



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

To cite: Huang Y, Guzy R, Ma S-F, *et al*. Central lung gene expression associates with myofibroblast features in idiopathic pulmonary fibrosis. *BMJ Open Res* 2023;**10**:e001391. doi:10.1136/bmjresp-2022-001391

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2022-001391>).

YH, RG and S-FM contributed equally.

Received 4 August 2022
Accepted 13 January 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Imre Noth;
IN2C@hscmail.mcc.virginia.edu

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 genes 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 ($r=0.88$, $p=1.6 \times 10^{-4}$). 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

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.

HOW 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.

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 the deposition of excessive extracellular matrix (ECM).² The mechanism underlying aberrant myofibroblast activation and matrix production remains elusive.

The main histological pattern of IPF is usual interstitial pneumonia (UIP), characterised by dense regions of scarring separated by areas of relatively preserved lung architecture, fibroblastic foci, clusters of inflammatory cells and honeycombing.³ Central portions of the lungs are typically spared from fibrosis,⁴ while immediate subpleural parenchyma shows advanced scarring and microscopic honeycombing.⁵ Peripheral pulmonary lesions are generally found in the peripheral one-third of the lung, although a consensus definition and radiological anatomical

landmarks delineating central and peripheral lesions do not yet exist.⁵ The distribution of honeycombing in UIP on high-resolution CT is typically basal and peripheral.⁶ In contrast, the central portion of the pulmonary lobule shows delicate alveolar septa without significant inflammation and fibrosis.⁷ Nearly all studies of the mechanisms underlying IPF pathogenesis in human lungs have examined peripheral, scarred areas of the lung.⁸ 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.^{4,9,10}

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.¹¹ 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.¹ 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

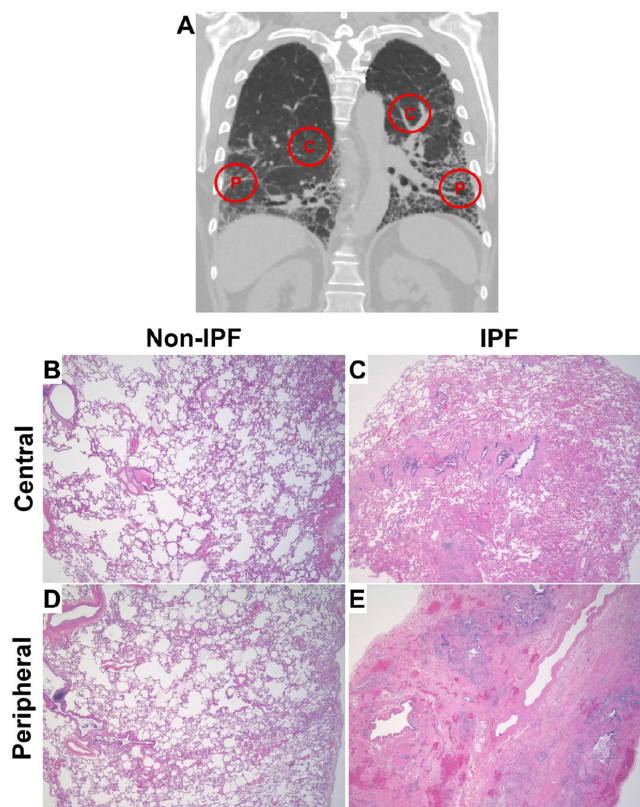


Figure 1 Spatial differences and features of IPF lung biopsies. (A) Representative CT image of IPF lungs illustrating the locations of the central (C) and peripheral (P) tissues, respectively. (B–E) H&E staining of central and peripheral biopsies derived from non-IPF (both with normal lung architecture) and IPF lungs (central section with mild or no focal fibrosis and peripheral with rearrangement scarring, microscopic honeycombing and fibroblastic foci). IPF, idiopathic pulmonary fibrosis.

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 , μ is the group mean, σ is the SD. Signature score = $\sum(Z_i) / N$, 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'¹⁵ 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 'rwikipathways'.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

Table 1 Histological scoring of central and peripheral IPF lung explant biopsies

	Central (N=13)	Peripheral (N=13)	Paired t-test p value
*Pathological classification			
UIP pathology	3 (patchy)	13 (subpleural)	
NSIP	2	0	
UIP/NSIP overlap	2	0	
CIPU	5	0	
E/PIF	1	0	
Fibrosis score (average±SD)	1.15±0.375	2.85±0.375	<0.0001
1	11	0	
2	2	2	
3	0	11	
Fibroblastic foci (average±SD)	1.08±0.862	2.08±0.862	0.0059
0	3	0	
1	7	4	
2	2	4	
3	1	5	
Honeycomb (average±SD)	0.31±0.48	1±0	0.0002
0	9	0	
1	4	13	
Inflammation foci (average±SD)	6.46±3.971	12.08±6.601	0.0103
1–7	7	3	
8–14	6	5	
15–21	0	5	
Pulmonary hypertension grade (average±SD)	1.69±0.63	2.69±0.48	0.0019
1	5	0	
2	7	4	
3	1	9	
CIPU, chronic interstitial pneumonitis, unclassifiable; E, emphysema; IPF, idiopathic pulmonary fibrosis; IPF.C, IPF lung explants of the central regions; IPF.P, IPF lung explants of the peripheral regions; NSIP, non-specific interstitial pneumonitis; PIF, patchy non-specific interstitial fibrosis; UIP, usual interstitial pneumonia.			

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,

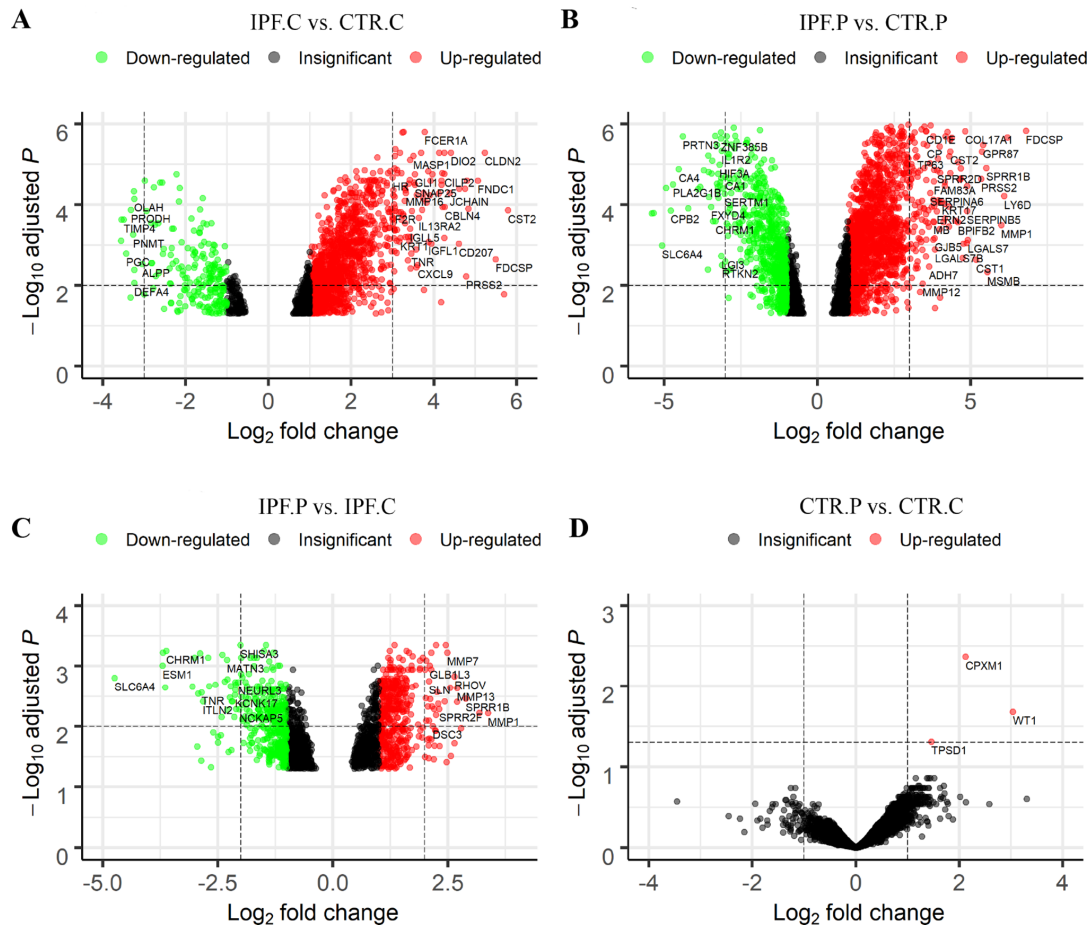


Figure 2 Volcano plot of differentially expressed genes (DEGs) by two-group comparison. (A) IPF central (IPF.C) versus non-IPF central (CTR.C); 1522 genes were upregulated and 201 were downregulated; (B) IPF peripheral (IPF.P) versus non-IPF peripheral (CTR.P); 1634 gene were upregulated and 686 were downregulated; (C) IPF peripheral (IPF.P) versus IPF central (IPF.C); 464 gene were upregulated and 379 were downregulated; (D) non-IPF peripheral (CTR.P) versus non-IPF central (CTR.C); only three upregulated genes were identified. Red and green dots represent DEGs. The horizontal dash lines represent FDR <1% in A–C and FDR <5% in D. The vertical dash lines represent FC >8 in A and B; FC >4 in C, and FC >2 in D. The full gene lists of upregulated (red font) and downregulated (green font) and statistical details of two-group comparison (A–C). Detailed DEGs lists can be found in online supplemental table E1 and DEGs passed the vertical lines defined above were in bold. CTR, non-IPF control; FC, fold change; FDR, false discovery rate; IPF, idiopathic pulmonary fibrosis.

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-peripheral versus IPF-central 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

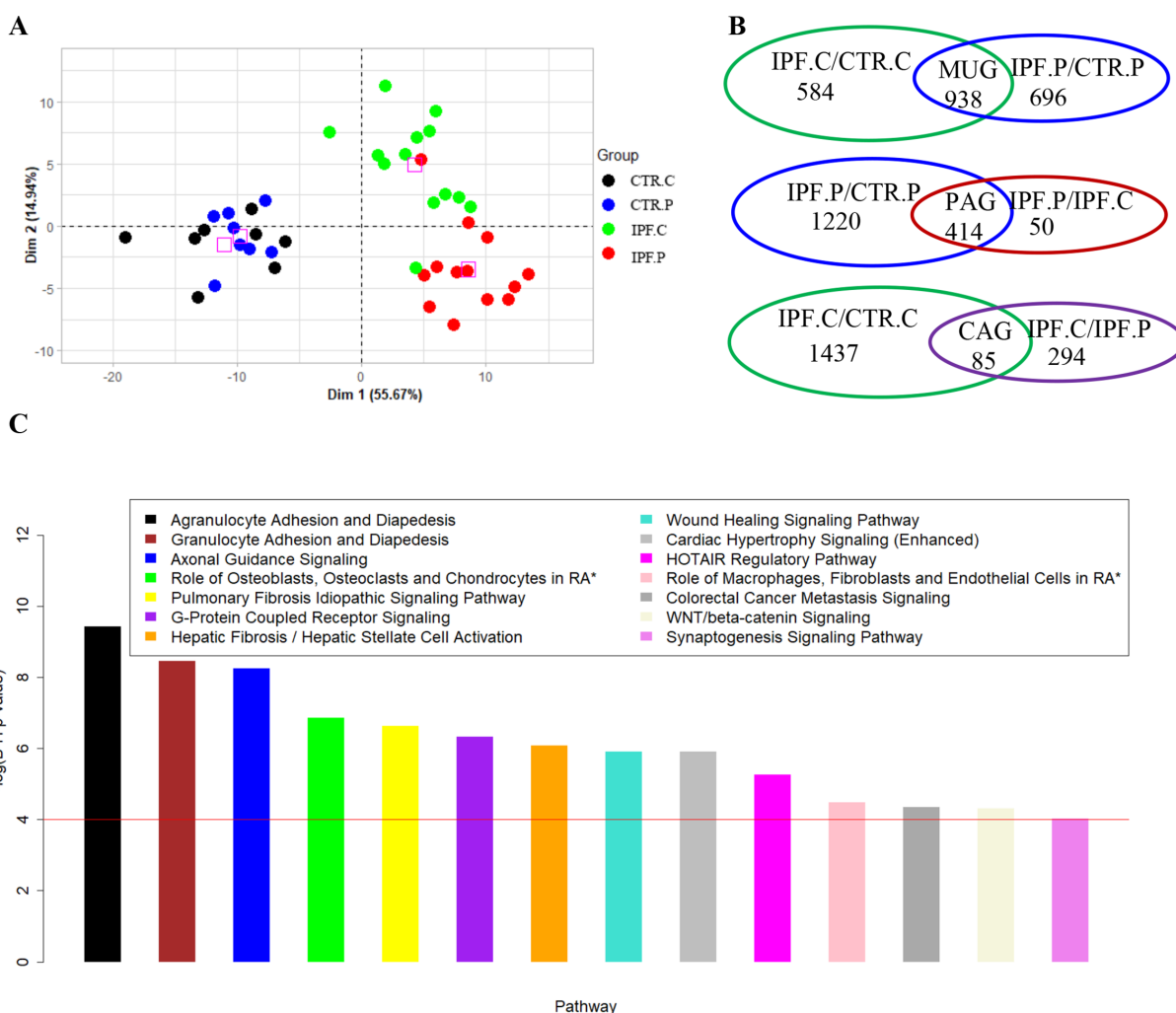


Figure 3 Regional specific and consensus upregulated genes. (A) Principle component analysis plot separates all samples into three dimensions: IPF-peripheral (IPF.P, red) group, IPF-central (IPF.C, green) group, non-IPF control peripheral (CTR.P, blue) and non-IPF control-central (CTR.C, black) group. (B) Venn diagram illustrates 938 mutually upregulated genes (MUG, top panel) increased in both central and peripheral IPF lungs compared with non-IPF control lungs; 414 peripheral-associated genes (PAG, middle panel) upregulated in peripheral IPF compared with peripheral non-IPF control and central 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.

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 non-IPF-central and IPF-peripheral (figure 3B, lower panel). We mapped PAG into IPF scRNA-seq data set GSE135893,⁸ in which cells derived from biopsies of multiple regions of IPF lung tissues. Thirty-one cell clusters were identified and summarised for cell numbers and percentages in IPF and control,

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.¹⁸ 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

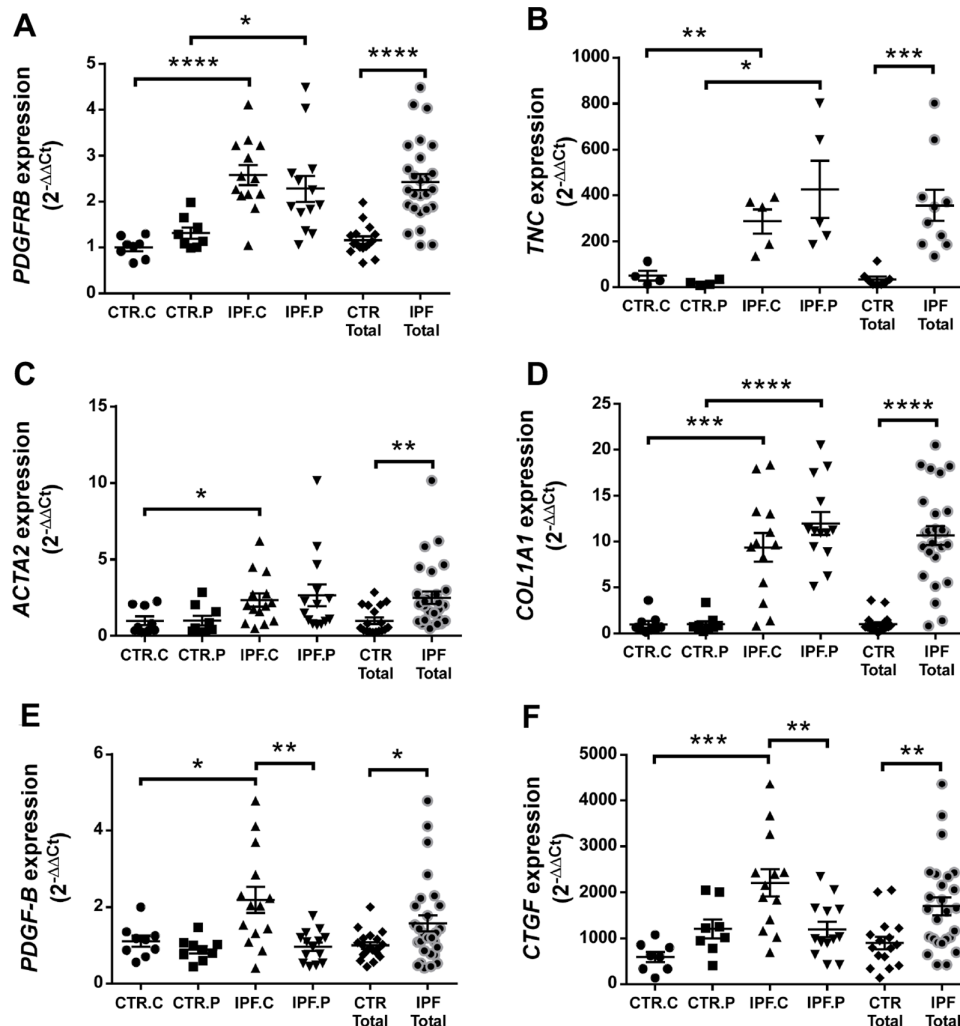


Figure 4 Quantitative reverse transcription-PCR validation of dysregulated genes in IPF lungs. (A) *PDGFRB* displays more significant upregulation in IPF central (IPF.C) than in IPF peripheral (IPF.P) when compared with their corresponding non-IPF regions. (B) *TNC* is mutually upregulated in both IPF 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) *PDGF-B* 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; *PDGF-B*, platelet derived growth factor subunit B; *PDGFRB*, platelet derived growth factor receptor beta; *TNC*, tenascin C gene; IPF, idiopathic pulmonary fibrosis.

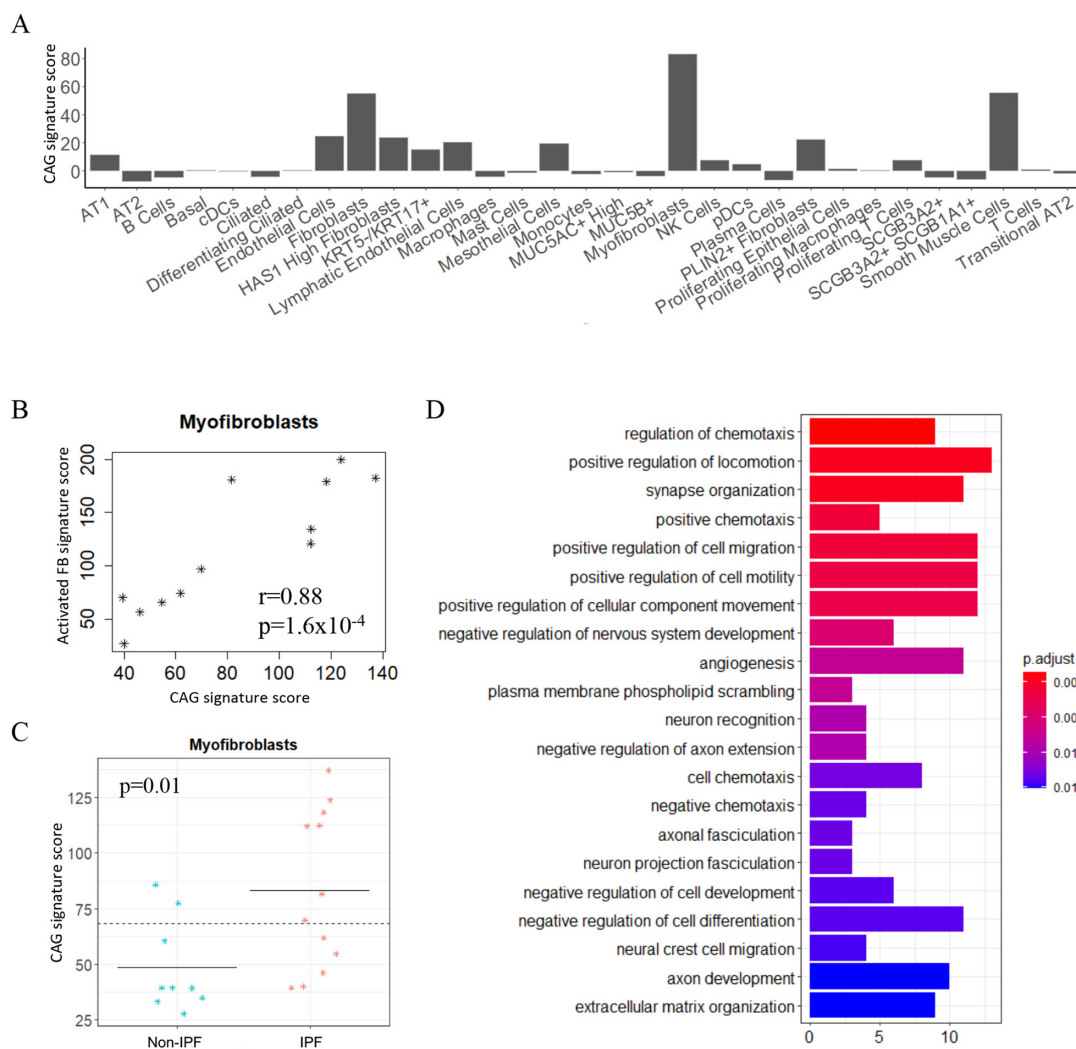


Figure 5 Central-associated genes (CAG) signature characterises myofibroblasts in IPF. CAG and activated fibroblasts signature genes were mapped to patients with IPF and non-IPF donors of single-cell RNA sequencing data GSE135893. (A) Z-scores of the 85 CAG were summed in each cell and then averaged in each cell type of patient with IPF as a signature score. (B) Correlation of CAG and activated fibroblasts signature scores in myofibroblasts of 12 patients with IPF. (C) Comparison of CAG signature scores between IPF and donor's myofibroblasts. (D) GO biological process significantly enriched with CAG. FB, fibroblasts; GO, Gene Ontology; 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, MatrisomeDB¹⁹ (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, $r=0.88$, $p=1.6 \times 10^{-4}$). 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 MatrisomeDB in CAG, including angiogenesis genes angiopoietin 1

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

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

ANGPT1, angiotensinogen 1; *BGN*, biglycan; *COL4A1*, collagen type IV alpha 1 chain; CTR, Non-IPF control; CTR.C, Non-IPF control central; IPF.C, idiopathic pulmonary fibrosis central; *LTBP2*, latent transforming growth factor beta binding protein 2; *MFAP5*, microfibril associated protein 5; *NTNG2*, Netrin G2; *PXDN*, peroxidase; *SEMA3A*, semaphorin 3A; *SEMA5A*, semaphorin 5A; *SEMA7A*, semaphorin 7A (John Milton Hagen blood group); *SPARC*, secreted protein acidic and cysteine rich; *TNR*, tenascin R; *VEGFC*, vascular endothelial growth factor C; *WNT2B*, Wnt family member 2B.

(*ANGPT1*), vascular endothelial growth factor C (*VEGFC*) and enhanced collagen type IV alpha 1 chain (*COL4A1*) level as markers of IPF central (table 2).

Gene co-expression module associates central myofibroblasts with peripheral inflammation foci

The Green module was positively correlated with myofibroblasts abundance ($r=0.8$, $p=0.001$) and two epithelial cell type (SCGB3A2 hi, KRT5 lo/KRT17 hi) proportions in IPF central, but negatively correlated with inflammation foci in IPF peripheral (figure 6A, $r=-0.73$, $p=0.004$). In addition, the Brown module demonstrated the most significant correlation with myofibroblasts proportion (figure 6A, $r=0.91$, $p=10^{-5}$). Accordingly, pathway analysis of Green and Brown module genes demonstrated similar functional profiles (figure 6B), 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 6B, 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 microarray^{22 23} and RNA-seq^{24 25} assays using IPF patient-derived lung tissue have identified differentially regulated genes and pathways. Recent scRNA-seq studies have identified functionally distinct pulmonary myofibroblastic 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.^{5 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-central compared with paired IPF-peripheral and only three upregulated genes in non-IPF-central compared with

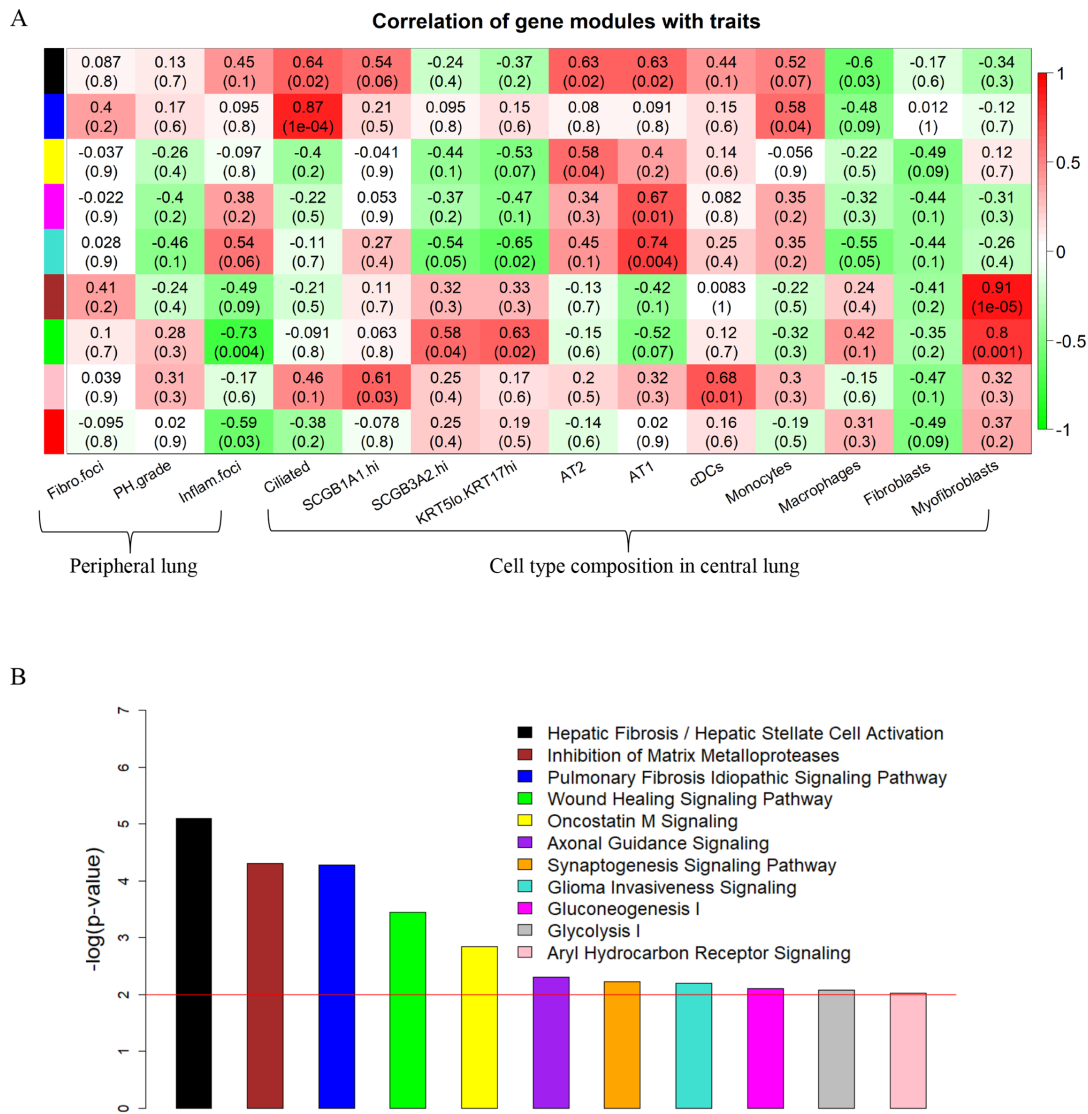


Figure 6 Correlation of IPF central gene modules with cell type compositions in central lungs and pathological changes in peripheral lungs. (A) Correlation matrix of gene modules with cell type abundance in IPF central lung and traits of peripheral lungs pathological alterations. Construction of gene 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.

non-IPF-peripheral specimens. Inflammation is one of the significant steps leading to fibrosis.¹⁶ 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 ,²⁸ which suppresses fibroblast-induced collagen synthesis and attenuates fibrosis.²⁹ As a commonly recognised opponent of Th1 cells, Th2 cells can alter Th1-associated IFN- γ expression levels.³⁰ Th17 cells are generated by exposure to transforming growth factor beta 1 (TGFB1) and IL-6 and are involved in

autoimmunity through their production of IL-17 family cytokines and IL-22.³¹ 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.

Notably, enriched pathways from MUG, including IPF signalling, hepatic fibrosis and wound healing signalling, suggest that IPF-central lung tissues may actively participate in IPF pathogenesis by reprogramming their transcriptome to a current status shared by fibrotic peripheral



lung tissues. 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.³² 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.¹⁹ Type IV collagen, an essential structural component of the basement membrane,³³ was deposited around α -smooth muscle actin (SMA)-positive myofibroblasts in fibroblastic foci of UIP. TGF β 1 stimulation enhances type IV collagen together with increased expression of α -SMA. Deposition of COL4A1 produced by 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.³⁴ Latent transforming growth factor beta binding protein 2 (LTBP2) belongs to the fibrillin/LTBP superfamily and plays a positive role in lung elastinogenesis.³⁵ 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 TGF β and highly upregulated in pulmonary myofibroblasts of the mouse bleomycin model, and human IPF.³⁶ Clinical study further shows that LTBP2 is associated with baseline % predicted forced vital capacity (FVC) values and the prognosis of patients with IPF.³⁶

CTGF is a secreted glycoprotein produced by various cell types, including fibroblasts, myofibroblasts and vascular endothelial cells.³⁷ Through these interactions with various regulators such as TGF β , VEGF and integrin, CTGF modulates cellular responses to their environment ECM, cell motility and adhesions involved in aberrant tissue repair and fibrosis.³⁸ 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.³⁹ 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.⁴⁰ 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 biospecimens. YH, RG, S-FM, JJ, JSK, AB, CB, MSE, CH, AIS and IN interpret data. RG, CB, S-FM, JJ and IN participated in the sample and data collection. AIS Laboratory acquired and banked non-idiopathic pulmonary fibrosis donor lungs. YH, RG, S-FM and IN wrote the draft of the manuscript. All authors participated in the critical revision and final approval of the manuscript. IN accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding This study was supported by National Heart, Lung and Blood Institute; R01HL130796 and UG3HL145266 (IN); K08HL125910 (RG); K23HL143135 (CB); K23HL150301 (JSK); 5R01AI125644 (AIS).

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, consenting 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.

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

- Raghu G, Collard HR, Egan JJ, *et al*. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788–824.
- Sgalla G, Iovene B, Calvello M, *et al*. Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir Res* 2018;19:32.
- Ahluwalia N, Shea BS, Tager AM. New therapeutic targets in idiopathic pulmonary fibrosis: aiming to rein in runaway wound-healing responses. *Am J Respir Crit Care Med* 2014;190:867–78.
- Todd NW, Galvin JR, Sachdeva A, *et al*. Microscopic organizing pneumonia and cellular non-specific interstitial pneumonia are widespread in macroscopically normal-appearing lung tissue in idiopathic pulmonary fibrosis. *J Heart Lung Transplant* 2016;35:S1053-2498(16)30254-6:1367–70..
- Touman AA, Vitsas VV, Koulouris NG, *et al*. Gaining access to the periphery of the lung: bronchoscopic and transthoracic approaches. *Ann Thorac Med* 2017;12:162–70.
- Devaraj A. Imaging: how to recognise idiopathic pulmonary fibrosis. *Eur Respir Rev* 2014;23:215–9.
- Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. *Annu Rev Pathol* 2014;9:157–79.
- Habermann AC, Gutierrez AJ, Bui LT, *et al*. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020;6:eaba1972.
- Luzina IG, Salcedo MV, Rojas-Peña ML, *et al*. Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis. *Cell Immunol* 2018;325:S0008-8749(18)30002-9:1–13..
- McDonough JE, Ahangari F, Li Q, *et al*. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* 2019;4:e131597:22..
- Moore MW, Herzog EL. Regulation and relevance of myofibroblast responses in idiopathic pulmonary fibrosis. *Curr Pathobiol Rep* 2013;1:199–208.
- Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995;57:289–300. 10.1111/j.2517-6161.1995.tb02031.x Available: <http://doi.wiley.com/10.1111/rssb.1995.57.issue-1>
- Husson F, Le S, Pagès J. *Exploratory multivariate analysis by example using rebook*. 2017.
- Wang X, Park J, Susztak K, *et al*. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 2019;10:380.
- Wu T, Hu E, Xu S, *et al*. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation* 2021;2:100141.
- Slenter DN, Kutmon M, Hanspers K, *et al*. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* 2018;46:D661–7.
- Jonsdottir HR, Arason AJ, Palsson R, *et al*. Basal cells of the human airways acquire mesenchymal traits in idiopathic pulmonary fibrosis and in culture. *Lab Invest* 2015;95:1418–28.



- 19 Shao X, Taha IN, Clauser KR, *et al.* MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Res* 2020;48:D1136–44.
- 20 Peyser R, MacDonnell S, Gao Y, *et al.* Defining the activated fibroblast population in lung fibrosis using single-cell sequencing. *Am J Respir Cell Mol Biol* 2019;61:74–85.
- 21 Xie T, Wang Y, Deng N, *et al.* Single-Cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep* 2018;22:S2211–1247(18)30339–5:3625–40..
- 22 Selman M, Pardo A, Barrera L, *et al.* Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 2006;173:188–98.
- 23 Konishi K, Gibson KF, Lindell KO, *et al.* Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2009;180:167–75.
- 24 Nance T, Smith KS, Anaya V, *et al.* Transcriptome analysis reveals differential splicing events in IPF lung tissue. *PLoS One* 2014;9:e92111.
- 25 Kusko RL, Brothers JF 2nd, Tedrow J, *et al.* Integrated genomics reveals convergent transcriptomic networks underlying chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;194:948–60.
- 26 Reyfman PA, Walter JM, Joshi N, *et al.* Single-Cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;199:1517–36.
- 27 Adams TS, Schupp JC, Poli S, *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020;6:eaba1983.
- 28 Wynn TA. Fibrotic disease and the T (H) 1/T (H) 2 paradigm. *Nat Rev Immunol* 2004;4:583–94.
- 29 Wynn TA, Cheever AW, Jankovic D, *et al.* An IL-12-based vaccination method for preventing fibrosis induced by schistosome infection. *Nature* 1995;376:594–6.
- 30 Roderfeld M, Rath T, Pasupuleti S, *et al.* Bone marrow transplantation improves hepatic fibrosis in abcb4^{-/-} mice via Th1 response and matrix metalloproteinase activity. *Gut* 2012;61:907–16.
- 31 Zhang S. The role of transforming growth factor β in T helper 17 differentiation. *Immunology* 2018;155:24–35.
- 32 Murray LA, Habiels DM, Hohmann M, *et al.* Antifibrotic role of vascular endothelial growth factor in pulmonary fibrosis. *JCI Insight* 2017;2:e92192:16..
- 33 Timpl R, Brown JC. Supramolecular assembly of basement membranes. *Bioessays* 1996;18:123–32.
- 34 Urushiyama H, Terasaki Y, Nagasaka S, *et al.* Role of $\alpha 1$ and $\alpha 2$ chains of type IV collagen in early fibrotic lesions of idiopathic interstitial pneumonias and migration of lung fibroblasts. *Lab Invest* 2015;95:872–85.
- 35 Shipley JM, Mecham RP, Maus E, *et al.* Developmental expression of latent transforming growth factor beta binding protein 2 and its requirement early in mouse development. *Mol Cell Biol* 2000;20:4879–87.
- 36 Enomoto Y, Matsushima S, Shibata K, *et al.* Ltbp2 is secreted from lung myofibroblasts and is a potential biomarker for idiopathic pulmonary fibrosis. *Clin Sci (Lond)* 2018;132:1565–80.
- 37 Lipson KE, Wong C, Teng Y, *et al.* CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis. *Fibrogenesis Tissue Repair* 2012;5(Suppl 1):S24.
- 38 Ziesche R, Hofbauer E, Wittmann K, *et al.* A preliminary study of long-term treatment with interferon gamma-1b and low-dose prednisolone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 1999;341:1264–9.
- 39 Richeldi L, Fernández Pérez ER, Costabel U, *et al.* Pamrevlumab, an anti-connective tissue growth factor therapy, for idiopathic pulmonary fibrosis (praise): a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet Respir Med* 2020;8:S2213–2600(19)30262–0:25–33..
- 40 Hung CF. Origin of myofibroblasts in lung fibrosis. *Curr Tissue Microenviron Rep* 2020;1:155–62.