Central lung gene expression associates with myofibroblast features in idiopathic pulmonary fibrosis

Yong Huang,1 Rob Guzy,2 Shwu-Fan Ma,1 Catherine A Bonham,1 Jonathan Jou,3 Jefree J Schulte,4 John S Kim,1, Andrew J Barros,1 Milena S Espindola,5 Aliya N Husain,6 Cory M Hogaboam,5 Anne I Sperling,1 Imre Noth1

ABSTRACT

Rationale Contribution of central lung tissues to pathogenesis of idiopathic pulmonary fibrosis (IPF) remains unknown.

Objective To ascertain the relationship between cell types of IPF-central and IPF-peripheral lung explants using RNA sequencing (RNA-seq) transcriptome.

Methods Biopsies of paired IPF-central and IPF-peripheral along with non-IPF lungs were selected by reviewing H&E data. Criteria for differentially expressed genes (DEG) were set at false discovery rate <5% and fold change >2. Computational cell composition deconvolution was performed. Signature scores were computed for each cell type.

Findings Comparison of central IPF versus non-IPF identified 1723 DEG (1522 upregulated and 201 downregulated). Sixty-two per cent (938/1522) of the mutually upregulated genes in central IPF were also upregulated in peripheral IPF versus non-IPF. Moreover, 85 IPF central-associated genes (CAG) were upregulated in central IPF versus both peripheral IPF and central non-IPF. IPF single-cell RNA-seq analysis revealed the highest CAG signature score in myofibroblasts and significantly correlated with a previously published activated fibroblasts signature (r=0.88, p=1.6×10−9). CAG signature scores were significantly higher in IPF than in non-IPF myofibroblasts (p=0.013). Network analysis of central-IPF genes identified a module significantly correlated with the deconvoluted proportion of myofibroblasts in central IPF and anti-correlated with inflammation foci trait in peripheral IPF. The module genes were over-represented in idiopathic pulmonary fibrosis signalling pathways.

Interpretation Gene expression in central IPF lung regions demonstrates active myofibroblast features that contributes to disease progression. Further elucidation of pathological transcriptomic state of cells in the central regions of the IPF lung that are relatively spared from morphological rearrangements may provide insights into molecular changes in the IPF progression.

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is the most severe form of interstitial lung disease and a common indication for lung transplantation.1 The commonly accepted scheme of IPF pathogenesis is centred around fibroblast and myofibroblast dysregulation in response to recurrent epithelial injuries from unknown stimuli, subsequently leading to deposition of excessive extracellular matrix.2

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Idiopathic pulmonary fibrosis (IPF) pathogenesis relies on fibroblast and myofibroblast dysregulation in response to recurrent epithelial injuries from unknown stimuli, subsequently leading to deposition of excessive extracellular matrix.

WHAT THIS STUDY ADDS

⇒ IPF central lung regions demonstrate massive dysregulated gene expression involving enhanced T-cell differentiation, IPF and wound healing signalling pathways.

H ow This Study Might Affect Research, Practice or Policy

⇒ Enhancement of myofibroblast features without apparent extracellular matrix deposition in the central region of the IPF lungs renders a window for cell-based molecular targeting therapy.
landmarks delineating central and peripheral lesions do not yet exist. The distribution of honeycombing in UIP on high-resolution CT is typically basal and peripheral. In contrast, the central portion of the pulmonary lobule shows delicate alveolar septa without significant inflammation and fibrosis. Nearly all studies of the mechanisms underlying IPF pathogenesis in human lungs have examined peripheral, scarred areas of the lung. Recent studies of gene expression in lesser fibrotic areas of IPF lungs revealed that expression of numerous immune-related, inflammation-related and ECM-related genes are altered compared with healthy control tissues.

Our study’s objective is to examine if gene expression changes in IPF central lungs, relatively spared from morphological change, play an active role in IPF peripheral lung tissue remodelling and pathogenesis. We obtained bulk RNA sequencing (RNA-seq) data from a cohort of paired central and peripheral lung explant biopsies from 13 IPF and 8 non-IPF donors. These patients with IPF were subjected to lung transplants, indicating they already have a long history of clinical course and severe disease. Using cell composition deconvolutions and publicly available single-cell RNA-seq (scRNA-seq) data, we show that IPF central tissues demonstrate enriched myofibroblast features and are involved in multiple pathways related to IPF development. Myofibroblasts have various functions, from a beneficial role in wound healing to a pathological role in ECM deposition, architectural remodelling and irreversible fibrosis. Therefore, comprehending where myofibroblasts become aberrantly activated is an essential question to understanding mechanisms of IPF pathogenesis and developing practical therapeutic approaches in the early stages.

METHODS
Patient selection
The University of Chicago Institutional Review Board reviewed and approved this study. Patients with IPF were diagnosed based on multidisciplinary reviews and published criteria. Non-IPF donor lungs were obtained from the Gift of Hope Organ and Tissue Donor Network (Ithaca, Illinois, USA).

Patient and public involvement
Patients and the public were not involved in this study’s design, recruitment and conduct.

Lung tissue collection and histological evaluation
IPF lungs were grossly examined by the time of explant, and multiple 1 cm³ biopsies were obtained from 13 paired central (IPF.C) and peripheral (IPF.P) areas (figure 1). Non-IPF donor lung biopsies of 8 paired central and peripheral obtained similarly were used as control (CTR.C and CTR.P, respectively). All the samples underwent pathological classification and scoring to quantify fibrosis (score 1=<25% of lung tissue on slide with fibrosis; score 2=25–75% fibrosis; score 3=>75% fibrosis), fibroblastic foci (greatest number in one 4× field; score 1=1–2 fibroblastic foci; score 2=3–4 fibroblastic foci; score 3=>5 fibroblastic foci), honeycombing (score 1=present; score 0=absent), inflammation (number of inflammatory cell aggregates present on entire slide) and pulmonary hypertension (score 1=mild (intimal fibrosis); score 2=moderate (intimal fibrosis with smooth muscle hypertrophy); score 3=severe (luminal narrowing)). The histopathological scored tissues were subjected to RNA extraction and sequencing. There is no systematic difference in RNA integrity number among all the lung tissues extracted.

RNA-seq library preparation and sequencing
The R/Bioconductor package ‘ComBat-seq’ package was used to correct batch effects. The counts were normalised by TPM (transcript per million). See online supplemental data for details.
Identification of differentially expressed genes

The differentially expressed genes (DEG) prioritised with false discovery rate (FDR) <5% and fold change >2 was identified using empirical Bayesian-moderated t-test implemented in R/Bioconductor package ‘limma’.12 13 P values were adjusted for multiple comparisons using the Benjamini-Hochberg method.13 Principle component analysis (PCA) was performed using R package ‘FactoMine’.14

Quantitative reverse transcription PCR

Total RNA was reverse transcribed to generate complementary DNA. Significant differences in mean values were calculated with GraphPad Prism. P value<0.05 was considered significant. See online supplemental data for detail.

Signature scoring

The Z-score of each gene was computed as $Z_i = (X_i - \mu) / \sigma$, whereas $X_i$ is the expression level of gene $i$, $\mu$ is the group mean, $\sigma$ is the SD. Signature score $= \Sigma Z_i / N$, where N is the number of cells in each cell cluster or patient. See online supplemental data for gene signatures description.

Deconvolution of cell compositions by scRNA-seq

Deconvolution of cell type compositions in IPF central and peripheral tissues was performed using R packages ‘MuSiC’ with IPF scRNA-seq reference data.8

Weighted gene co-expression network analysis

An unsupervised ‘Weighted gene co-expression network analysis (WGCNA)’ package was used to correlate co-expression modules with pathological traits and the deconvoluted cell type abundance. Gene expression data were normalised and filtered to remove redundant genes and genes with minimum variation (SD <0.66) across samples. See online supplemental data for details.

Pathway analyses

Pathway analysis was performed with Gene Set Enrichment Analysis (GSEA) algorithm by R/Bioconductor package ‘clusterProfiler’,16 which uses the entire transcriptome without fold change or FDR cut-off for a robust estimate of the functional pathways associated with the condition. Gene ontology analysis was performed using R/Bioconductor package ‘wikipathways’.17 Significant pathways of WGCNA gene module were identified using Ingenuity Pathway Analysis with the criterion of Fisher’s exact test p<0.01.

RESULTS

Clinical, macroscopic and microscopic features of biopsies

Examples of paired central and peripheral regions of IPF lungs are illustrated in figure 1A. H&E-stained paired non-IPF central (CTR.C) and non-IPF peripheral (CTR.P) sections revealed evidence of mild emphysema with no differences in histology alterations, including fibrosis and inflammation (figure 1B and D). Biopsies from IPF central (IPF.C) lungs revealed pathological abnormalities classified as UIP, non-specific interstitial pneumonitis, chronic interstitial pneumonia unclassifiable or emphysema with patchy non-specific interstitial fibrosis (figure 1C, table 1). Despite these abnormalities, IPF central had significantly lower fibrosis scores,
honeycombing, inflammation and milder grade of intimal fibrosis with pulmonary hypertension compared with paired IPF peripheral (IPF-P) samples (table 1). Notably, 11 of 13 IPF-central samples had a fibrosis score of 1 and exhibited fewer fibroblastic foci than IPF-peripheral (p=0.0059, table 1). In contrast, IPF-peripheral samples demonstrated advanced fibrosis and architectural disruption (figure 1E), and 11 out of 13 were classified as UIP with a fibrosis score of 3, and all of them had significant histological honeycombing (table 1). In the non-IPF group, only two of the eight control donors demonstrated inflammation. None of them displayed fibrosis (fibroblastic foci=0).

**IPF central tissues demonstrated massive gene dysregulations**

The DEG were identified by two group comparisons and illustrated in volcano plots (figure 2). The number of upregulated DEGs in IPF-central and IPF-peripheral samples was similar compared with non-IPF-central and non-IPF-peripheral, respectively, (figure 2A, B, 1522 and 1634, respectively). Paired comparison of IPF-central versus IPF-peripheral identified 464 upregulated and 379 downregulated genes, respectively, (figure 2C). GSEA revealed that Th1, Th2 and Th17 cell differentiation, and T-cell receptor signalling pathways were enriched in IPF-central (online supplemental figure E1, left panel). There were only three DEGs between non-IPF-peripheral and non-IPF-central (figure 2D). Detailed DEGs lists can be found in online supplemental table E1 and DEGs passed the vertical lines defined above were in bold.

**Mutually upregulated genes in IPF central and IPF peripheral tissues**

The PCA plot showed that IPF and non-IPF samples were separated at the first dimension, while IPF-central and
IPF-peripheral samples were separated at the second dimension (figure 3A). Sixty-two per cent (938/1522) of genes defined as mutually upregulated genes (MUG) were increased in both central and peripheral IPF regions when compared with corresponding non-IPF regions (figure 3B, top panel). Pathway analysis of these regionally independent MUG revealed enrichment in the pulmonary fibrosis idiopathic signalling pathway (figure 3C, online supplemental table E2). These findings suggest that IPF central tissues with absent or minimal fibrosis still undergo gene expression reprogramming in concordance with the initial transcriptomic stage of IPF pathogenesis.

Peripheral-associated genes and central-associated genes were mapped to different cell types
We identified 414 peripheral-associated genes (PAG), defined as genes uniquely increased in IPF-peripheral compared with non-IPF-peripheral and IPF-central (figure 3B, middle panel). Eighty-five central-associated genes (CAG) defined as genes upregulated in IPF-central compared with central non-IPF control and peripheral IPF; 85 central-associated genes (CAG, bottom panel) upregulated in central IPF compared with central non-IPF control and peripheral IPF. IPF.C/CTR.C: genes upregulated in IPF central compared with non-IPF control central tissues; IPF.P/CTR.P: genes upregulated in IPF peripheral compared with non-IPF control peripheral tissues; IPF.C/IPF.P: genes upregulated in IPF central compared with IPF peripheral tissues; IPF.P/IPF.C: genes upregulated in IPF peripheral compared with IPF central tissues. (C) Ingenuity Pathway Analysis of MUG with adjusted p value<0.0001. Detailed lists of MUG, PAG, CAG, and pathways enriched from MUG can be found in online supplemental table E2. CTR, non-IPF control; IPF, idiopathic pulmonary fibrosis.
respectively, (online supplemental table E3). We found that 74% of the PAG were primarily expressed in cell type cluster 1 (ciliated) and cluster 10 (differentiating ciliated) cells (online supplemental figure E2A). The combined proportion of the two ciliated cell clusters (C1 and C10) is 22.4% in the IPF but only 7.4% in the non-IPF donors. We speculate that the increased expressions of PAG may be attributed mainly to the increased honeycomb cysts with ciliated and differentiated ciliated cells in IPF peripheral regions. In contrast to PAG, CAG were mapped to diverse cell types (online supplemental figure E2B), suggesting their involvement in various cellular activities during IPF development. Detailed lists of MUG, PAG, CAG, and pathways enriched from MUG can be found in online supplemental table E2.

Reverse transcription-PCR validation of dysregulated genes in IPF lungs

A pericyte marker gene, platelet-derived growth factor receptor beta (PDGFRB) demonstrated more significant upregulation in IPF-central versus non-IPF-central compared with IPF-peripheral versus non-IPF-peripheral (figure 4A). Tenascin C gene encoding an ECM protein was upregulated in both IPF-central and IPF-peripheral compared with non-IPF (figure 4B). Smooth muscle alpha (α)–2 actin (ACTA2) was only upregulated in IPF-central versus non-IPF-central, but not in IPF-peripheral versus non-IPF-central due to larger variations (figure 4C). Collagen type I alpha 1 chain gene was mutually upregulated in both IPF-central and IPF-peripheral when

![Figure 4](http://bmjopenrespres.bmj.com/)

**Figure 4** Quantitative reverse transcription-PCR validation of dysregulated genes in IPF lungs. (A) PDGFRB displays more significant upregulation in IPF central (IPF.C) than in IPF peripheral (IPF.P) when compared with their corresponding non-IPF regions. (B) TNC is mutually upregulated in both IPF central and peripheral regions compared with the non-IPF regions. (C) ACTA2 is only upregulated in IPF central (IPF.C) compared with non-IPF control central (CTR.C). (D) COL1A1 is mutually upregulated in IPF central and peripheral regions compared with non-IPF regions. (E and F) PDGFB and CTGF are upregulated in IPF central (IPF.C) compared with both IPF peripheral (IPF.P) and non-IPF control central (CTR.C). *p<0.05, **p<0.01, ***p<0.001 and ****p<0.0001 are compared between groups as specified. ACTA2, alpha (α)–2 actin; COL1A1, collagen type I alpha 1; CTGF, connective tissue growth factor; CTR, non-IPF control; IPF, idiopathic pulmonary fibrosis; PDGFB, platelet derived growth factor subunit B; PDGFRB, platelet derived growth factor receptor beta; TNC, tenascin C gene; IPF, idiopathic pulmonary fibrosis.
compared with their corresponding non-IPF regions (figure 4D). Platelet derived growth factor subunit B and connective tissue growth factor (CTGF) were significantly upregulated in IPF-central compared with non-IPF-central and IPF-peripheral (figure 4E,F). Notably, 11.8% (111/938) of the MUG were mapped to the Extracellular Matrix (ECM)-protein knowledge database, MatriomeDB19 (online supplemental table E4), indicating IPF central underwent ECM remodelling partially common to IPF peripheral.

**CAG signature correlates with myofibroblasts in IPF**

Next, we computed CAG signature scores based on Z-scores in each cell type of IPF scRNA-seq data (GSE135893). CAG signature demonstrated the highest score in myofibroblasts (figure 5A). Concordantly, the previously published activated fibroblasts (FB) signature consisting of 49 genes retrieved from the bleomycin-induced IPF mouse model20 also displayed the highest signature score in myofibroblasts, followed by other mesenchymal cell types (online supplemental figure E3A). However, PDGFRB-high gene signature21 was preferentially expressed in smooth muscle cells rather than myofibroblasts (online supplemental figure E3B). CAG and the activated FB signature scores were significantly correlated with each other in myofibroblasts across 12 patients with IPF (figure 5B, r=0.88, p=1.6×10⁻⁴). Moreover, the CAG signature score of myofibroblasts was significantly higher in IPF than in non-IPF donors (figure 5C, p=0.01). Gene Ontology (GO) analysis revealed enrichment of CAG in biological processes regulating cell migration, mobility, chemotaxis and angiogenesis (figure 5D). We further identified 14 ECM genes from MatriomeDB in CAG, including angiogenesis genes angiopoietin 1.

In addition, the Brown module demonstrated the most significant correlation with myofibroblasts proportion foci in IPF peripheral (figure E4B), although there were no overlapping genes between the two modules. For example, ACTA1 and COL11A1 were found in the Green module, while ACTA2 and COL15A1 were present in the Brown module. Specifically, the main functions shared between the two gene modules were mainly related to myofibroblast features, including inhibition of matrix metalloproteases, pulmonary fibrosis idiopathic signaling pathway and wound healing pathway (figure E6B, online supplemental figure E4A). Network analysis revealed pro-inflammatory cytokines and ECM remodeling in the Brown module (online supplemental figure E4B).

Peripheral lymphatic endothelium correlates with multiple central cell types

Cell composition of the paired IPF central and peripheral lungs were illustrated in a correlation matrix plot. The central and peripheral cell types revealed significant correlation of peripheral lymphatic endothelium with multiple central cell types, including ciliated cells, proliferating T cells, MUC5AC+high cells, fibroblasts and plasma cells, and negatively correlated with myofibroblasts of IPF-central (online supplemental figure E5).

DISCUSSION

Our study systematically investigated the transcriptomic profiles of paired central and peripheral biopsies from IPF and non-IPF lungs. This design allowed us to perform pairwise comparisons and correlation analyses of gene co-expression networks with cell compositions and pathological alterations within and between spatial biopsies. Patients with IPF in our study are subjected to lung transplants, indicating they already have severe clinical disease. IPF-central samples displayed mild or no fibrosis and had a lesser degree of architectural distortion and pulmonary hypertension than their paired IPF-peripheral samples. We observed enhanced myofibroblast features and T-cell differentiation without apparent ECM deposition in central IPF biopsies compared with scarred peripheral IPF lungs.

IPF is a fatal lung disease manifested by scarred peripheral and basilar regions and mild or non-fibrotic central lung areas.7 The cardinal features of IPF on histology are fibrosis (ie, architectural distortion with honeycombing, traction bronchiectasis, fibroblastic foci) and its spatial and temporal heterogeneity. Therapeutic approaches targeting inflammatory, fibroblast proliferation and tissue remodelling pathways in IPF have been largely unsuccessful due to a limited understanding of the interaction between molecular profiles and cell type compositions in tissue remodelling. Numerous microarray22 23 and RNA-seq24 25 assays using IPF patient-derived lung tissue have identified differentially regulated genes and pathways. Recent scRNA-seq studies have identified functionally distinct pulmonary myofibrogenic mesenchymal cell types in patients without IPF and patients with IPF.26 27 However, a potential limitation of these studies has been their reliance on tissue samples from peripherally-scarred regions of the lung and our study attempts to address this gap.3 6

Our data suggest gene dysregulation in the central IPF lungs is driven mainly by pathological changes specific to IPF rather than spatial-related differences since 464 upregulated genes were identified in IPF-centric compared with paired IPF-peripheral and only three upregulated genes in non-IPF-centric compared with peripheral biopsies from patients with IPF.

Table 2 Fourteen extracellular matrix-specific genes identified in central associated genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Symbol</th>
<th>IPF.C/CTR.C</th>
<th>Adjusted p value</th>
<th>IPF.C/IPF.P</th>
<th>Adjusted p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANGPT1</td>
<td>2.9</td>
<td>0.02</td>
<td>2.2</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>BGN</td>
<td>4.4</td>
<td>0.01</td>
<td>2.8</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>COL4A1</td>
<td>3.2</td>
<td>0.01</td>
<td>2.0</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>LTBP2</td>
<td>7.6</td>
<td>0.00</td>
<td>2.6</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>MFAP5</td>
<td>5.1</td>
<td>0.00</td>
<td>2.2</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>NTNG2</td>
<td>2.2</td>
<td>0.04</td>
<td>2.1</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>PXDN</td>
<td>2.7</td>
<td>0.02</td>
<td>2.1</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>SEMA3A</td>
<td>4.2</td>
<td>0.00</td>
<td>2.1</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>SEMA5A</td>
<td>2.5</td>
<td>0.01</td>
<td>2.2</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>SEMA7A</td>
<td>3.4</td>
<td>0.00</td>
<td>2.1</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>SPARC</td>
<td>3.3</td>
<td>0.00</td>
<td>7.2</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>TNR</td>
<td>10.9</td>
<td>0.00</td>
<td>2.6</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>VEGFC</td>
<td>2.4</td>
<td>0.02</td>
<td>2.0</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>WNT2B</td>
<td>2.4</td>
<td>0.02</td>
<td>2.0</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

(ANGPT1), vascular endothelial growth factor C (VEGFC) and enhanced collagen type IV alpha 1 chain (COL4A1) level as markers of IPF central (table 2).
non-IPF-peripheral specimens. Inflammation is one of the significant steps leading to fibrosis. Pathway analysis revealed enrichment in Th1, Th2 and Th17 cell differentiation and T-cell receptor signalling pathways. Interleukin (IL)-12 induces the differentiation of naïve CD4 cells to Th1 cells to produce the pro-inflammatory cytokine interferon (IFN)-γ, which suppresses fibroblast-induced collagen synthesis and attenuates fibrosis. As a commonly recognised opponent of Th1 cells, Th2 cells can alter Th1-associated IFN-γ expression levels. Th17 cells are generated by exposure to transforming growth factor β1 (TGFβ1) and IL-6 and are involved in autoimmunity through their production of IL-17 family cytokines and IL-22. Therefore, an imbalance of T-cell status and differentiation in the central region may have a profound role in IPF development. The large degree of gene dysfunction and the associated pathways in IPF-central may support changes in disease involvement and progression in lieu of regional differences.

Figure 6 Correlation of IPF central gene modules with cell type compositions in central lungs and pathological changes in peripheral lungs. (A) Correlation matrix of gene modules with cell type abundance in IPF central lung and traits of peripheral lungs pathological alterations. Construction of gene modules and correlation analysis were performed using the R/Bioconductor package ‘WGCNA’. Cell type deconvolution of IPF central lung tissues was performed using the R package ‘MuSiC’. Co-expressed gene modules were depicted in unique colour bars on the left y-axis. Correlation coefficient values from −1 (green) to 1 (red) were depicted on the right y-axis. Coloured squares in the correlation matrix plot represent positive (red) or negative (green) correlation between module eigengene with pathological traits or cell type compositions. (B) Ingenuity Pathway Analysis of Green module genes anti-correlated with the trait of peripheral lung inflammation foci. The red line indicates the significance criterion at p<0.01. IPF, idiopathic pulmonary fibrosis; WGCNA, weighted gene co-expression network analysis.
lungs. Therefore, MUG may serve as common candidate molecular targets for fibrotic peripheral and mild or non-fibrotic central tissues in IPF lungs.

CAG were distributed across multiple cell types, indicating their involvement in diverse cell activities. In contrast, PAG were expressed primarily on ciliated cells due to cell cluster expansion in IPF. The top 10 GO biological processes of CAG included the regulation of chemotaxis, positive regulation of cell migration and angiogenesis. The angiogenesis genes in CAG included VEGFC and ANGPT1. VEGF expression plays an antifibrotic role in disease progression, while reduced VEGF in IPF may promote fibroproliferation. Our findings in IPF central shed light on the pathological transcriptomic state and may provide insights into molecular changes in early IPF lesions.

We further prioritised a set of ECM markers from CAG according to MatrisomeDb. Type IV collagen, an essential structural component of the basement membrane, was deposited around α-smooth muscle actin (SMA)-positive myofibroblasts in fibroblastic foci of UIP. TGFβ1 stimulation enhances type IV collagen secretion and deposition. Although LTBP2 is presented as an ECM protein, myofibroblasts in early fibrotic lesions of UIP may be implicated in refractory pathophysiology, including migration of lesion fibroblasts via a focal adhesion kinase (FAK) pathway. Latent transforming growth factor beta binding protein 2 (LTBP2) belongs to the fibrillin/LTBP superfamily and plays a positive role in lung elastogenesis. Although LTBP2 is presented as an ECM protein, myofibroblasts in non-fibroblastic foci also stain positive for LTBP2. They are thus associated with tissue remodelling and fibrogenesis. LTBP2 is inducible by TGFβ and highly upregulated in pulmonary myofibroblasts of the mouse bleomycin model, and human IPF. Clinical study further shows that LTBP2 is associated with baseline % predicted forced vital capacity (FVC) values and the prognosis of patients with IPF.

CTGF is a secreted glycoprotein produced by various cell types, including fibroblasts, myofibroblasts and vascular endothelial cells. Through these interactions with various regulators such as TGFβ, VEGF and integrin, CTGF modulates cellular responses to their environment ECM, cell motility and adhesions involved in aberrant tissue repair and fibrosis. Our reverse transcription-PCR confirmed CTGF upregulation in IPF-central compared with both IPF-peripheral and non-IPF-central tissues. In concordance with our findings, CTGF expression in transbronchial biopsy specimens of IPF lung is approximately four times higher than in patients without IPF. CTGF in plasma is elevated in patients with IPF, and the increased concentration correlates with the change in FVC. Furthermore, the CTGF monoclonal antibody, pamrevlumab, may attenuate the progression of IPF.

Collectively, our findings and prior work demonstrate the promise of dysregulated genes in central lungs as molecular candidates for targeted therapy in IPF.

Current paradigms of pulmonary fibrosis pathogenesis suggest that recurrent epithelial injuries and prolonged inflammation lead to activated myofibroblasts. Myofibroblasts play multiple roles in tissue remodelling. They exhibit enhanced mobility and contractility to promote epithelial wound closure and tissue repair. Dysregulated activation of myofibroblasts is believed to lead to the over-exuberant secretion of ECM proteins and promote pathological fibrogenesis. Here, we specifically characterised a novel status of myofibroblasts features without excessive ECM secretion and deposition. Our meta-analysis of CAG signature scores prioritised myofibroblasts as the top cell type in IPF scRNA-seq data. In concordance, WGCNA also identified Green and Brown gene modules significantly correlated with myofibroblasts within central tissues. Although there was no overlapping gene between the Green and Brown module, the functional enrichment of the two modules within central tissues was similar to that of MUG, consisting of IPF signalling, wound healing signalling and ECM remodelling pathways. Note that myofibroblasts markers ACTA1 and ACTA2 were present in the pulmonary fibrosis idiopathic signalling pathway of the Green and Brown module, respectively, further supporting the myofibroblast features in central tissues.

Currently, the mechanism for the activation of myofibroblasts without evidence of fibrosis is unclear. One could speculate that active myofibroblasts in the central regions are just representative of the disease process evident only at a ‘transcriptomic level’ since patients in our study are subjected to lung transplants, indicating they already have an end-stage disease. Therefore, the transcriptomic alterations in central IPF lung characterise a particular early-stage of IPF initiation but without further progression into scar formation. It is also possible that the differential expression and increased myofibroblast activation in central regions may represent areas of more viable fibroblast and myofibroblast cells that are viable when compared with the more ‘burnt out’ and paucicellular regions of peripheral honeycomb lung.

Our study has several limitations. Our patient cohort is limited to only 13 patients with IPF and 8 non-IPF donors. Demographic data about the cohort and other variables such as pulmonary function or other variables are not available. The explanted lungs were derived from more severe and end-staged patients. Therefore, they cannot represent the full spectrum of disease. Our study only has one female patient with IPF. Regardless, it still represents the most significant number of patients reported using pairwise transcriptomic analysis of paired central and peripheral lung biopsies. There is a lack of analysis of transcriptional profiles concerning pathological scores for different features and a more quantitative assessment of pathology to anchor transcriptional signatures. These limitations can be addressed with larger cohorts in future studies. Moreover, our data cannot clarify whether the massive gene dysregulations and the specific myofibroblasts feature in mild or non-fibrotic central lung regions indicate an early stage of fibrogenesis or cross-talk with...
peripheral regions. The relatively greater amount of honeycomb in the peripheral lung could yield a higher bronchiolar/cilia signature in PAG. While the decomposition of bulk RNA-seq data revealed a correlation of the proportion of myofibroblasts in the central lung with inflammatory foci and lymphatic endothelial cells in the peripheral lung, scRNA-seq assay of central tissues would provide invaluable molecular cell type information to illustrate the pathogenic role of central lung tissues in IPF. Staining and quantifying myofibroblasts in different regions profiled by RNA-seq would help interpret the findings.

In summary, our study demonstrates differential histopathological and transcriptomic profiles of lung tissue in IPF by central versus peripheral regions. Enhanced myofibroblast features with IPF signalling pathway indicate that the central lung regions are subjected to the primary injury associated with IPF initiation despite the non-fibrotic appearance. However, the central regions of the lungs undergo a different clinical course than the peripheral regions of the lungs in terms of lung fibrosis progression. Mild ECM deposition in central portions of the lung indicate a modifiable phase of fibrogenesis for intervention of myofibroblasts activity. Hence, the central lung-associated molecular profiles and myofibroblast activity render possible candidate targets and novel mechanisms for future antifibrotic strategies in IPF.

Author affiliations
1Division of Pulmonary & Critical Care Medicine, University of Virginia, Charlottesville, Virginia, USA
2Section of Pulmonary & Critical Care Medicine, University of Chicago, Chicago, Illinois, USA
3Department of Surgery, University of Illinois, Peoria, Illinois, USA
4Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, Wisconsin, USA
5Division of Pulmonary & Critical Care Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA
6Department of Pathology, University of Chicago, Chicago, Illinois, USA

Contributors YH, RG, S-FM and IN supervised the study. YH and RG performed the analyses. JJS and AH reviewed and scored the clinical features of the patients with idiopathic pulmonary fibrosis were prospectively enrolled in the institutional review board-approved ILD registry (IRB#14163A). Participants gave informed consent to participate in the study before taking part.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval For individuals recruited at the University of Chicago, consent was obtained from patients with idiopathic pulmonary fibrosis, were prospectively enrolled in the institutional review board-approved ILD registry (IRB#14163A).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs Catherine A Bonham http://orcid.org/0000-0001-9132-4083
John S Kim http://orcid.org/0000-0002-8887-150X

References
Central Lung Gene Expression Associates with Myofibroblast Features in Idiopathic Pulmonary Fibrosis

Yong Huang1, Robert Guzy2, Shwu-Fan Ma1, Catherine A Bonham1, Jonathan Jou3, Jefree J Schulte4, John S Kim1, Andrew J Barros1, Milena S Espindola5, Aliya N Husain2, Cory M Hogaboam5, Anne I Sperling1, Imre Noth1#

*Contributed equally to this project

#Corresponding author

Online Supplement

Supplementary Materials and Methods

RNA-Seq library preparation and sequencing. Frozen lung biopsies were homogenized and extracted with TRIzol (Invitrogen) following manufacturer's protocol. RNA was re-precipitated by sodium acetate/ethanol to remove residual organic solvents. RNA quality and integrity were confirmed by Nanodrop (A260/A280 ratios between 1.7 and 2.2) and Bio-Analyzer mini-gel assay (Agilent, Santa Clara, CA). All RNA samples displayed a RNA Integrity Number (RIN) > 7 were proceeded to cDNA library preparation at the Genomics Core Facility of the University of Chicago (https://fgf.uchicago.edu/). There was no systematic difference of RIN between central and peripheral lung tissues. Total RNA in the amount of 1μg per sample was depleted of ribosomal RNA using the Ribo-Zero kit (Epicentre, Madison, WI). The directional (first strand) cDNA libraries were
prepared following the guide of TruSeq Stranded Total RNA Sample Preparation kit. RNA was fragmented at 94°C for 6 min, followed by the first strand cDNA generation. Deoxy-UTP was incorporated in second strand synthesis in order to effectively quench the second strand during PCR amplification. After adenylation of the 3' end and ligation of adapters, fragments were selected and enriched with 10 cycles of PCR amplification. Clusters were generated by bridge amplification within paired-end flow cells using Illumina HiSeq PE Cluster Kit v4 cBot according to manufacturer's instructions (Illumina, San Diego, CA). The clusters on flow cells were then sequenced on the Illumina HiSeq4000 using TrueSeq SBS Kit. A total of 780Gb reads were generated for cDNA libraries prepared from 42 samples using high output mode of 100bp paired-end (PE) sequencing. Around 94% sequences passed quality checked (> Q30), yielding ~87 M passing filter clusters per sample, and 3.7G clusters in total. Raw sequencing data in fastq format were processed using RNA-seq aligner STAR v2.5.3 (E1). GenCode GRCh38 release 24 was used for sequence alignment and transcriptome annotation. Batch effect associated with different library preparations was corrected using R/Bioconductor package ‘Comba-seq’ package (E2). The counts data were normalization by TPM (transcript per million) (E3, E4). The raw reads data in fastq format have been deposited in Sequence Read Archive (SRA) database. The reads count and normalized data have been deposited in NCBI Gene Expression Omnibus (GEO) database (Accession GSE199949).

**Quantitative reverse transcription PCR.** cDNA was made from 1ug of total RNA using BioRad iScript Reverse Transcription Supermix (170-8841, BioRad, Hercules, CA). Quantitative RT-PCR was performed on a Roche LightCycler 480 II using Taqman® gene expression assays (Thermo Fisher Scientific). All samples were normalized to Gapdh and then scaled relative to non-IPF controls using the standard delta-Ct (ΔCt) method. Data is reported as $2^{-\Delta\Delta C_t}$. Individual data points presented in graphs represent biological replicates (independent biopsies). Significant differences in mean values were calculated with GraphPad Prism using unpaired Student’s paired or unpaired t-tests. A $p$-value of less than 0.05 was considered to be significant.

**Tissue processing for generation of the scRNA-seq library.** IPF single cell data was downloaded from GSE135893 (E5). Briefly, biopsies of multiple regions of lung tissues
were digested in an enzymatic cocktail consisting of collagenase I/dispase II (1 ug/ml) tissue or Miltenyi Multi-Tissue Dissociation Kit according to the manufacturer’s protocol. A gentleMACS Octo Dissociator (Miltenyi Inc.) was used to dissociate lungs. Tissue lysates were serially filtered through sterile gauze, 100- and 40-um sterile filters (Fischer). Single-cell suspensions then underwent cell sorting using serial columns (Miltenyi Microbeads, CD235a and CD45) at Vanderbilt University Medical Center or fluorescence-activated cell sorting at the Translational Genomics Research Institute. CD45− and C45+ populations were mixed 2:1 in samples processed at VUMC and used as input for generating scRNA-seq libraries. Only the cells derived from the whole lungs of IPF patients and donors were included in the current study.

**Fibroblasts gene signatures derived from bleomycin-induced pulmonary fibrosis model in mouse.** Peyser and coauthors published a 49-gene signature defining the Activated sub-population of fibroblasts in lung fibrosis. Lung cells are isolated at Day 11 after bleomycin induction of lung fibrosis in mice. Gene expression profiles were analyzed by three complementary techniques including single cell RNAseq (scRNA-seq) to generate the gene signature (E6). Xie and coauthors analyzed mesenchymal preparations of lung fibrosis and identified a novel cell cluster uniquely present in bleomycin-treated mice. This cell cluster can be characterized by 38 genes including *Pdgfrb*, termed as Pdgfrb-high cell cluster (E7). We mapped these signature genes to human homolog and computed their z-scores in each cell type of a published human IPF scRNA-seq data GSE135893 (E5).

**Weighted gene co-expression network analysis.** Gene expression data were normalized and filtered to remove redundant genes and genes with minimum variation (standard deviation <0.66) across samples. An unsupervised “Weighted gene co-expression network analysis (WGCNA)” package (E8) was used to correlate co-expression modules with pathological traits and the deconvoluted cell type abundance.
References


Figure E1. GSEA pathway analysis of global gene expression profile between IPF peripheral and IPF central lung tissues. Top 10 significant KEGG pathways in peripheral (activated) or central (suppressed) were identified using the R/Bioconductor package “clusterProfiler”.
Figure E2. Dot plots of IPF regional lung-specific genes across cell types in IPF scRNA-seq data. A. 414 Peripheral-Associated Genes (PAG); B. 85 Central-Associated Genes (CAG); C. Annotation of each cell cluster type corresponding to cluster number in the order from top to bottom of the Y-axis. The X-axis is the gene symbol in PAG denoted by ticks or in CAG denoted by gene names. The cell numbers and percentages of the different cell types obtained from normal or fibrotic tissues of single cell data downloaded from GSE135893 is listed in Table E3.

<table>
<thead>
<tr>
<th>Cell cluster# from top to bottom</th>
<th>Cell cluster type</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Smooth Muscle Cells</td>
</tr>
<tr>
<td>28</td>
<td>Mesothelial Cells</td>
</tr>
<tr>
<td>16</td>
<td>Myofibroblasts</td>
</tr>
<tr>
<td>27</td>
<td>HAS1 High Fibroblasts</td>
</tr>
<tr>
<td>26</td>
<td>Fibroblasts</td>
</tr>
<tr>
<td>12</td>
<td>PLIN2+ Fibroblasts</td>
</tr>
<tr>
<td>0</td>
<td>Macrophages</td>
</tr>
<tr>
<td>4</td>
<td>T Cells</td>
</tr>
<tr>
<td>5</td>
<td>Monocytes</td>
</tr>
<tr>
<td>19</td>
<td>B Cells</td>
</tr>
<tr>
<td>22</td>
<td>Plasma Cells</td>
</tr>
<tr>
<td>20</td>
<td>Mast Cells</td>
</tr>
<tr>
<td>13</td>
<td>cDCs</td>
</tr>
<tr>
<td>29</td>
<td>pDCs</td>
</tr>
<tr>
<td>9</td>
<td>NK Cells</td>
</tr>
<tr>
<td>23</td>
<td>Proliferating T Cells</td>
</tr>
<tr>
<td>15</td>
<td>Proliferating Macrophages</td>
</tr>
<tr>
<td>3</td>
<td>Endothelial Cells</td>
</tr>
<tr>
<td>17</td>
<td>Lymphatic Endothelial Cells</td>
</tr>
<tr>
<td>14</td>
<td>Transitional AT2</td>
</tr>
<tr>
<td>25</td>
<td>Proliferating Epithelial Cells</td>
</tr>
<tr>
<td>11</td>
<td>cDCs</td>
</tr>
<tr>
<td>30</td>
<td>Macrophages</td>
</tr>
</tbody>
</table>

Cell cluster annotation:
- **PAG** (Peripheral-Associated Genes)
- **CAG** (Central-Associated Genes)
Figure E3. Fibroblasts signature genes from mouse bleomycin model were assessed and signature scores computed across human IPF cell types reported in scRNA-seq data GSE135893. A. Signature scores of the 49 genes characterizing Activated Fibroblasts cluster. B. Signature scores of the 38 genes characterizing PDGFRB-high Fibroblasts.
Figure E4. Ingenuity Pathway Analysis of the Brown gene module correlated with myofibroblasts. Construction of gene co-expression modules and correlation analysis were performed using R/Bioconductor package “WGCNA”. Cell type deconvolution of central lung tissues were performed using R package “MuSiC”. Brown module demonstrated significant correlation with myofibroblasts in IPF.C (Figure 4A). A. Significant pathways enriched with Brown module genes with adjusted p-value < 0.01. B. Gene-gene interaction network of Brown module genes.
Figure E5. Correlation matrix of cell type compositions derived from deconvolution. Cell type compositions were computed using R packages “MuScic”. Human IPF lung scRNA.seq data GSE135893 were used as a reference for cell type deconvolution. "_P" or "_C" represent cell types in peripheral or central lungs, respectively. Blue box and * highlight cell types correlated and anti-correlated with Lymphatic endothelial cells in peripheral lung, respectively.