverged from a node that was previously characterized by increased gene expression). The tracks also differed in their slope of gene expression. Increased tracks A, B and all of the decreased tracks (I-M) exhibited their largest changes between control and IPF1, potentially reflecting changes that occurred in IPF tissue before overt changes in tissue histology were present (Figure 3B: A, B, I-M). Tracks C, D, E and H exhibited the majority of gene expression changes between IPF1 and IPF2, potentially reflecting changes occurring during the development of more advanced disease (Figure 3B: C-E, H). Tracks F and G were characterized by genes that changed mostly at the transition of IPF2 and IPF3, potentially reflecting the changes that occur in the transition to end stage disease (Figure 3B: F, G).

The gene expression tracks differed in their functional enrichments (Figure 3C and dynamic Sankey image in supplemental website (10)) suggesting distinct trajectory for changes in molecular function associated with increased fibrosis severity. Both increased tracks A and B were highly enriched for extracellular matrix (ECM) genes (Figure 3C, BIOCARTA M5889: Ensemble of genes encoding ECM and ECM-associated proteins; track A: $p = 1.79 \times 10^{-12}$; track B: $p = 2.46 \times 10^{-16}$). Genes in track A included the fibrillar collagen types 1 and 3 (COL1A1, COL1A2, COL3A1) and track B included genes for FACIT collagens type 14, 15, and 16 (COL14A1, COL15A1, COL16A1). The third track (M) to emerge from node 1 (Figure A-B) was decreased in expression and enriched with antimicrobial peptides (Reactome:1457777, $p = 1.21 \times 10^{-3}$) that included the neutrophil defensins (DEFA1, DEFA3, DEFA4). This track also included AGER, expressed

in type I alveolar epithelium and previously shown to be decreased in IPF (19). The main regulator of node 1 was the antifibrotic microRNA miR-29c in the mild stage (IPF1, node 1), and the transcriptions factors NF1, NKX6-1, RUNX2, and POU2AF1. As mentioned before, we observed the biphasic regulation of tracks A and B with miR-29 mainly regulating node 1 while let-7 family taking over the advanced stages (nodes 5 and 6, Figure 3A and supplementary website). The key transcription factor for node 6 (increased genes from IPF 1 to IPF 2, track B) was POU2AF1, a lymphocyte transcription factor never been associated with lung fibrosis which we went on to validate (see below). HMGA2, the transcription factor and target gene of Let-7 and previously been reported to be involved in fibrosis, was also a regulator of this node (34). The transcription tracks C, D, and E, with gene expression mainly increased at the moderately affected regions between minimally and most affected regions (IPF2, Figure 3B) were enriched for genes related to activation of the lymphocyte immune response. Specifically, track C was associated with regulation of lymphocyte activation (GO:0051249, $p = 8.98x10^{-10}$) and included the WNT gene WNT10A, track D with immunological synapse (GO:0001772, p = 0.027), and track E with hematopoietic or lymphoid organ development (GO:0048534, $p = 3.18x10^{-4}$) (Figure 3C). The transcription factor LEF1, the canonical WNT signaling transcription factor (36), was the main regulator for node 2. Tracks F, G, and H, where gene expression was mainly increased in most affected IPF regions (IPF3) were enriched with DNA related functions; track F for cell division (GO:0051301, p = 0.049) and chromosome organization (GO:0051276, p = 0.049), track G with chromosome breakage (GO:0031052, p = 0.047) and DNA elimination (GO:0031049, p = 0.047), and track H for ion channel proteins, (GO:0006812, $p = 3.46x10^{-5}$) (Figure 3C). These tracks were regulated by numerous microRNAs, including the miR-30 family (miR-30A/B/D/E) (node 3). Tracks I-M were comprised of genes decreased in minimally affected IPF tissue (IPF1) and reflected the extent of functions lost in the IPF lung even before overt histological changes were observed. Track I was enriched for olfactory signaling (Reactome:1269583, $p = 4.02x10^{-17}$), track J for mitochondrial transport (GO:0006839, p = 0.030), protein targeting to mitochondria (GO:0006626, p = 0.038), and mitochondrial

organization (GO:0007005, p = 0.049), potentially reflecting the loss of normal mitochondrial function recently reported in IPF (37). Track K was enriched for alcohol metabolism (GO:1902652, $p = 4.01x10^{-4}$) and cholesterol biosynthesis (Reactome:1270037, $p = 1.02x10^{-4}$) potentially reflecting the metabolic shift that occurs in the IPF lung (Figure 3C). Track L was highly enriched for genes related to the circulatory system (GO:0003013, circulatory system process) ($p = 7.21x10^{-7}$) reflecting inhibition of angiogenesis in progressive pulmonary fibrosis (Figure 3C). The primary regulator of node 4 that regulated tracks I-L was the microRNA miR-205, which had increased expression in IPF compared to control and has been demonstrated to have an extensive role in inhibiting angiogenesis (38, 39). Interestingly, regulators for node 15, that regulated path K, included the thryroid hormone receptors A and B (THRA and THRB) suggesting their role in altered metabolic regulation in IPF. The decreased track M was enriched with antimicrobial peptides (Reactome:1457777, $p = 1.21x10^{-3}$) that included the neutrophil defensins (DEFA1, DEFA3, DEFA4) potentially reflecting early changes in innate immunity during fibrosis.

Analysis of cellular enrichment of gene expression tracks. To identify the contribution of shifts in cellular populations to the gene expression tracks, we used the annotated cell types from a publicly available single cell RNA-seq dataset of lung tissues (40). In general, decreased tracks were enriched for AT1, AT2 and endothelial vascular cell markets (tracks J-M, Figure 3D), early increased tracks were fibroblasts and ciliated epithelial cell markers (track B, Figure 3D), and progressive tracks were enriched with B lymphocytes, basal cell and ciliated cells markers (tracks C-E, Figure 3D). Macrophage and monocyte genes were enriched in the early decreased tracks (tracks J-L, Figure 3D) and in the advanced increased tracks (tracks D & E, Figure 3D) potentially reflecting the enhanced recruitment of these cells to areas with advanced remodeling. Track A, G and H were not enriched with cell markers potentially representing modular changes in fibrosis that are not dependent in shifts in cell poppulations.

Pou2af1$^{-/-}$ mice exhibit attenuated lung fibrosis in bleomycin model. As our DREM model identified POU2AF1 as being a regulator of node 6 in the fibrotic track B, we sought to confirm a mechanistic link

between POU2AF1 and fibrosis using a transgenic mouse model. For this, we administered bleomycin to Pou2af1$^{-/-}$ mice and wild-type (Pou2af1$^{+/+}$) littermates and measured collagen and hydroxyproline levels. Pou2af1$^{-/-}$ mice exhibited significantly lower expression of Col1α1 mRNA level in comparison with wild-type counterpart with saline or after administration of bleomycin (Figure 4A). Saline treatment of wild-type and Pou2af1$^{-/-}$ mice showed no difference in hydroxyproline levels. At day 14 after intratracheal bleomycin administration, wild-type littermates developed pulmonary fibrosis while Pou2af1$^{-/-}$ mice exhibited lower hydroxyproline levels indicative of reduced deposition of collagen in the lung (Figure 4B). In line with these findings, Pou2af1$^{-/-}$ mice showed reduced histological evidence of fibrosis by H&E, α-SMA, and Masson's trichrome staining as compared to wild-type littermates (Figure 4C).

Human POU2AF1 was localized in lymphocytic aggregates in the IPF lung. As we confirmed that POU2AF1 was able to directly regulate fibrosis, we sought to identify the cell types which may be expressing this gene in the human lung. In examining human lung samples from patients with IPF using immunostaining, we found extensive POU2AF1/BOB1 staining in lymphocytic aggregates present in fibrotic regions but no staining in bronchial epithelium (Figure 4D), suggesting lymphocytes are the primary source of POU2AF1. Reanalyzing a single-cell RNA sequencing dataset from Adams. et al. (41) showed POU2AF1 is highly expressed in B-cells (idenitfied by expression of CD79) with some expression in ciliated epithelial cells (identified by FOXJ1). In IPF, POU2AF1 is increased only in B-cells but not the ciliated airway bronchial cells (Figure 4E).

Discussion

In this study, we quantitatively characterized differentially affected regions in human lungs to model progression of IPF in the human lung. We performed RNA-seq and microRNA profiling of microCT characterized regions in IPF lungs and compared gene expression between normal histology lungs from donors with IPF lungs where the lung tissue was quantified and categorized into minimally affected, moderately affected, and end-stage disease severity. Our differential expression analysis identified a core set of genes which changed even before fibrosis is histologically evident and continue to change as disease progresses. Our systems biology analysis allowed us to identify distinct tracks of gene expression that increase or decrease with progressive fibrosis in the lung, as well as their regulation by microRNAs and transcriptions factors. Among these were regulators previously identified such as the miR-29, let-7, and miR-21 microRNAs. These regulators drive profibrotic programs but seemed to have distinct and previously unrecognized roles at different stages of the histological progression of IPF. Finally, we confirmed our model prediction that POU2AF1, a lymphocyte transcriptional co-activator previously not associated with lung disease, was a regulator of profibrotic transcriptional programs using immunohistochemistry and the bleomycin-induced fibrosis model.

Studies in IPF using transcriptomic technologies have led to the identification of numerous genes and microRNAs involved in pulmonary fibrosis such as MMP7, CCL18, COMP, SPP1, RXFP1 , miR-29 , miR-21, Let-7, among others (16, 24, 32, 33). However, few studies have considered these results in the context of the tissue microenvironment and whether they relate to early or late-stage events in disease progression. This is important as IPF is a disease that progresses from the periphery to the center with different lung regions having distinct pathological features likely corresponding to different stages of disease, as occurs in a heterogeneous and progressive disease (42). In contrast to previous studies, we focused on analyzing gene expression in distinct IPF microenvironments, by cutting out differentially affected regions in the same lung, then characterizing these samples by microCT and performing RNAseq

and microRNA profiling on exactly the same core that was imaged. Thus, we were able to correlate gene expression changes with the extent of fibrosis in the tissues which led to some unexpected and novel insights. The first was that normal histology tissue in IPF is very different from similar histology tissue obtained from a human without IPF. The changes are impressive with over a thousand genes differentially expressed including known IPF genes. A recent study examining mild and severely fibrotic regions in 3 IPF explant lungs also showed changes were present in macroscopically normal but microscopically fibrotic tissues (43). We extended these findings by showing that even microscopically normal tissues, as quantitatively measured by microCT and pathological assessment, showed these molecular changes and accounted for the multiple samples examined per subject providing a more accurate description of these changes.These gene expression changes continue to progress, as can be seen in tracks A and B, corresponding with increased fibrosis, but these changes are already present in normal histology lung. For instance, the gene expression of many collagens is increased in the minimally affected IPF lung, before changes at the protein levels are seen. Similarly, miR-29 microRNA, potentially the most universal antifibrotic microRNA (32, 44–46), is also decreased at this stage demonstrating that histologically normal tissue from patients with IPF is molecularly abnormal. The second finding was that different genes, commonly thought as characteristic of IPF, showed distinct patterns, some increasing early and with minimal further changes (those in track A and B), some consistently increasing with progression of fibrosis (tracks C and D), and some changing only in the most severe disease (tracks F and G). The third finding was that the regulation of fibrosis is, at least to some extent, modular and thus the end results, increased ECM proteins, are driven by different microRNAs and transcription factors in minimally affected tissue compared with progressively affected or end stage tissue (Figure 5). For example, the regulation of fibrillar collagens is particularly relevant for IPF as it is the primary component of fibrotic extracellular matrix. The microRNA miR-29c, considered a master regulator for fibrosis (47, 48), has a direct effect on inhibiting fibrillar collagen expression and has been studied as a possible therapeutic for pulmonary fibrosis (33).

Despite high levels of collagen expression throughout IPF disease progression, we found that miR-29c is mostly a regulator at early stages of disease with the let-7 family of microRNAs taking over at later stages. Various members of the let-7 family of microRNAs are decreased in the IPF lung at progressive stages and regulate HMGA2, collagen 1 and potentially other pro-fibrotic molecules (32, 34), however, little was known about their regulation in the IPF lung. Taken together, these results uncover for the first time the diversity of gene expression, highlight the complexity and potential stability of the fibrotic networks, and the need to develop therapeutics that target the early events and potentially prevent the development of disease, as well as therapeutics that target late events to potentially reverse the areas of progressive disease.

A common difficulty in studies that apply high throughput profiling approaches to systems biology models of disease, is the requirement to provide some validations to the model predictions. In our study we applied several layers of validation. The first was that we assessed whether findings of our model were consistent with what was previously known. This was done at the level of the genes and showed our results were highly consistent with previous transcriptomic analyses of IPF lungs (22, 49, 50), including a recent paper that applied gene expression module networks to identify co-expression pathways in IPF (23). A second layer of validation was whether our model would identify pathways and regulators consistent with what is known in the literature – two examples are the WNT pathway and the presence of mitochondrial dysfunction. The WNT pathway is well known to be involved in pulmonary fibrosis (29, 51). We found increased WNT10A, in track C, associated with the major transcription factor for WNT signaling, LEF1, at node 2. The observed decrease in WNT3A was also consistent with previous studies (29). In recent years, mitochondrial dysfunction has been of significant focus in IPF (20, 37). Impressively, we have found that two decreased tracks J and K were enriched for mitochondrial organization, structure and function, as well metabolic pathways required for cellular function. The third layer was experimental and as it was impossible to check every single prediction, we chose to focus on POU2AF1, never previously associated

with fibrosis. POU2AF1, a transcriptional coactivator that associates with POU2F1 or with POU2F2 is mainly expressed in B-cells in relation to the formation of germinal centers, and at much lower levels in T-cells (52, 53). Because POU2AF1 was identified from our model as a regulator at node 6 of the fibrotic track B, we confirmed that it was increased in the IPF lung using immunohistochemistry and single cell RNAseq data. Using a knockout mouse demonstrated that it has a role in bleomycin induced pulmonary fibrosis. We had previously found POU2AF1 to be coexpressed with B-cell related genes in IPF (23) and reanalyzing IPF single cell data (41) we also showed that the predominant expression and increase in of POU2AF1 in IPF is in B-cells. While these data support a role for POU2AF1 in B-lymphocytes, that are increasingly implicated in fibrosis (18, 54, 55), they are of course limited, especially considering that POU2AF1 was also recently found in airway epithelial cells (56). Thus, while our results establish a functional link between POU2AF1 to pulmonary fibrosis, and confirm our model predictions, more detailed studies using cell specific loss and gain of function experiments will be required to assess the specific role of POU2AF1 in fibrosis.

Probably the key limitation of our study is our inability to directly determine whether changes we observed are the result of a change in gene expression in tissue resident cells or a shift in the cellular composition of the tissue. To address this concern, we used a recently published single-cell IPF RNAseq dataset (40) and computational approaches to determine the source of the signal in each of our tracks. This does not resolve the issue of potential rare cell populations but does allow us to assess the source of the signal and incorporate it in our model. Some events, such as the increase in airway basal cells or B-cell related genes in certain tracks are, of course, the result of cellular infiltrations, but such an analysis was not done so far, and the identification of early and late molecular changes may have importance regardless of the cell type they happen in. Additionally, some of the findings are probably not the result of major change in cell populations. For example, one of our most impressive novel finding, the significant molecular changes in IPF tissue with conserved histology, is probably not significantly affected by cellular

infiltration, as these were observed in samples that were histologically similar to controls without IPF. Last, it is important to note that our findings probably represent the exact gene expression in the lung, as explanted lungs were inflated, snap frozen over liquid N_2 and then carefully phenotyped without induction of post harvesting gene expression changes, whereas many of the common approaches to single cell profiling of tissue require dissociation and sometimes sorting, processes that are not yet standardized, and make the detailed imaging based quantitation of tissue fibrosis impossible (57). While clearly, analysis of carefully characterized, differentially affected tissues at the single cell resolution would be ideal, we believe that our insights, as well as our dataset with the accompanying user-friendly interactive website will have significant impact on our understanding of the progression of fibrosis within the human lung.

In this study, we provide the first dynamic regulatory model of human IPF based on mRNA and microRNA profiles of carefully characterized differentially affected microenvironments in the IPF lung. Our analysis allowed us to determine that histologically normal IPF tissue is dramatically different from lung tissue from individuals without IPF, identify molecular changes that are specific to tissue with early, progressive and end stage histological changes, and identify microRNA and transcriptional regulators of these stages. We have compared our results to previously available datasets, incorporated single cell data and validated the importance of a previously unrecognized transcription factor. Our results suggest that different regulatory mechanisms are active in minimally affected, progressively affected and end-stage tissues in IPF. This finding should drive the identification of novel biomarkers and therapeutic interventions that address the specific stages of disease in IPF.

Methods

Subject selection and study approval. Patients with IPF undergoing lung transplantation at University Hospital Leuven, Belgium were selected. Donor lungs that were not suitable for transplantation were collected to be used as "healthy" controls.

Lung processing and sample imaging. In total, we analyzed 95 samples from 10 IPF lungs and 6 donor lungs for this study. Following transplant surgery, explanted lungs were inflated to 30 cm H_2O and held at 10 cm H_2O pressure inflation while frozen over liquid nitrogen vapor according to previously established protocols (58). The frozen lung was imaged using a high-resolution CT scanner (Siemens Somatom) and cut into 2 cm thick slices along the transverse plane for sampling. Cores were systematically sampled from each slice using a 1.4 cm diameter coring drill. For this study, 2 cores were randomly selected from each of upper, mid, and lower lung regions for 6 samples per lung (Figure 1 A and B). Frozen lung samples were scanned by microCT (Skyscan 1172; Bruker, Belgium) using a cooling stage according to a previously established protocol (59). MicroCT scans were set at 40kV, 226mA, and 0.5° rotation step at -30°C with temperature maintained throughout the scan. Scans were reconstructed using NRecon software (Bruker, Belgium) and images analyzed using CTAn software (Bruker, Belgium) with a manual threshold to segment tissue from air to measure tissue percentage (tissue%) and surface area/volume (surface density, SD) from each core (Figure 1C). Tissue% and SD changes are both correlated with increased tissue from fibrosis and reduction in surface area due to loss of normal parenchymal tissues allowing these measures to quantify the extent of disease within each sample. As expected, more severely fibrotic regions were in the basal regions of the lung (supplemental Figure S3).

A portion of each core was embedded for sectioning with additional samples cut for gene expression. Sections were stained for hematoxylin and eosin, picrosirius red (total collagen), and Van Gieson (elastin). Antibodies were used to detect collagen type I (ab6308; Abcam, USA), collagen type III (2150-0100; Bio-

Rad, USA) and POU2AF1/BOB1 (ab238036; Abcam, USA). Detailed methods for histology and mRNA and microRNA labelling and processing steps are available in the supplemental data.

Single cell RNA-sequencing analysis. Using a previously published dataset (41), we examined the cell types expressing POU2AF1 using the violin plot function in Seurat (version 3.0.2) and R (version 3.5.1) in 243,472 cells from 29 healthy control lungs and 32 IPF lungs.

Differential expression analysis. Samples were divided based on principal component analysis of surface density and collagen 1 and collagen 3 volume fractions followed by expectation-maximization clustering. Three groups were defined representing early stage disease with normal lung structure (IPF1), moderate stage disease where progressive changes are occurring (IPF2), and late stage disease with severe fibrosis and end-stage pathology (IPF3). Genes were filtered and log2 transformed then differential expression calculated using linear mixed-effects models for each IPF group with a false discovery rate (FDR) adjusted $p < 0.05$ considered significant for gene expression and $p < 0.1$ for microRNA expression used in DREM analysis and $p < 0.05$ used for identifying miR target genes with greater than 2-fold increased/decreased expression. Detailed descriptions are available in the supplement.

Fibrosis progression analysis using the Dynamic Regulatory Events Miner. The dynamic regulatory event miner (DREM) (60, 61) was applied to our dataset to develop a model of regulatory tracks associated with disease progression in IPF. DREM uses a machine learning method based on graphical models to identify places in the gene expression profile where a set of genes that were previously co-expressed start to diverge. These locations denote the branching points and are associated with regulators (TFs and miRNAs) predicted to impact these events. Cell types associated with each track were determined using single cell expression data from lung tissues (40). Detailed descriptions of these methods are in the supplement.

Mouse model of lung fibrosis. C57BL/6 wild-type mice (WT) and Pou2af1$^{-/-}$ mice (KO) on the C57BL/6 background (62) were administered bleomycin or saline then sacrificed at day 14. A total of 104 mice from

3 batches were used in this study (WT-saline (WTS), n = 26; WT-bleomycin (WTB), n = 26; KO-saline (KOS), n = 23; KO-bleomycin (KOB), n = 29). Total RNA was extracted from 30-50 mg of frozen lung tissue by tissue disruption and homogenization. Gene expression was determined by real-time PCR (TaqMan) (Life Technologies, Thermo Scientific Inc. Rockford, IL, USA). β-glucuronidase was employed as an internal standard control and Collagen 1α1 primer was obtained from Life Technologies (Thermo Scientific Inc. Rockford, IL, USA). Each reaction was performed in triplicate and control reactions without RNA performed as negative control. Relative gene expression was normalized to the unstimulated control group and a linear mixed-effects model was used to compare relative expression of Col1α1 between groups using batch as the random effect. Lung hydroxyproline content was analyzed with the hydroxyproline colorimetric assay kit (BioVision Inc, Milpitas, CA). The micrograms of hydroxyproline per right lung was calculated with each assay normalized to the unstimulated control group. A linear mixed-effects model was used to compare relative hydroxyproline levels between groups using batch as the random effect. Paraffin-embedded tissue sections were stained with Masson's trichrome (collagen/connective tissue) or H&E using two slices per animal and two animals per group. Immunohistochemistry for α-smooth muscle actin (α-SMA) (ab5694; Abcam, USA) was performed with non-immune serum used for negative controls. Detailed methods are further described in the supplement.

Data sharing. All RNAseq data is available at GSE124685. The complete DREM model is available as an interactive model online (10). This website allows for detailed track analysis and data mining of differentially expressed genes and associated regulators.

Statistics. Differential expression was calculated using linear mixed-effects models for each IPF group with a false discovery rate (FDR) adjusted $p < 0.05$ considered significant for gene expression and $p < 0.1$ for microRNA expression. DREM modelling is based on input-output hidden Markov modelling to identify regulatory tracks (see supplement for additional details). For comparing groups in the mouse model, a

linear mixed-effects model was used with treatment group as the fixed effect and each batch considered as the random effect; significance was considered as $p < 0.05$.

Study Approvals. Human lungs were collected following local hospital ethical committee approval (ML6385) and informed patient consent. According to Belgian legislation, declined donor lungs can be used for research purposes. All animals included in this study were approved by the Institutional Animal Care and Use Committee (IACUC) at Yale University (2016-20084).

Author Contributions

PVB, ZBJ, JCH, BMV, NK contributed to study conception and design. JEM, FA, QL, SJ, MV, GD, AT, NT, FC, XY contributed to the acquisition of data and conducted the experiments. DT provided the transgenic mouse model. JEM, SJ, XY, RJH, DVM, JZ, JD, PVB, ZBJ, NK contributed to the analysis and writing. Samples were collected and processed by JEM, SEV, JV, KM, LDS, RV, AN, WAW. All authors reviewed and contributed comments that were incorporated into the final version of this manuscript. Study supervision was provided by NK.

Acknowledgment

The work was funded by NIH grants R01HL127349, U01HL145567, U01HL122626, and U54HG008540 to NK, U01HL137159 and R01LM012087 to PVB, and KU Leuven grant C24/15/30 to BV. JEM is a senior research fellow of the ERS (RESPIRE2-2015-9192). WW and RV are senior clinical investigators and SEV a senior research fellow of the FWO.

References

1. Kim DS, Collard HR, King TE. Classification and Natural History of the Idiopathic Interstitial Pneumonias. *Proc. Am. Thorac. Soc.* 2006;3(4):285–292.

2. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *N. Engl. J. Med.* 2018;378(19):1811–1823.

3. Lynch DA et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir. Med.* 2018;6(2):138–153.

4. Scadding JG, Hinson KF. Diffuse fibrosing alveolitis (diffuse interstitial fibrosis of the lungs). Correlation of histology at biopsy with prognosis. *Thorax* 1967;22(4):291–304.

5. Crystal RG, Bitterman PB, Rennard SI, Hance AJ, Keogh BA. Interstitial lung diseases of unknown cause. Disorders characterized by chronic inflammation of the lower respiratory tract (first of two parts). *N. Engl. J. Med.* 1984;310(3):154–166.

6. King TE et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 2014;370(22):2083–2092.

7. Richeldi L et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 2014;370(22):2071–2082.

8. Raghu G et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* 2018;198(5):e44–e68.

9. Carrington R, Jordan S, Pitchford SC, Page CP. Use of animal models in IPF research. *Pulm. Pharmacol. Ther.* 2018;51:73–78.

10. IPF regulatory iDREM model [Internet]www.sb.cs.cmu.edu/IPFReg. cited

11. McDonough JE et al. A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. *Respir. Res.* 2018;19(1):132.

12. Ashcroft T, Simpson JM, Timbrell V. Simple method of estimating severity of pulmonary fibrosis on a numerical scale.. *J. Clin. Pathol.* 1988;41(4):467–470.

13. Augsten M et al. CXCL14 is an autocrine growth factor for fibroblasts and acts as a multi-modal stimulator of prostate tumor growth. *Proc. Natl. Acad. Sci.* 2009;106(9):3414–3419.

14. Prasse A et al. A vicious circle of alveolar macrophages and fibroblasts perpetuates pulmonary fibrosis via CCL18. *Am. J. Respir. Crit. Care Med.* 2006;173(7):781–792.

15. Jia G et al. CXCL14 is a candidate biomarker for Hedgehog signalling in idiopathic pulmonary fibrosis. *Thorax* 2017;72(9):780–787.

16. Zuo F et al. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc. Natl. Acad. Sci.* 2002;99(9):6292–6297.

17. Richards TJ et al. Peripheral Blood Proteins Predict Mortality in Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 2012;185(1):67–76.

18. Schiller HB et al. Deep Proteome Profiling Reveals Common Prevalence of MZB1-positive Plasma B Cells in Human Lung and Skin Fibrosis. *Am. J. Respir. Crit. Care Med.* [published online ahead of print: June 27, 2017]; doi:10.1164/rccm.201611-2263OC

19. Englert JM et al. A role for the receptor for advanced glycation end products in idiopathic pulmonary fibrosis. *Am. J. Pathol.* 2008;172(3):583–591.

20. Yu G et al. Thyroid hormone inhibits lung fibrosis in mice by improving epithelial mitochondrial function. *Nat. Med.* 2018;24(1):39–49.

21. Vukmirovic M, Kaminski N. Impact of Transcriptomics on Our Understanding of Pulmonary Fibrosis. *Front. Med.* 2018;5:87.

22. Bauer Y et al. A Novel Genomic Signature with Translational Significance for Human Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* 2014;52(2):217–231.

23. McDonough JE et al. Gene correlation network analysis to identify regulatory factors in idiopathic pulmonary fibrosis. *Thorax* 2019;74(2):132–140.

24. Seibold MA et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N. Engl. J. Med.* 2011;364(16):1503–1512.

25. Hunninghake GM et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N. Engl. J. Med.* 2013;368(23):2192–2200.

26. Fingerlin TE et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* 2013;45(6):613–620.

27. Staab-Weijnitz CA et al. FK506-Binding Protein 10, a Potential Novel Drug Target for Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 2015;192(4):455–467.

28. Baarsma HA, Königshoff M. "WNT-er is coming": WNT signalling in chronic lung diseases. *Thorax* 2017;72(8):746–759.

29. Königshoff M et al. Functional Wnt Signaling Is Increased in Idiopathic Pulmonary Fibrosis. *PLOS ONE* 2008;3(5):e2142.

30. Ru Y et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.* 2014;42(17):e133.

31. Liu G et al. miR-21 mediates fibrogenic activation of pulmonary fibroblasts and lung fibrosis. *J. Exp. Med.* 2010;207(8):1589–1597.

32. Cushing L et al. miR-29 Is a Major Regulator of Genes Associated with Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* 2011;45(2):287–294.

33. Montgomery RL et al. MicroRNA mimicry blocks pulmonary fibrosis. *EMBO Mol. Med.* 2014;6(10):1347–1356.

34. Pandit KV et al. Inhibition and Role of let-7d in Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 2010;182(2):220–229.

35. Pottier N et al. Identification of keratinocyte growth factor as a target of microRNA-155 in lung fibroblasts: implication in epithelial-mesenchymal interactions. *PloS One* 2009;4(8):e6718.

36. Eastman Q, Grosschedl R. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr. Opin. Cell Biol.* 1999;11(2):233–240.

37. Mora AL, Bueno M, Rojas M. Mitochondria in the spotlight of aging and idiopathic pulmonary fibrosis. *J. Clin. Invest.* 2017;127(2):405–414.

38. Salajegheh A et al. Modulatory role of miR-205 in angiogenesis and progression of thyroid cancer. *J. Mol. Endocrinol.* 2015;55(3):183–196.

39. Vosgha H, Ariana A, Smith R, Lam A. miR-205 targets angiogenesis and EMT concurrently in anaplastic thyroid carcinoma. *Endocr. Relat. Cancer* [published online ahead of print: January 9, 2018]; doi:10.1530/ERC-17-0497

40. Reyfman PA et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* [published online ahead of print: December 15, 2018]; doi:10.1164/rccm.201712-2410OC

41. Adams TS et al. Single Cell RNA-seq reveals ectopic and aberrant lung resident cell populations in Idiopathic Pulmonary Fibrosis. *bioRxiv* 2019;759902.

42. Mai C et al. Thin-Section CT Features of Idiopathic Pulmonary Fibrosis Correlated with Micro-CT and Histologic Analysis. *Radiology* 2017;283(1):252–263.

43. Luzina IG et al. Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis. *Cell. Immunol.* 2018;325:1–13.

44. van Rooij E et al. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105(35):13027–13032.

45. Qin W et al. TGF-β/Smad3 Signaling Promotes Renal Fibrosis by Inhibiting miR-29. *J. Am. Soc. Nephrol. JASN* 2011;22(8):1462–1474.

46. Roderburg C et al. Micro-RNA profiling reveals a role for miR-29 in human and murine liver fibrosis. *Hepatol. Baltim. Md* 2011;53(1):209–218.

47. Cushing L, Kuang P, Lü J. The role of miR-29 in pulmonary fibrosis. *Biochem. Cell Biol.* 2014;93(2):109–118.

48. Deng Z et al. MicroRNA-29: A Crucial Player in Fibrotic Disease. *Mol. Diagn. Ther.* 2017;21(3):285–294.

49. Kim S et al. Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* 2015;16:924.

50. Steele MP et al. Relationship between gene expression and lung function in Idiopathic Interstitial Pneumonias. *BMC Genomics* 2015;16(1):869.

51. Burgy O, Königshoff M. The WNT signaling pathways in wound healing and fibrosis. *Matrix Biol.* 2018;68–69:67–80.

52. Schubart DB, Rolink A, Kosco-Vilbois MH, Botteri F, Matthias P. B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. *Nature* 1996;383(6600):538–542.

53. Zwilling S, Dieckmann A, Pfisterer P, Angel P, Wirth T. Inducible expression and phosphorylation of coactivator BOB.1/OBF.1 in T cells. *Science* 1997;277(5323):221–225.

54. François A et al. B lymphocytes and B-cell activating factor promote collagen and profibrotic markers expression by dermal fibroblasts in systemic sclerosis. *Arthritis Res. Ther.* 2013;15(5):R168.

55. Xue J et al. Plasma B lymphocyte stimulator and B cell differentiation in idiopathic pulmonary fibrosis patients. *J. Immunol. Baltim. Md 1950* 2013;191(5):2089–2095.

56. Zhou H et al. POU2AF1 Functions in the Human Airway Epithelium To Regulate Expression of Host Defense Genes. *J. Immunol. Baltim. Md 1950* 2016;196(7):3159–3167.

57. Schiller HB et al. The Human Lung Cell Atlas - A high-resolution reference map of the human lung in health and disease. *Am. J. Respir. Cell Mol. Biol.* [published online ahead of print: April 17, 2019]; doi:10.1165/rcmb.2018-0416TR

58. McDonough JE et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N. Engl. J. Med.* 2011;365(17):1567–1575.

59. Vasilescu DM et al. Nondestructive cryomicro-CT imaging enables structural and molecular analysis of human lung tissue. *J. Appl. Physiol. Bethesda Md 1985* 2017;122(1):161–169.

60. Ding J, Hagood JS, Ambalavanan N, Kaminski N, Bar-Joseph Z. iDREM: Interactive visualization of dynamic regulatory networks. *PLoS Comput. Biol.* 2018;14(3):e1006019.

61. Ding J et al. Integrating multi-omics longitudinal data to reconstruct networks underlying lung development. *Am. J. Physiol. Lung Cell. Mol. Physiol.* [published online ahead of print: August 21, 2019]; doi:10.1152/ajplung.00554.2018

62. Qin XF, Reichlin A, Luo Y, Roeder RG, Nussenzweig MC. OCA-B integrates B cell antigen receptor-, CD40L- and IL 4-mediated signals for the germinal center pathway of B cell development. *EMBO J.* 1998;17(17):5066–5075.

Figures and figure legends

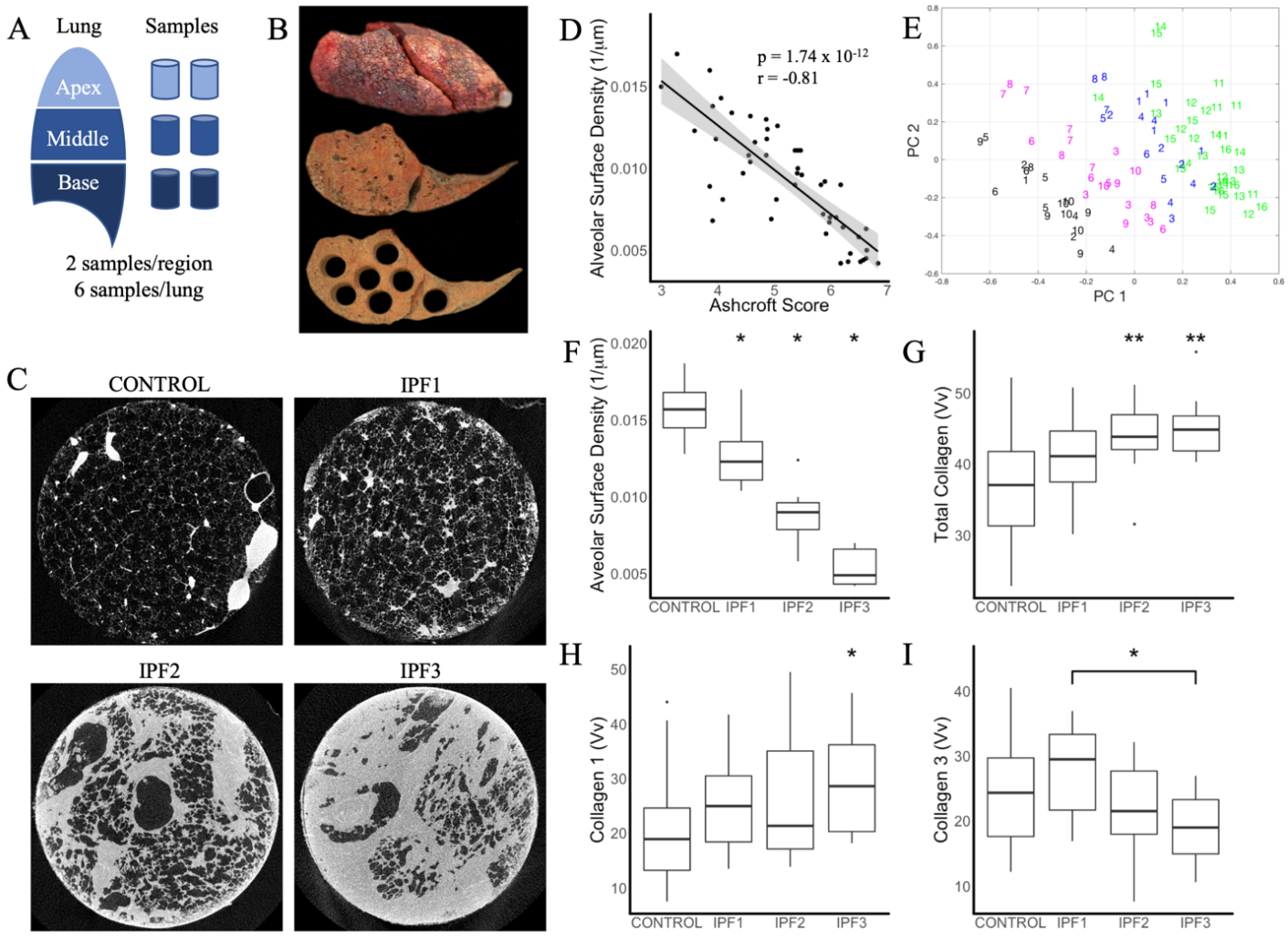

Figure 1. Sampling of the lung. (A) Schematic for sampling of the lung. (B) Image of the inflated whole IPF lung with a slice taken before and after sampling. (C) Representative microCT images of a control sample and for each stage of IPF. (D) A significant negative correlation (r = -0.81, p = $1.74x10^{-12}$) was found between alveolar surface density (ASD) compared to Ashcroft scores. Grey area represents the 95% confidence interval. (E) Principal component and expectation-maximization clustering separated samples into control (green) and early stage (IPF1; blue), progressive (IPF2; magenta) and end-stage (IPF3; black) IPF. Each subject is identified by number (IPF: 1-10; Control 11-16). (F) ASD was significantly reduced in IPF stages compared to control but a large overlap remained between IPF1 and control. (G) Volume fraction (Vv) of total collagen was significantly increased in IPF2 and IPF3 stages compared to control. (H) Collagen 1 Vv was significantly increased in IPF3 compared to controls. (I) Collagen 3 Vv showed no

difference to control but a significant decrease in IPF1 compared to IPF3. For panels F-I, data are presented as boxplots showing median and interquartile ranges for each group compared using linear mixed-effects models. * $p < 0.05$; ** $p < 0.001$

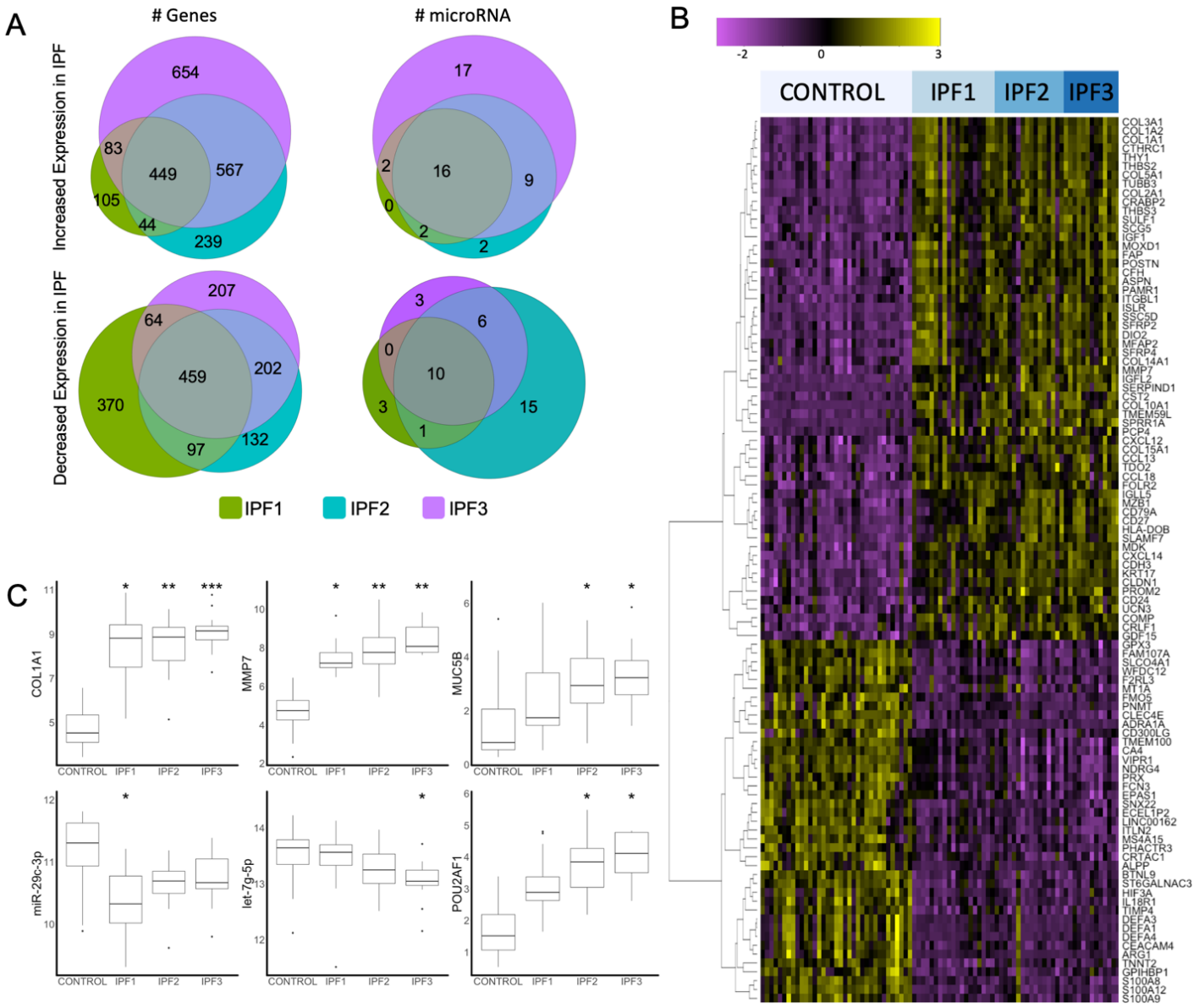

Figure 2. Expression of genes and regulators. (A) Euler diagrams showing the number of upregulated or downregulated genes or microRNAs at each stage of IPF. (B) Heatmap of the 100 genes with the greatest increased or decreased expression in all stages of IPF. (C) Boxplots for expression of key genes and regulators showing median and interquartile ranges for each group compared using linear mixed-effects models. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$ vs control.

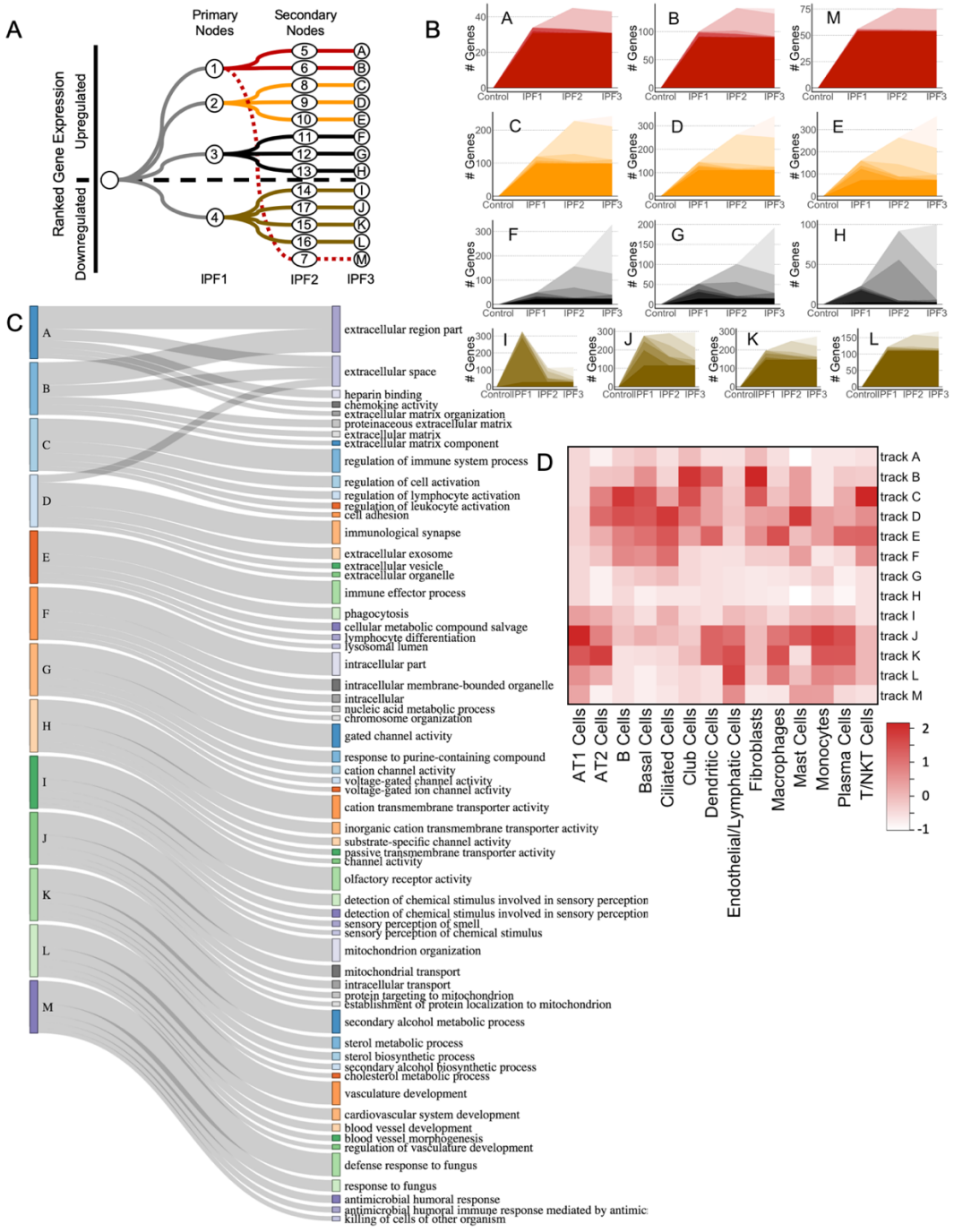

Figure 3. DREM tracks at each stage of IPF. (A) Flow diagram showing the four primary tracks established in IPF1 which split into 13 secondary tracks in the IPF2 stage. Numbers denote regulatory node for that junction where regulators and gene expression diverge. Letters denote track name. (B) Number of differentially expressed genes for each track at each stage of IPF. (C) Sankey diagram of biological function terms identified for each DREM track. (D) Heatmap of enriched genes between DREM tracks and a published IPF single cell gene marker list scaled to the number of genes per cell type (40).

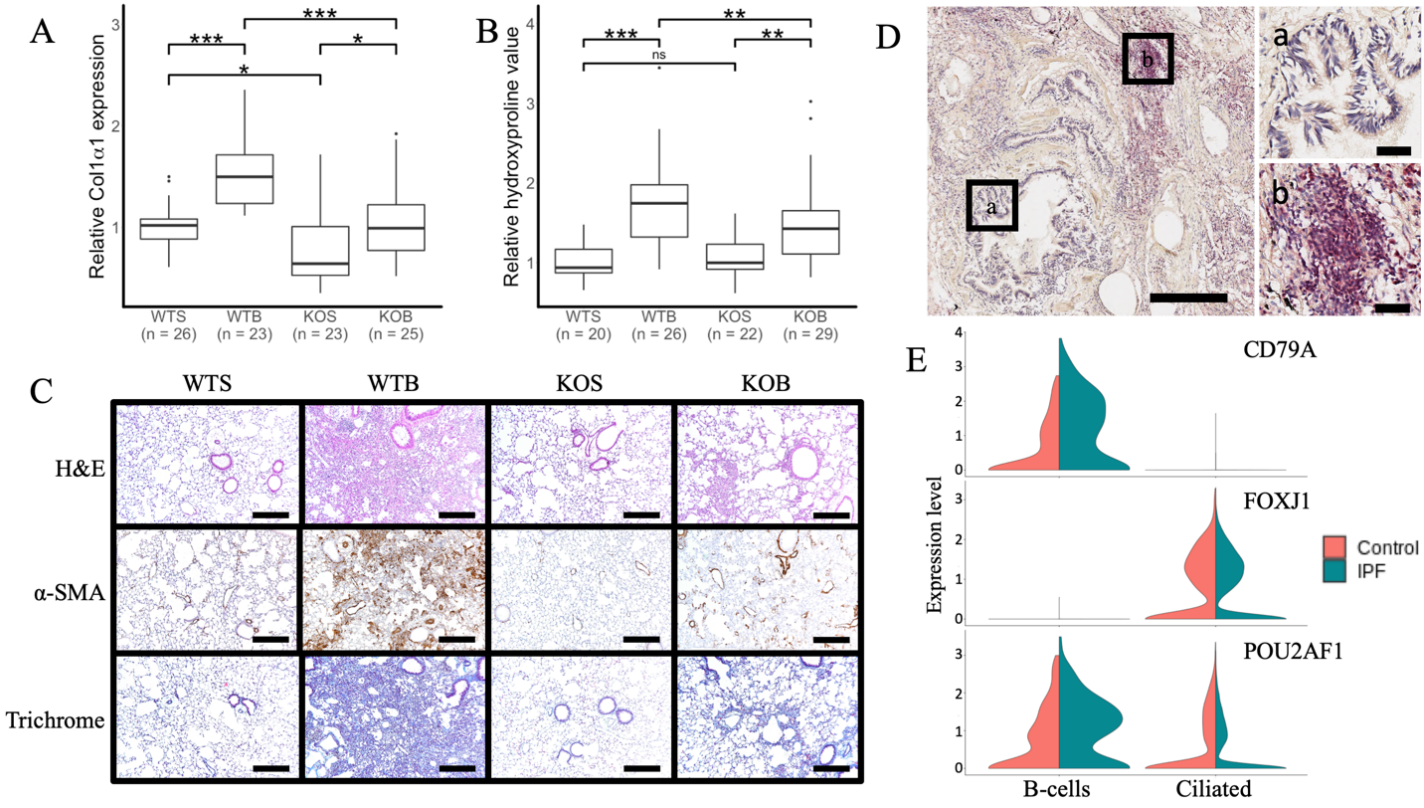

Figure 4. Pou2af1$^{-/-}$ mice show less fibrotic severity in bleomycin (Bleo)-induced lung fibrosis. (A) qRT-PCR analysis of relative change in collagen type I, α1 (Col1α1) mRNA levels in wild-type (Pou2af1$^{+/+}$) and Pou2af1$^{-/-}$ mice at day 14 after treatment with saline or bleomycin. Data are presented as boxplots showing median and interquartile ranges for each group with groups compared using linear mixed-effects models. Wild-type saline (WTS) n = 26; wild-type bleomycin (WTB) n = 23; Pou2af1 knock-out saline (KOS) n = 23; and Pou2af1 knock-out bleomycin (KOB) n = 25. (B) Relative collagen deposition assessed by hydroxyproline content per right lung in wild-type (Pou2af1$^{+/+}$) and Pou2af1$^{-/-}$ mice as indicated by treatment groups. Data are shown as boxplots with median and interquartile ranges and compared using linear mixed-effect models. WTS n = 20; WTB n =26; KOS n = 22; and KOB n = 29. (C) H&E staining (upper panels), α-SMA (middle panels) and Masson's trichrome (lower panels) staining of representative lung sections (n = 2 per group). Scale bars = 200 µm. (D) POU2AF1 immunostain (red) in human IPF lung tissue showed no staining in bronchial epithelium (a) but lymphocytic aggregates (b) were positive. Scale bars = 400 µm (left) or 50 µm (right two images). (E) Violin plots from single-cell RNA-sequencing data of control (orange) and IPF (green) lung tissues showing specific cell-types expressing POU2AF1. Upper panel shows CD79 expression identifying the B-cell population, middle panel shows FOXJ1 expression identifying the ciliated epithelial cell population. POU2AF1 (bottom panel) was

highly expressed in CD79 expressing B-cells with some expression in ciliated epithelium (FOXJ1). In IPF, POU2AF1 showed increased expression in B-cells and a slight reduction of expression in the ciliated cells. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$, ns = no significance.

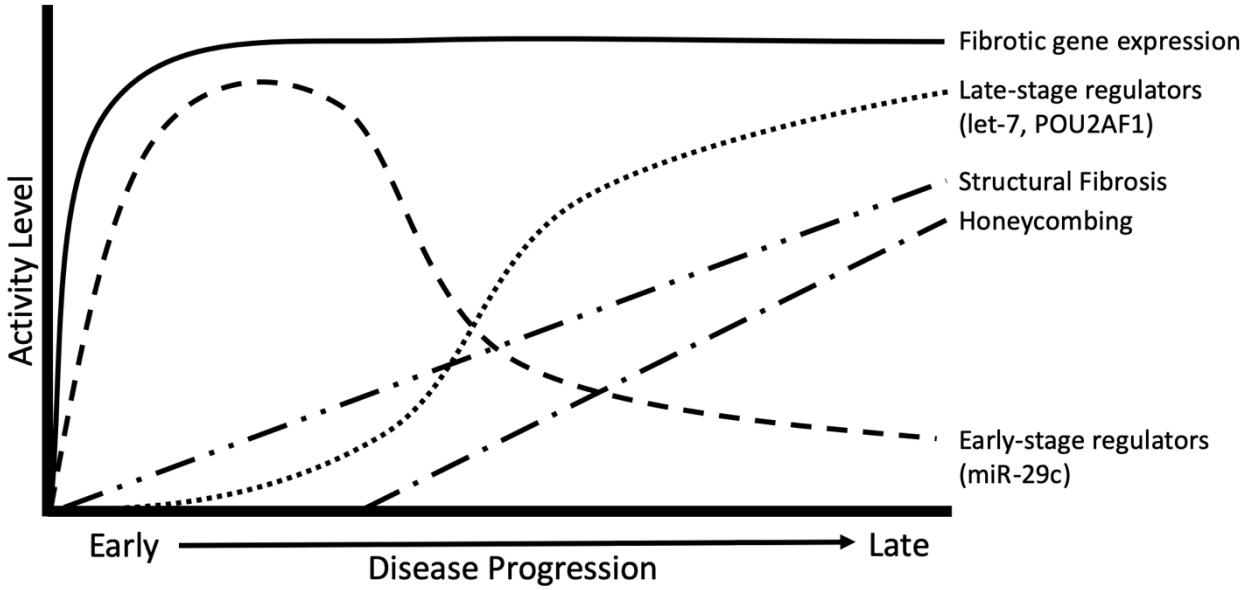

Figure 5. A model of fibrotic activity in IPF.

Tables

Table 1: Demographic data for human subjects.

		Control	IPF
N		6	10
Sex:	**female**	0	0
	Male	6	10
Age (year)		57.8 ± 10.7	57.0 ± 5.1
Height (cm)		175.8 ± 5.9	172.6 ± 6.7
Weight (kg)		83.2 ± 13.6	72.9 ± 10.3
BMI		26.8 ± 3.5	24.4 ± 2.2
Smoking History:	**non smoker (ns)**	4	0
	former smoker (fs)	1	10
	current smoker (cs)	1	0
Pack Year		fs = 24 cs= 39	22.8 ± 11.8
FEV_1 (L)		—	1.99 ± 0.51
FVC (L)		—	2.40 ± 0.77
FEV_1/FVC		—	0.85 ± 0.08
TLC (L)		—	3.6 ± 1.2
FEV_1pp (%)		—	60.7 ± 15.5
FVCpp (%)		—	58.7 ± 19.7
TLCpp (%)		—	54.9 ± 16.9
DLCOpp (%)		—	27.6 ± 7.8