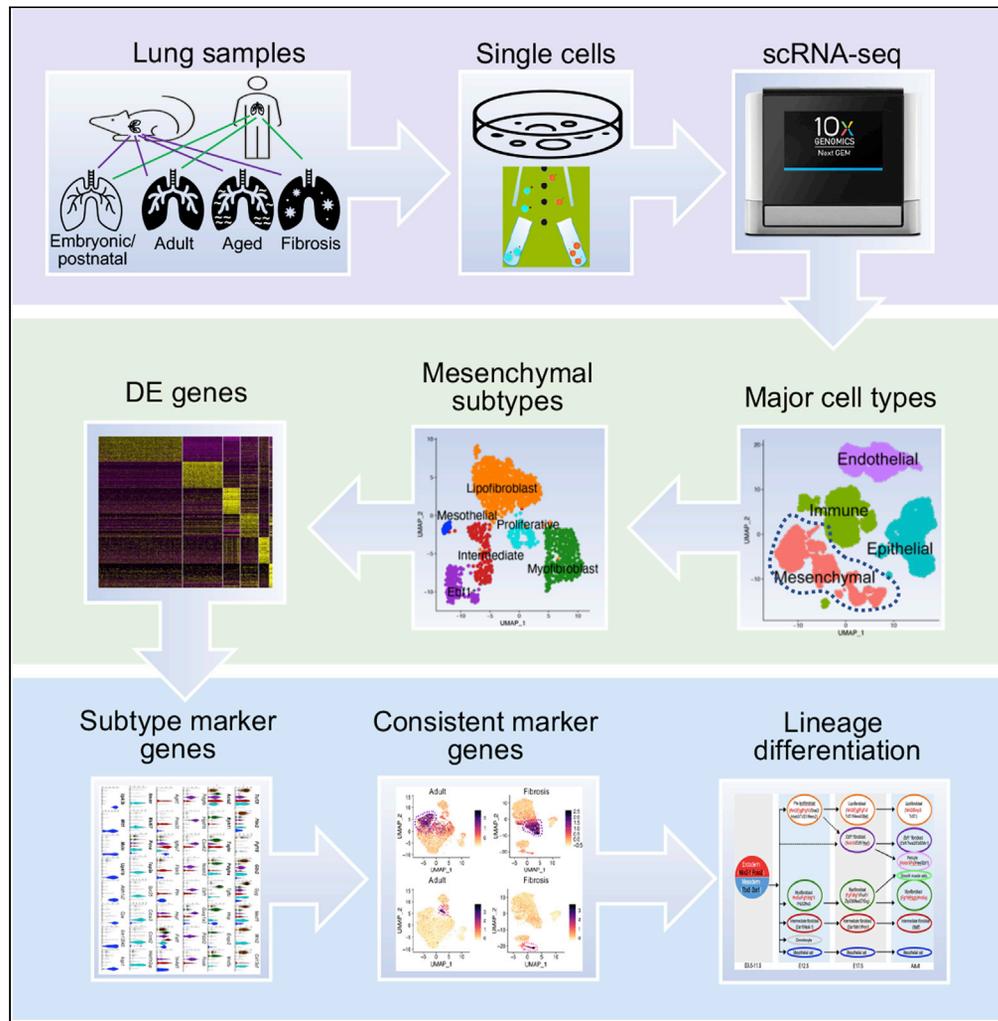


Article

Categorization of lung mesenchymal cells in development and fibrosis



Xue Liu, Simon C. Rowan, Jiurong Liang, ..., Barry R. Stripp, Paul W. Noble, Dianhua Jiang

paul.noble@cshs.org (P.W.N.)
dianhua.jiang@cshs.org (D.J.)

Highlights

Comprehensively profiling of lung mesenchymal subtypes in development and fibrosis.

Identification of novel lung mesenchymal cell subtypes in development and fibrosis.

Delineation of novel discriminative and consistent mesenchymal subtype markers.

Redefinition of the cell contributors of ECM in pulmonary fibrosis.

Liu et al., iScience 24, 102551
June 25, 2021 © 2021 The Authors.
<https://doi.org/10.1016/j.isci.2021.102551>



Article

Categorization of lung mesenchymal cells in development and fibrosis

Xue Liu,^{1,7} Simon C. Rowan,^{1,2,7} Jiurong Liang,¹ Changfu Yao,¹ Guanling Huang,¹ Nan Deng,³ Ting Xie,¹ Di Wu,³ Yizhou Wang,³ Ankita Burman,¹ Tanyalak Parimon,¹ Zea Borok,⁴ Peter Chen,¹ William C. Parks,¹ Cory M. Hogaboam,¹ S. Samuel Weigt,⁵ John Belperio,⁵ Barry R. Stripp,¹ Paul W. Noble,^{1,*} and Dianhua Jiang^{1,6,8,*}

SUMMARY

Pulmonary mesenchymal cells are critical players in both the mouse and human during lung development and disease states. They are increasingly recognized as highly heterogeneous, but there is no consensus on subpopulations or discriminative markers for each subtype. We completed scRNA-seq analysis of mesenchymal cells from the embryonic, postnatal, adult and aged fibrotic lungs of mice and humans. We consistently identified and delineated the transcriptome of lipofibroblasts, myofibroblasts, smooth muscle cells, pericytes, mesothelial cells, and a novel population characterized by *Ebf1* expression. Subtype selective transcription factors and putative divergence of the clusters during development were described. Comparative analysis revealed orthologous subpopulations with conserved transcriptomic signatures in murine and human lung mesenchymal cells. All mesenchymal subpopulations contributed to matrix gene expression in fibrosis. This analysis would enhance our understanding of mesenchymal cell heterogeneity in lung development, homeostasis and fibrotic disease conditions.

INTRODUCTION

Lungs of vertebrates consist of two intertwined and highly branched tree-like tubular systems—one conducting air and the other blood (Morrissey and Hogan, 2010). The epithelium and surrounding mesenchyme are two of the major cell components of the lung and are derived from endoderm and mesoderm during embryonic gastrulation, respectively (Herriges and Morrissey, 2014). The lineage diversity and differentiation of the pulmonary mesoderm is largely unknown, in spite of its critical functions during development and disease (Morrissey and Hogan, 2010). The pulmonary mesenchyme includes multiple distinct cell lineages with various functions in lung development and the pathogenesis and progression of debilitating respiratory conditions like idiopathic pulmonary fibrosis (IPF) (McCulley, et al., 2015; Rock, et al., 2011). Pulmonary mesenchymal cells, including commonly identified subtypes, like myofibroblasts, and poorly described groups, for instance adventitial fibroblasts, undergo dynamic structural, biochemical, and functional changes during development and disease (Lee, et al., 2017; Zepp, et al., 2017; Xie, et al., 2016a; Kumar, et al., 2014).

Recent studies utilizing single cell omics technologies, including single cell RNA-sequencing (scRNA-seq), have focused on defining the transcriptome of different cell types including lung mesenchymal cells (Guo, et al., 2019; Raredon, et al., 2019; Reyfman, et al., 2019; Valenzi, et al., 2019; Xie, et al., 2018). While lipofibroblasts, myofibroblasts, smooth muscle cells (SMCs), pericytes, and mesothelial cells are commonly reported, the transcriptomic signatures differ in these descriptive studies. Frequently publications identify subtypes by a mixture of location and/or discriminative gene expression (Adams, et al., 2020; Habermann, et al., 2020; Mayr et al., 2020; Travaglini, et al., 2020; Morse, et al., 2019; Reyfman, et al., 2019). Further, the use of different databases and cells of divergent developmental or disease stages has been confounding, resulting in a range of different transcriptomic signatures being attributed to the same cell population (Park, et al., 2019; Peyser, et al., 2019; Raredon, et al., 2019; Xie, et al., 2018). An array of cell “specific” markers has been reported. Highly discriminative markers, especially for fibroblast subpopulations, remain elusive and the majority of markers are non-specific. To add to the confusion, clusters identified by high expression of a delineating gene in one publication are subsequently identified by others as a predefined

¹Department of Medicine and Women’s Guild Lung Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

²UCD School of Medicine, Conway Institute, University College Dublin, Belfield, Ireland

³Genomics Core, Cedars-Sinai Medical Center, CA 90048, USA

⁴Division of Pulmonary and Critical Care Medicine, Keck School of Medicine of University of Southern California, Los Angeles, CA 90033, USA

⁵Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA

⁶Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

⁷These authors contributed equally

⁸Lead contact

*Correspondence: paul.noble@cshs.org (P.W.N.), dianhua.jiang@cshs.org (D.J.) <https://doi.org/10.1016/j.isci.2021.102551>



mesenchymal cell type (Travaglini, et al., 2020; Morse, et al., 2019). The current approach likely leads to overlap of distinct clusters and does little to resolve the controversies regarding the definitive transcriptomic signature of the pulmonary mesenchymal populations.

Achieving a comprehensive understanding of the different mesenchymal populations in the murine and human lung is critical for advancing our understanding of the contribution of the different populations to cell lineage contribution in development and fibrotic disease processes. In this study, we undertook a comprehensive and longitudinal scRNA-seq analysis of mesenchymal cells from developing, healthy and fibrotic murine and human lungs. We characterized all known and novel molecularly distinct mesenchymal populations, examined known marker gene expression and specification, and identified novel marker genes that were highly discriminative for each subtype. This resource provides a basis for the interrogation and investigation of each subtype in development, health and disease by the academic and clinical research community. It will enhance our understanding of lung development and may aid the development of targeted therapies for the treatment of pulmonary fibrosis.

RESULTS

scRNA-seq on E17.5 murine lungs identified mesenchymal cell subtypes

Among all the known mesenchymal subpopulations, lipofibroblasts are the least well described but are known to emerge during late embryonic stages in mouse lung, when myofibroblasts are also prominent (Kugler, et al., 2017; Al Alam, et al., 2015; Chao, et al., 2015; Bostrom, et al., 1996). To comprehensively profile these subpopulations, we first performed scRNA-seq on E17.5 murine lungs. Sequencing libraries were prepared using the 10x genomics chromium system from FACS-purified cells (Figure S1A). Samples were integrated after quality control (Figures S1B, S1C, and 1A), and cells were visualized in two dimensions according to their gene expression profiles using Uniform Manifold Approximation and Projection (UMAP) (Figure S1D). 2,002 mesenchymal cells were subset from distinct immune, epithelial and endothelial clusters and checked for purity (Figures 1B and S1E–S1G, Table S1). We identified discrete clusters of lipofibroblasts (*Plin2*⁺, *Tcf21*⁺), myofibroblasts (*Acta2*⁺, *Pdgfra*⁺), proliferating fibroblasts (*Hmnr*⁺, *Mki67*⁺), mesothelial cells (*Wt1*⁺, *Upk3b*⁺), a novel population defined by *Ebf1* expression and an intermediate subtype that displayed low expression of genes from multiple populations (Figure 1C) (Park, et al., 2019; Li, et al., 2018; Xie, et al., 2018; Al Alam, et al., 2015; Rock, et al., 2011). Heatmap visualization of the top 30 differentially expressed (DE) genes revealed a distinct pattern of gene expression in each identified cluster (Figure 1D) and representative known and novel delineating genes for each subtype were visualized (Figures 1E and S1H).

To validate the identified subpopulations, two further linear dimensional reduction assays, t-SNE and PCA, were performed using independent component analysis on module data (k-means) and metagene data. Corresponding mesenchymal clusters were identified with the proliferative fibroblast and mesothelial cell populations excluded (Figures S2A and S2B). A customizable suite of single-cell R-analysis tools (SCRAT) based on self-organizing maps (SOMs) machine learning was used to analyze for sample similarity and perform pseudotime analysis. The correlation-spanning tree and trajectory report suggested a directed hierarchical relationship between the fibroblast subpopulations. The correlation-spanning tree and k-nearest neighbor graph began from the lipofibroblast cluster, bifurcated to intermediate fibroblasts and finally bifurcated to *Ebf1*⁺ fibroblasts and myofibroblasts (Figures S2C–S2F).

SMCs and pericytes are known components of lung mesenchyme but were not identified in the initial scRNA-seq analysis, possibly due to the number of cells analyzed. Therefore, scRNA-seq analysis was performed on much more cells from three more embryonic lungs together with tracheas and main bronchus (Table S1). The samples were integrated after quality control and clustered (Figure S3A), and major cell types and gene signatures were identified (Figures S3B and S3C). The mesenchymal fraction containing 9,076 cells was subset, clustered, and the fraction purity confirmed (Figure S3D). The mesenchymal subtypes were identified using the subtype specific signature genes identified in the initial analysis (Figure S3E) and each subtype displayed a distinct pattern of gene expression (Figure S3F). Two distinct SMC clusters were identified in this supplementary analysis (Figure S3E). The molecular signature of the mesenchymal subpopulations was homologous to those in the initial analysis (Figures S3G and S3H). No clear pericyte cluster was identified in either data set.

scATAC-seq on E17.5 murine lungs confirmed the mesenchymal cell subtypes

To validate the definition of the mesenchymal subpopulations in the scRNA-seq data sets at E17.5 lungs, we performed a further unbiased analysis, single-cell ATAC-seq (scATAC-seq), on an additional three E17.5

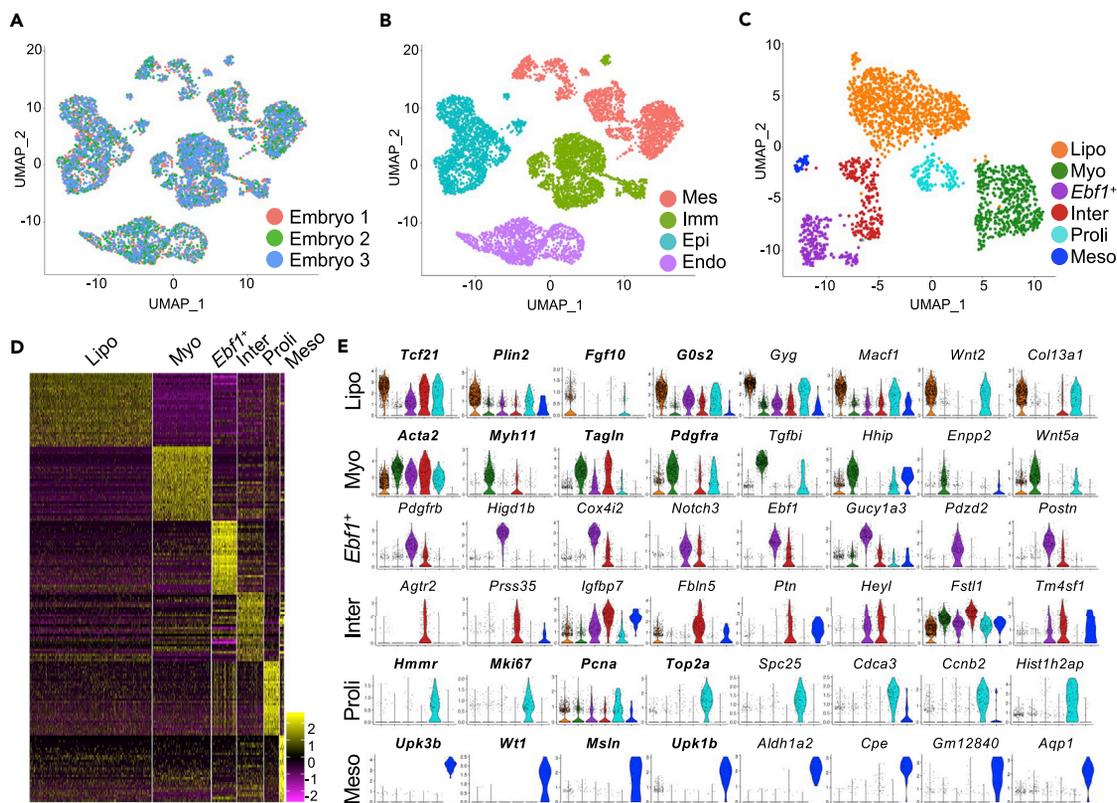


Figure 1. Single-cell RNA profiling of E17.5 mouse lung mesenchymal cells

(A and B) UMAP visualization of sample integration (A) and cell type definition (B) of E17.5 mouse lung scRNA-seq data. Mes, mesenchymal cells; Imm, immune cells; Epi, epithelial cells; Eryth, erythrocyte; Endo, endothelial cells.

(C) Six mesenchymal cell clusters were defined shown by UMAP.

(D) Heatmap presentation of top 30 DE genes (rows) for individual cells (columns) in each subtype.

(E) Violin plot representation showing relative expression of the fibroblast cluster classical (Sia, et al., 2019) and novel (non-bold) signature genes.

Lipo, lipofibroblasts; Myo, myofibroblasts; *Ebf1*⁺, *Ebf1*⁺ fibroblasts; Inter, intermediate fibroblasts; Proli, proliferative fibroblasts; Meso, Mesothelial cells. See also Figures S1–S3.

mouse lungs. After quality control (Figures S4A and S4B), sample integration using Harmony (Figure 2A), and major cell type clustering (Figures S4C, S4D, and 2B), 4,287 mesenchymal nuclei were extracted (Figure S4E) and the fraction purity was confirmed (Figure S4F). The mesenchymal populations in the scATAC-seq data set were identified by comparing the gene expression level (scRNA-seq) and gene activity (scATAC-seq) data in these E17.5 data sets to identify the shared characters (Figure 2C). Subtype signatures and two representative genes of each subtype were visualized by UMAP (Figures 2D and S4G) and a heatmap of the top 30 subcluster specific genes confirmed the distinct gene expression pattern in each subcluster (Figure S4H). Additional subtype specific genes were illustrated using dot plots (Figure S4I).

To determine whether similar mesenchymal cell subtypes would be consistently defined by the two assays, the scRNA-seq and scATAC-seq datasets were then integrated and batch effect was corrected (Figure 2E). The major populations were identified using cell type specific genes (Figures 2F, S5A, and S5B) and the mesenchymal cells were extracted and re-clustered (Figure 2G). Mesenchymal cell subtypes (Figure 2H) were identified by the expression of subtype specific gene signatures (Figure S5E) with the top 30 genes in each cluster visualized by heatmap (Figure S5D). Although differences in relative levels of gene accessibility (scATAC-seq) and transcript (scRNA-seq) were present in the integrated data set, the previously identified subtype-specific signatures were consistently identified (Figure 2I).

Identification of lung mesenchymal subtypes throughout development and fibrosis

To comprehensively profile the early lineages of the subpopulations identified at E17.5, scRNA-seq data sets from earlier developmental stages (E9.5, 10.5, 11.5, 12.5, 14.5, 16.5) were examined. 215 cells from

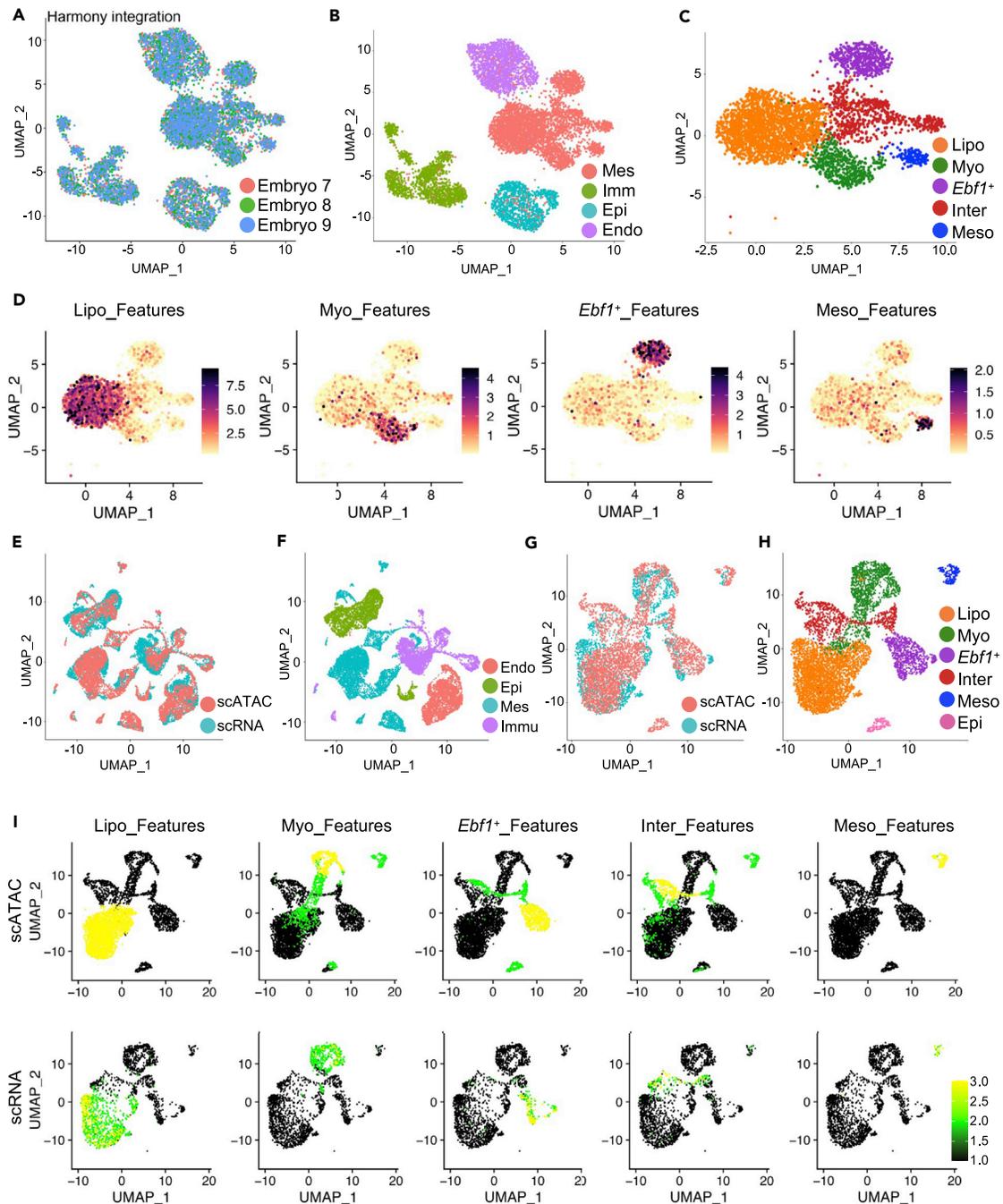


Figure 2. Single-cell ATAC-seq of E17.5 mouse lungs

(A and B) UMAP visualization of sample integration by Harmony (A) and cell type definition (B) of the scATAC-seq data of E17.5 mouse lungs. Mes, mesenchymal cells; Imm, immune cells; Epi, epithelial cells; Endo, endothelial cells.

(C) Mesenchymal cell subtype identification on E17.5 mesenchymal cells.

(D) Average expression of mesenchymal cell subtype feature genes (Lipo_Features: *Col13a1*, *Macf1*, *Limch1*, *Wnt2*; Myo_Features: *Tgfb1*, *Adcy7*, *Lgr6*, *Egfm1*; *Ebf1*⁺_Features: *Higd1b*, *Pdgfrb*, *Heyl*, *Gucy1a3*; Meso_Features: *Upk3b*, *Wt1*, *Krt19*, *Lrrc52*) were visualized by UMAPs.

(E and F) Integration of the scATAC-seq and scRNA-seq data of E17.5 mouse lungs (E) and major cell type definition in the integrated data (F).

(G) Mesenchymal cells were extracted and scATAC-seq and scRNA-seq cell distribution was visualized by UMAP.

(H and I) Mesenchymal cell subtypes were identified (H) by the average expression of the signature genes (Lipo_Features: *Col13a1*, *Wnt2*, *Macf1*, *Gyg*; Myo_Features: *Tgfb1*, *Hhip*, *Enpp2*, *Wnt5a*; *Ebf1*⁺_Features: *Ebf1*, *Higd1b*, *Pdzd2*, *Postn*; Inter_Features: *Agr2*, *Prss35*, *Fbln5*, *Ptn*; Meso_Features: *Msln*, *Upk3b*, *Lrrm4*, *Wt1*) (I).

Lipo, lipofibroblasts; Myo, myofibroblasts; *Ebf1*⁺, *Ebf1*⁺ fibroblasts; Inter, intermediate fibroblasts; Meso, Mesothelial cells.

See also Figures S4 and S5.

E9.5, E10.5, and E11.5 data sets were integrated after quality control (Figure S6A) (Pijuan-Sala, et al., 2019). Distinct endoderm (*Nkx2-1*⁺, *Foxa2*⁺) and mesoderm (*Tbx5*⁺, *Osr1*⁺) clusters were identified (Figures 3A, 3B, S6B, and S6C). However, the subtype-specific transcriptomic profiles identified at E17.5 were indistinct, suggesting the differentiation fate of the mesodermal cells was not yet determined.

Using published data sets (Cohen, et al., 2018; Han, et al., 2018), 1,158 and 4,246 mesenchymal cells were obtained from E12.5 and E14.5 lungs (Figures S6D–S6F and S6H–S6I, Table S1). 2,981 α SMA-GFP; *Tbx4*-Cre; *Rosa26*-tdTomato (*Tbx4*-lineage⁺, α SMA⁺) fibroblasts were FACS-purified from E16.5 murine lungs as per our previous study (Xie, et al., 2018) (Figures S6K–S6L). 3,664 mesenchymal cells from postnatal day 1 (P1) lungs and 3,125 and 1,097 mesenchymal cells from P7 and P15 *Pdgfra*-GFP⁺ lungs (Guo, et al., 2019; Li, et al., 2018) were accessed from published studies (Figures S7A–C, S7E and S7G–I). All mesenchymal cell purity was confirmed, and subtypes were identified (Figures 3C–3H). The resulting mesenchymal clusters were distinct and displayed homologous signatures to the mesenchymal populations identified at E17.5 (Figures 3C–3H, S6G, S6J, S6M, S7D, S7F, and S7J, Table S1).

Adult murine scRNA-seq data from our and others' previously published studies were accessed (Aran, et al., 2019; Parimon, et al., 2019; Raredon, et al., 2019; Reyfman, et al., 2019; Xie, et al., 2016a). 4,193 and 4,728 mesenchymal cells from adult normal and bleomycin injured murine lungs, respectively, were extracted and integrated after quality control (Figures S8A, S8B, S8D, and S8E, Table S1). In addition to the subtypes identified in earlier data sets, distinct SMC cluster and a pericyte cluster were also identified (Figures 3I, 3J, S8C, and S8F). As IPF is a disease of aging, mesenchymal cells (EPCAM/CD31/CD45⁺) were collected from three aged mouse lungs and three age matched lungs 14 days after bleomycin injury for scRNA-seq. 12,304 and 13,335 mesenchymal cells were extracted from aged normal and fibrotic lungs, respectively, following sample integration and quality control (Figures S8G–S8I and S8K–S8M). The mesenchymal cell subtypes identified in the aged normal and fibrotic lungs were similar to those identified in adult mouse lungs (Figures 3K, 3L, S8J, and S8N).

To determine if corresponding mesenchymal subpopulations were present in the human lungs, single cell lung suspensions of explanted healthy and IPF donor lung tissues were generated with scRNA-seq performed on the EPCAM⁺ FACS-purified population. After quality control, the major cell types were identified as described in mouse data sets. The results of scRNA-seq on P1, month 21 (M21), healthy and IPF donor human lung tissue from publicly available data sets (Adams, et al., 2020; Habermann, et al., 2020; Morse, et al., 2019; Reyfman, et al., 2019; Valenzi, et al., 2019) were re-analyzed and integrated, where appropriate, with the scRNA-seq data generated in our laboratory (Table S2). Up to seven molecularly distinct mesenchymal subpopulations were identified with distinct and highly conserved transcriptomic profiles that were orthologous to those in the corresponding murine lung subpopulations (Figures 3M–3O).

Conserved and time point-specific signature genes of lipofibroblast

To identify specific and consistent markers, we determined the DE genes of lipofibroblasts at each stage and visualized the top genes (Figures 4A and S9A). Accordingly, we identified genes *Limch1*, *Gyg*, *Macf1*, *Mfap4*, *Npnt*, *Wnt2*, *Col13a1*, and *Inmt* that were consistently expressed and discriminative in the lipofibroblast clusters at all data sets (Figures 4A and S9A). Among these novel genes, *Gyg*, *Macf1*, *Wnt2*, and *Col13a1* were the most specific and consistently expressed compared to canonical markers (Figures 4B and S9B) and might better distinguish lipofibroblasts *in vivo*.

Commonly reported lipofibroblast markers, represented by *Tcf21*, *Plin2*, *Fgf10*, and *G0s2* (Park, et al., 2019; Al Alam, et al., 2015; McGowan and McCoy, 2014), were prominent when lipofibroblasts emerged in the embryonic lung (E16.5–E17.5) and when the presence of lipofibroblasts reportedly peaks (P7–15) (Figure 4B). At all other developmental/disease stages, these genes, except *Tcf21*, were poorly discriminative for the lipofibroblast cluster (Figure 4B).

To validate the identified lipofibroblast signature, mesenchymal cells were FACS-purified from adult murine lungs using cell surface proteins encoded by prominently expressed lipofibroblast genes in the scRNA-seq data set, represented by CD249 (*Enpep*) (Figures S9C and S9D). The top DE genes in bulk-sequenced CD249⁺ fibroblasts in comparison to CD249⁻ fibroblasts overlapped substantially with the highly discriminative lipofibroblast genes in the scRNA-seq analysis, like *Limch1*, *Col13a1*, *Fgf10*, and *Tcf21* (Figure S9E). To further investigate the *in vitro* functions of lipofibroblasts, isolated murine lung fibroblasts were cultured

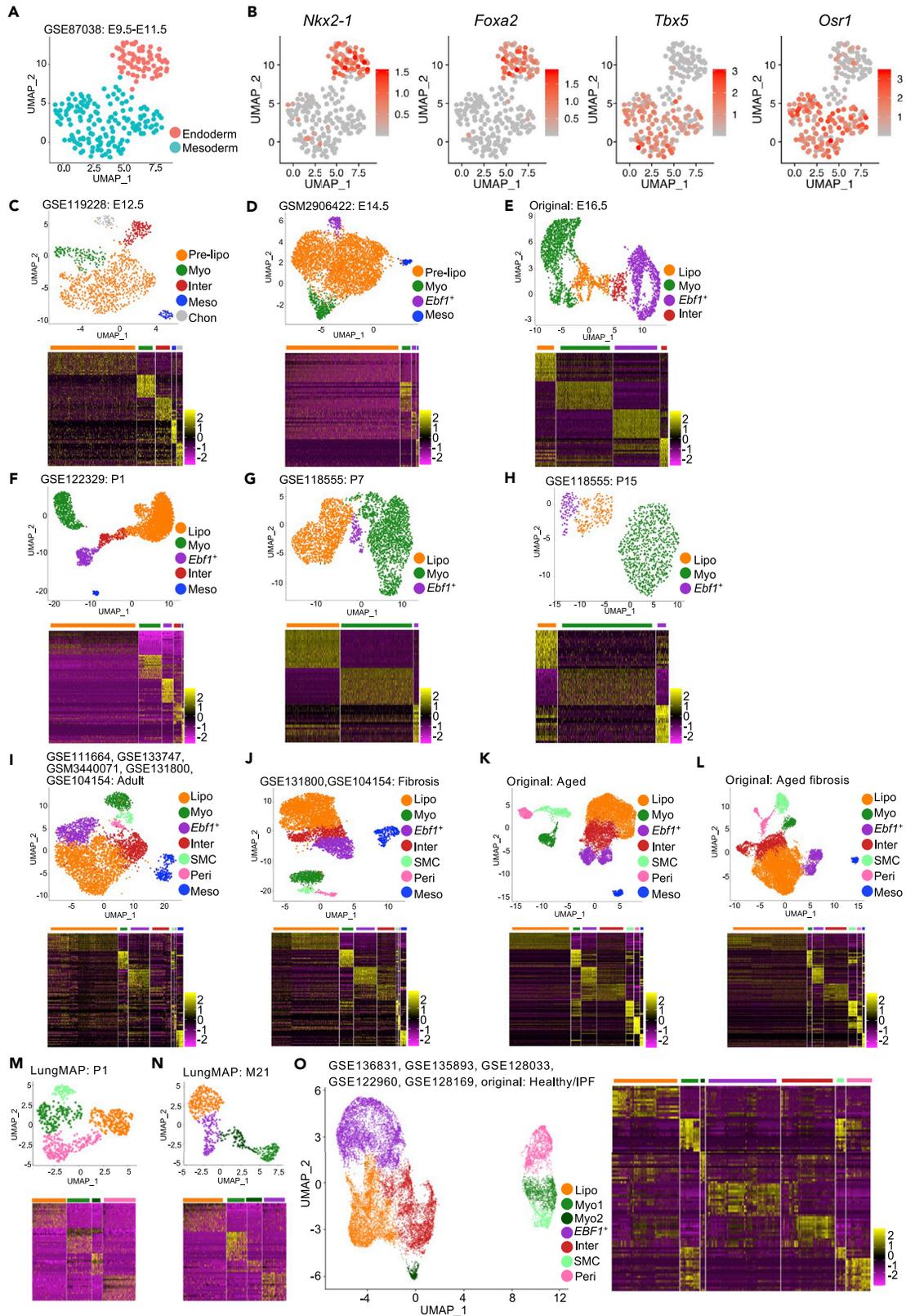


Figure 3. Classification of lung fibroblast subtypes in murine and human lungs

(A and B) UMAP visualization of E9.5-E11.5 lung endoderm and mesoderm (A) and the transcripts of specific transcription factors (B). (C–L) UMAP visualization of mesenchymal cell subtype classification and heatmaps of top 15 genes in E12.5 (C), E14.5 (D), E16.5 (E), P1 (F), P7 (G) and P15 (H) mouse lungs and in adult (I, J) and aged (K, L) mice lung before (I, K) and after (J, L) bleomycin injury. (M–O) UMAP visualization of mesenchymal cell subtype classification and heatmaps of top 15 genes in P1 (M), M21 (N) and integrated healthy and IPF human lungs (O). Pre-lipo/Lipo, Pre-lipofibroblasts/lipofibroblasts; Myo, myofibroblasts; *Ebf1*⁺/*EBF1*⁺, *Ebf1*⁺/*EBF1*⁺ fibroblasts; Inter, intermediate fibroblasts; SMC, smooth muscle cells; Peri, pericytes; Meso, Mesothelial cells; Chon, Chondrocytes. P1, Postnatal day 1; M21, 21 months. See also [Figures S6–S8](#).

and biochemically stimulated using methods described in the literature ([Figure 4C](#)). Stimulated cells displayed pronounced lipid inclusions ([Figure 4C](#)). Lipofibroblast-like cultures, analyzed following scRNA-seq, were entirely free of contaminating cell types and displayed high transcript expression of canonical makers, like *Plin2* and *Fgf10* ([Figures 4D](#) and [S9F](#)). However, their transcriptomic signature differed from lipofibroblasts *in vivo*. Colony forming assays using lipofibroblast-like cells demonstrated they were more supportive of AEC2 colony formation than unstimulated cells ([Figures 4E](#) and [4F](#)).

In humans, consistent with murine lung, many of the novel lipofibroblast signature genes identified in the murine lung consistently discriminated a specific cluster in the human data sets ([Figures 4G](#), [4H](#), and [S9G](#)). Novel marker genes and transcription factors for this population included many genes identified in murine lipofibroblasts, like *LIMCH1* and *MACF1* ([Figures 4G](#), [4H](#), and [S9G](#)). Again, these conserved novel signature genes were consistently discriminative in comparison to canonical markers. Commonly reported lipofibroblast markers, with the exception of *TCF21*, were found to poorly discriminate a distinct cluster ([Figure 4H](#)). Statistically identified genes, whose expression changed most significantly in healthy vs. IPF lipofibroblasts included, upregulated ECM-related genes and fibrosis promoting/protective genes ([Figure S9H](#)).

Delineation of novel discriminative markers for myofibroblasts and SMCs

The transcriptomic profiles of myofibroblasts and SMCs have not yet been definitively determined. In the current study, clear myofibroblast clusters were identified in all mouse data sets ([Figures 3C–3L](#)). The DE genes of these myofibroblasts clusters were visualized using volcano ([Figure 5A](#)) and violin plots ([Figures S10A](#) and [S10B](#)). *Tgfb1*, *Hhip*, *Enpp2*, *Egfm1*, *P2ry14*, *Wnt5a*, *Nnat*, *Mustn1*, *Actg2*, and *Cnn1* were among the top DE genes. These genes were found to be more discriminative and conserved between data sets in murine myofibroblasts ([Figures 5A–5C](#), [S10A](#), and [S10B](#)). The four most specific and highly expressed genes in the myofibroblast clusters were *Tgfb1*, *Hhip*, *Enpp2*, and *Wnt5a* ([Figures 5B](#), [5C](#), [S10C](#), and [S10E](#)). *Acta2*, *Tagln*, and *Pdgfra*, although widely reported as myofibroblast marker genes ([Li, et al., 2018](#); [Murgai, et al., 2017](#); [Hsia, et al., 2016](#); [Robin, et al., 2013](#); [Rock, et al., 2011](#); [Hinz, et al., 2007](#)), are highly expressed in other mesenchymal subtypes ([Figures 5B](#) and [S10A](#)). *Acta2*, *Myh11*, and *Tagln* were preferentially expressed in all early myofibroblast clusters (E12.5, E14.5, E16.5, E17.5, P7, P15) except P1 where these genes displayed limited transcript expression in murine lungs ([Figures 5B](#) and [S10A](#)). None of these genes was as discriminative and conserved as novel markers which we identified above ([Figures 5B](#) and [S10A](#)).

A distinct SMC cluster in the embryonic and early postnatal lung data sets could not be detected in most mouse data sets ([Figures 1C](#) and [2C–2H](#)), with the exception of the larger E17.5 data set that included the tracheas and main bronchi ([Figures S3E](#) and [S3G](#)). A limited number of genes associated commonly with SMCs were detected in the myofibroblast cluster of the earliest data sets (E12.5, E14.5). In the adult and aged normal and fibrotic murine lung, distinct SMC clusters were identified by higher expression of commonly associated genes, like *Acta2*, *Myh11* alongside more SMC-specific markers like *Actg2* and *Actc1* ([Figures 3I–3L](#), [5C](#), [S10D](#), and [S10F](#)).

In human lungs, myofibroblasts and SMCs were closely associated and had similar transcriptomic signatures ([Figure 5D](#)). High expression of SMC-related marker genes, represented by *CNN1*, *SYNPO2*, *ACTG2*, was used to differentiate these related, and closely associated, populations in the integrated adult data set ([Figures 5D](#), [5E](#), [S10G](#), and [S10H](#)). Unique myofibroblast specific genes, not expressed in SMCs could not be identified. In addition to commonly used SMC markers, the P1 dataset expressed the reported human airway SMC marker *HHIP* ([Figure 5E](#)) ([Danopoulos, et al., 2020](#)). To confirm the identification of distinct SMC and myofibroblast clusters a subset of the healthy cells were clustered. In this analysis, SMC and myofibroblast clusters were distinct, and SMC markers, including recently reported vascular smooth muscle specific genes *MEF2C* and *NTRK3*, were clearly discriminative ([Figures S10I](#) and [S10J](#)).

Figure 4. Continued

(E and F) Colony formation assays (E) and colony forming efficiency (CEF) quantification (F) were performed to examine the supporting potentials of the control and stimulated fibroblasts. Scale bar, 1 mm.

(G) UMAP visualization of averaged expression of novel human lipofibroblast signature genes (*A2M*, *LIMCH1*, *GPC3*, *SCN7A*, *RGCC*) in P1, M21 and integrated healthy/IPF lungs.

(H) Comparison of known and novel human lipofibroblast markers and cluster specific transcription factors (TFs) in each human data set. All listed genes were at $p < 10^{-5}$ and Avg_logFC > 1.

Pre-lipo/Lipo, Pre-lipofibroblasts/lipofibroblasts; Myo, myofibroblasts; *Ebf1*⁺/*EBF1*⁺, *Ebf1*⁺/*EBF1*⁺ fibroblasts; Inter, intermediate fibroblasts; SMC, smooth muscle cells; Peri, Pericytes; Prol, proliferative fibroblasts; Meso, Mesothelial cells; Chon, Chondrocytes. P1, postnatal day 1; M21, 21 months.

See also [Figures S9](#), [S14](#), and [S15](#).

([Danopoulos, et al., 2020](#)). However, despite distinct clustering the transcriptomic signatures, other than the expression of commonly reported SMC markers, were very similar.

Commonly used myofibroblast-related genes were prominent in both populations, but more highly expressed genes in the SMC clusters were present in the data set ([Figure 5E](#)). Two myofibroblast subpopulations were identified in M21 and adult human lungs ([Figures 5D and 5E](#)). The first (Myo1), highly expressed commonly reported myofibroblast marker genes. The second (Myo2) had a gene profile homologous to that of “classical myofibroblasts” in a recent publication ([Travaglini, et al., 2020](#)). The top DE genes and transcription factors in these clusters were determined statistically ([Figure 5E](#)). Comparative analysis of changes in gene expression between healthy and IPF myofibroblast and SMC populations highlighted altered expression of numerous genes implicated in the pathogenesis of IPF ([Figure 5F](#)).

Identification of an *Ebf1*⁺ fibroblast subtype and pericytes

A distinct, previously unidentified cluster of mesenchymal cells defined by *Ebf1* expression, emerged at E14.5 and was identified at mouse lung examined ([Figures 3D–3L](#)). DE analysis revealed time point-specific genes for the *Ebf1*⁺ cluster in each data set ([Figure 6A](#)). *Ebf1*⁺ cells displayed a distinct signature represented by *Ebf1*, *Gucy1a3*, *Pdzd2*, *Postn*, *Pdgfrb*, *Higd1b*, *Cox4i2* and *Notch3*, up to P1 ([Figures 6A](#), [6B](#), [S11A](#), and [S11C](#)). From P7 onward the transcriptomic profile was better represented by *Ebf1*, *Serpinf1*, *Postn*, *Col14a1* and *Pi16* ([Figures 6C](#), [S11B](#), and [S11C](#)). In adult and aged mesenchymal cells, two *Ebf1*⁺ clusters were identified ([Figures 3I–3L](#)). One *Ebf1*⁺ cluster we identified as pericytes due to condensed expression of known pericyte markers like *Cspg4* (Ng2) and *Pdgfrb* ([Figures 6D](#), [S11D](#), and [S11E](#)). The other distinct cluster, also *Ebf1*⁺, expressed the novel transcriptomic signatures identified ([Figures 6B–6D](#), and [S11C](#)). Discriminative genes for the *Ebf1*⁺ cluster in the E14.5-P1 lungs included genes, for example *Higd1b*, *Cox4i2*, and *Notch3*, subsequently identified among the top DE genes of adult/aged lung pericytes ([Figures 6B–6D](#), [S11C](#), and [S11D](#)). These data suggest that pericytes and adult *Ebf1*⁺ fibroblasts diverge during early development but may share a common lineage. Most specific and consistent genes for adult/aged *Ebf1*⁺ cluster were visualized ([Figures 6D](#) and [S11F](#)). Traditional pericyte markers, like *Cspg4*, or the common lineage marker *Foxd1* ([Chen and Fine, 2016](#)), displayed low transcript expression ([Figure S11E](#)), while *Pdgfrb* was condensed in pericytes but with high background in other clusters ([Figures 6D](#) and [S11G](#)). Novel pericyte markers identified in our analysis were expressed at a greater level, with greater specification than commonly used pericyte marker genes ([Figures 6D](#), [S11D](#), and [S11G](#)).

At the protein level, co-staining for EBF1, the endothelial cell marker, vWF, and SMC marker, α SMA, on E17.5 lung sections indicated that EBF1⁺ cells were not only perivascular (pericytes) but also in interstitial lung tissue (*Ebf1*⁺ fibroblasts) ([Figure 6E](#)). This was also true in P7 mouse lung indicated by another endothelial cell marker, CD31, and Ebf1 con-staining ([Figures S12A–S12C](#)). These data support the hypothesis that the *Ebf1*⁺ population may consist of both pericytes and a distinct fibroblast subtype.

In human lungs, we identified a mesenchymal population with an orthologous transcriptomic signature to the murine lung *Ebf1*⁺ cluster ([Figures 6F](#), [6G](#), and [S12D](#)). As noted in the murine lung, canonical pericyte markers, like *RGS5* and *CSPG4* were expressed in limited numbers of cells ([Figures 6F](#) and [S12E](#)). Conserved expression of more novel pericyte marker genes, represented by *HIGD1B*, *NDUFA4L2*, *COX4I2* and *FAM162B* among others, were identified ([Figures 6F](#), [6H](#), and [S12E](#)). The top DE genes and transcription factors in the *EBF1*⁺ and pericyte clusters were determined statistically ([Figures 6F–6H](#), [S12D](#), and [S12E](#)). Comparative analysis of changes in gene expression between healthy and IPF *EBF1* fibroblasts and pericytes again highlighted the genes related to fibroblast migration, proliferation and fibrosis ([Figure 6I](#)).

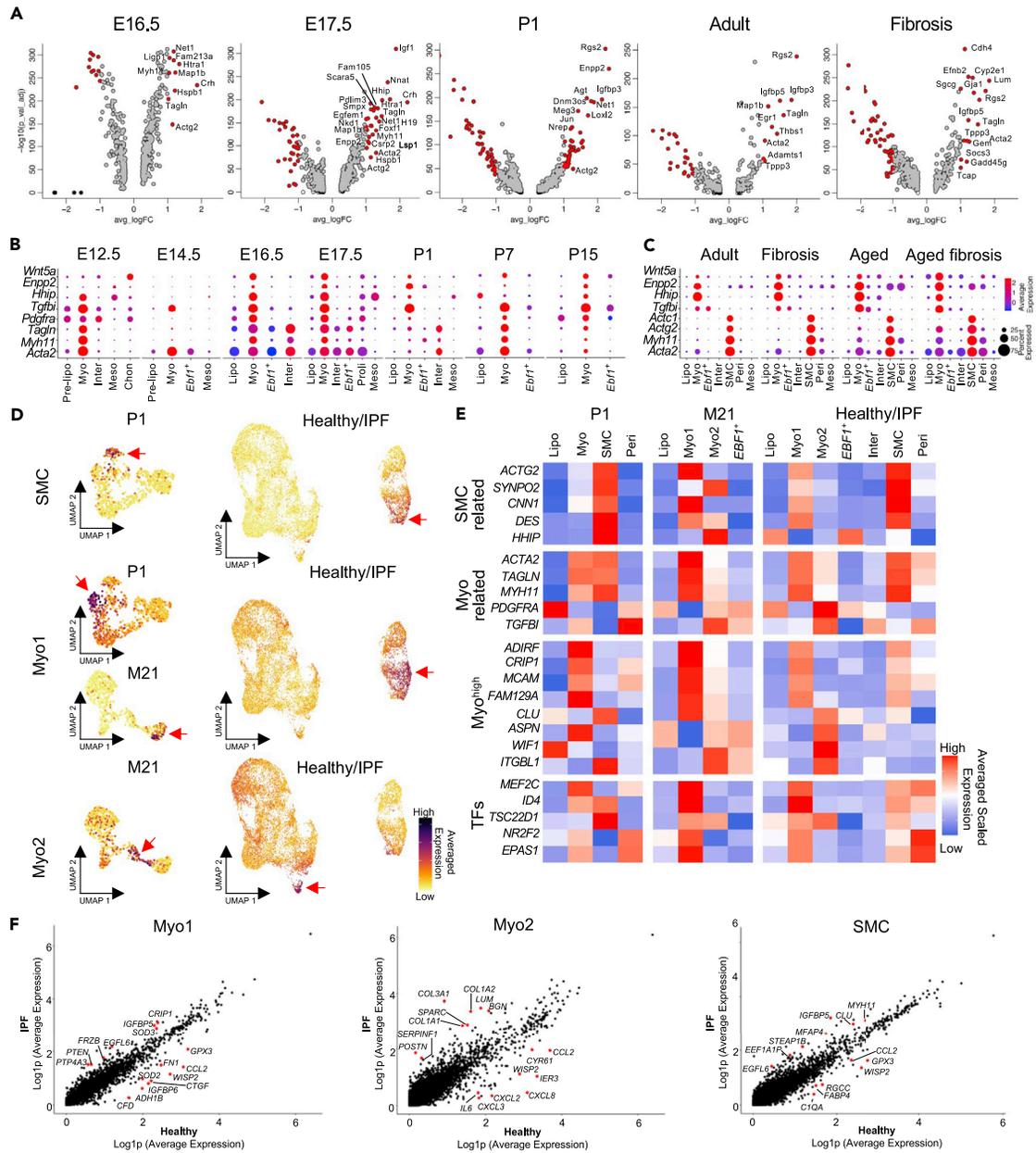


Figure 5. Identification novel markers for lung myofibroblast and SMC subtypes

(A) Visualization of DE genes of myofibroblasts in mouse lungs by volcano plots. Genes in red, $p < 10^{-5}$; Avg_logFC > 1. Genes in black, $p < 10^{-5}$; Avg_logFC < 1. Genes in gray, $p > 10^{-5}$; Avg_logFC > 1.

(B) Comparison of known and novel myofibroblast markers in embryonic and postnatal mouse lungs.

(C) Visualization of myofibroblast and SMC markers in adult and aged normal and fibrotic mouse lungs by dot plots.

(D) UMAP visualization of averaged expression of human SMC (ACTG2, SYNP O2, CNN1) and myofibroblast (Myo1: ADIRF, CRIP1, MCAM, FAM129A; Myo2: CLU, ASPN, WIF1, ITGBL1) signature genes in P1, M21 and integrated healthy/IPF lungs.

(E) Comparison of known and novel human lung relevant SMC and myofibroblast markers, and cluster specific transcription factors (TFs) in each human data set. All listed genes were at $p < 10^{-5}$ and Avg_logFC > 1.

(F) Comparative analysis of changes in gene expression in healthy and IPF myofibroblasts and SMC clusters.

Pre-lipo/Lipo, Pre-lipofibroblasts/lipofibroblasts; Myo, myofibroblasts; Ebf1+/EBF1+, Ebf1+/EBF1+ fibroblasts; Inter, intermediate fibroblasts; SMC, smooth muscle cells; Peri, Pericytes; Prolif, proliferative fibroblasts; Meso, Mesothelial cells; Chon, Chondrocytes. P1, postnatal day 1; M21, 21 months.

See also Figures S10, S14, and S15.

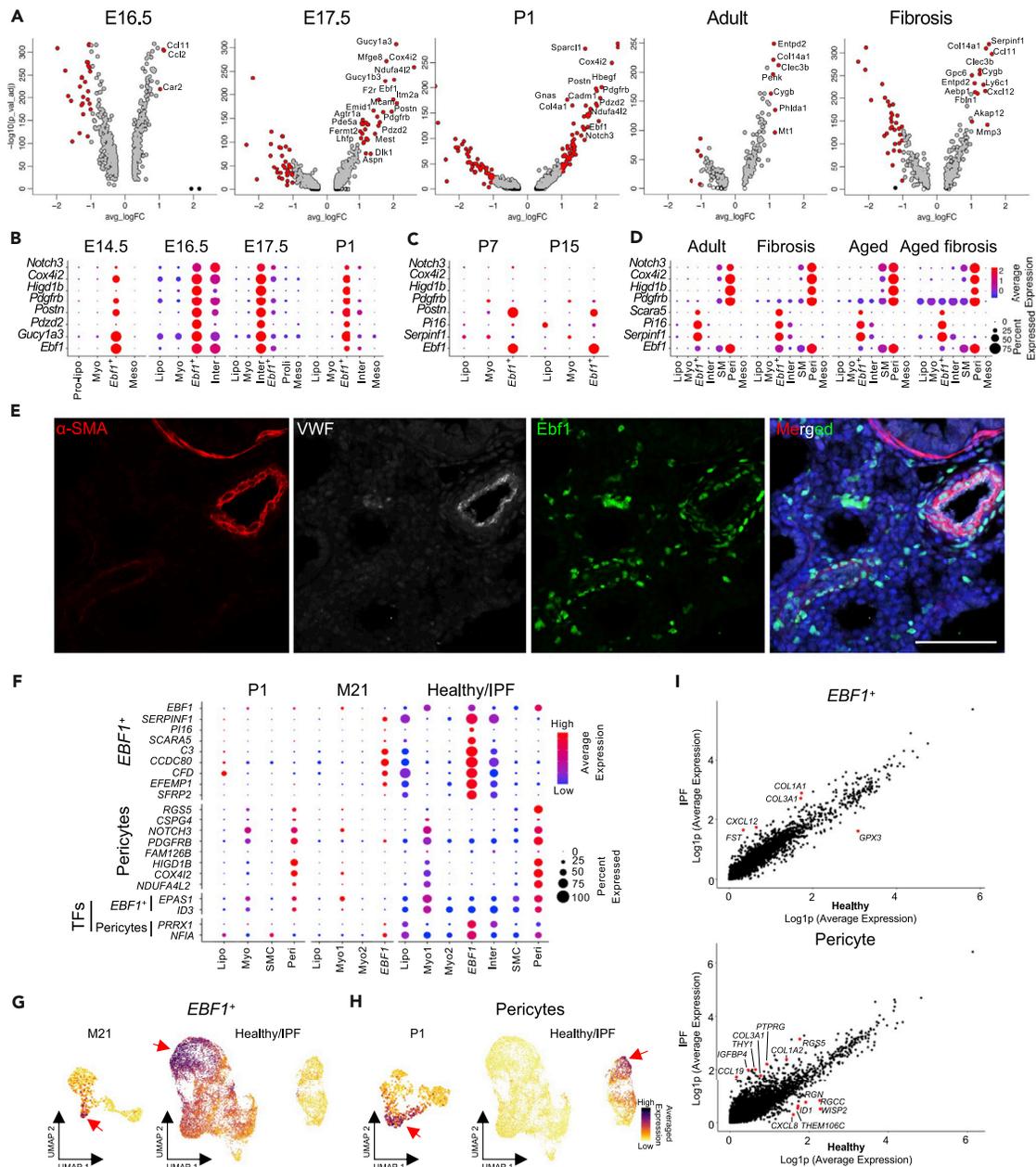


Figure 6. Identification of *Ebf1*⁺/*EBF1*⁺ and pericyte clusters

(A) Visualization of DE genes of *Ebf1*⁺ fibroblasts in mouse lungs by volcano plots. Genes in red, $p < 10^{-5}$; Avg_logFC >1. Genes in black, $p < 10^{-5}$; Avg_logFC <1. Genes in gray, $p > 10^{-5}$; Avg_logFC >1.

(B–D) Dot plots visualization of *Ebf1*⁺ fibroblast specific genes in embryonic (B) and postnatal (C) mouse lungs, and *Ebf1*⁺ fibroblasts and pericytes specific genes in adult and aged normal and fibrotic lungs (D).

(E) aSMA, VWF and *Ebf1* staining on E17.5 mouse lung section to locate *Ebf1* protein. Scale bar, 50 μ m.

(F) Dot plots visualization of *EBF1*⁺ fibroblast, known and novel pericyte specific genes and cluster specific transcription factors (TFs) in each human data set. All genes listed were $p < 10^{-5}$ and Avg_logFC >1.

(G and H) UMAP visualization of averaged expression of human *EBF1*⁺ (*CCDC80*, *SERPINF1*, *CFD*, *SCARA5*) and pericyte (*COX4I2*, *HIGD1B*, *NDUFA4L2*, *FAM162B*) signature genes in P1, M21, and integrated healthy/IPF lungs.

(I) Comparative analysis of changes in gene expression in healthy and IPF *EBF1*⁺ fibroblasts and pericyte clusters.

Pre-lipo/Lipo, Pre-lipofibroblasts/lipofibroblasts; Myo, myofibroblasts; *Ebf1*⁺/*EBF1*⁺, *Ebf1*⁺/*EBF1*⁺ fibroblasts, Inter, intermediate fibroblasts; SMC, smooth muscle cells; Peri, Pericytes; Prolif, proliferative fibroblasts; Meso, Mesothelial cells; Chon, Chondrocytes. P1, postnatal day 1; M21, 21 months.

See also [Figures S11](#), [S12](#), [S14](#), and [S15](#).

Differentiation potential of the embryonic mesenchymal cell clusters

To investigate the differentiation potential of the mesenchymal cell clusters, the mesoderm cells (E9.5–E11.5) and mesenchymal cells from E12.5 and E17.5 data sets were integrated. The integrated data were projected onto SCRAT for sample similarity and pseudotime analysis (Xie, et al., 2018). Mesodermal cells were dispersed throughout the other clusters, suggesting that the mesodermal cells may be pluripotent progenitor cells. E12.5 pre-lipofibroblasts and E17.5 lipofibroblasts were closely associated but did not integrate, suggesting a direct hierarchical relation between these two clusters (Figures 7A and 7B). This was confirmed by pseudotime analysis (Figure 7B). Myofibroblasts and intermediate fibroblasts from E12.5 integrated with the corresponding subpopulation from the E17.5 data set, suggesting that these cell types were terminally differentiated cells at the earlier embryonic stage (Figures 7A and 7B). E17.5 *Ebf1*⁺ fibroblasts were separated into two sub-clusters and displayed greater differentiation potential compared to myofibroblasts and intermediate fibroblasts (Figures 7A and 7B). It is possible that these two populations, at E17.5, are the progenitors of the corresponding population in the adult lung.

To summarize the genetic program of the mesenchymal subpopulations, we identified transcription factors and growth factors specific for each cluster that were conserved between developmental stages (Figure 7C).

Matrix gene expression in mesenchymal clusters in healthy and fibrotic lungs

We did not detect a significant increase in myofibroblast number, nor evidence of trans-differentiation of other mesenchymal populations into myofibroblasts, in either the fibrotic mouse or IPF human lungs. To investigate this further, adult and age mouse non-fibrotic and fibrotic lung mesenchymal cells were integrated, and all previously identified fibroblast subtypes were identified (Figures 7D and 7E). All mesenchymal clusters identified in both the human and mouse integrated data sets, not solely myofibroblasts, increased their expression of major ECM related genes in both species (Figures 7F–7H, S13A, and S13B). UMAP visualization confirmed the increased expression of the known matrix related genes and the novel myofibroblast markers in all subtypes in the integrated adult normal and fibrosis murine data sets (Figures S13C–13F). These data suggest that fibrotic injury increases the expression of the ECM-related genes in all mesenchymal cell subtypes.

Examination of commonly used cell type markers

Col14a1 and *Col13a1*, matrix fibroblast genes we previously reported (Xie, et al., 2018), were found to be expressed in the lipofibroblast (Figure S14A) or *Ebf1*⁺ fibroblast clusters, respectively, at different data sets (Figure S14B). We examined commonly used mesenchymal cell markers, *Pdgfra* and *Pdgfrb*, and found *Pdgfrb* expression overlapped with *Pdgfra* expression in some data sets while in others the expression of these two genes was well separated but with high background overall (Figures S14C and S14D). In adult and fibrotic mouse lungs, *Pdgfra* was well separated from *Acta2*⁺ cells and *Pdgfrb*⁺ cells but was co-expressed with *Tcf21* (Figures S15A, S15C–S15E, S15G, and S15H). *Pdgfrb* expression was preferentially expressed in one of the two *Ebf1*⁺ clusters, pericytes (Figures S15B and S15F). The expression of *Vim* (Vimentin), a frequently used mesenchymal cell marker and in some instances reportedly a gene specific for myofibroblasts (Rock, et al., 2011), was highest in endothelial cells and detectable in both mesenchymal and immune cells. *Vim* was rarely detected in epithelial cells (Figure S14E). These data suggest that *Vim* should not be used as a discriminative mesenchymal cell marker.

DISCUSSION

Our transcriptomic analysis, using both embryonic and adult tissues, encompasses nearly the full-time course of mesenchymal cell development. These data sets offer a valuable resource for identifying mesenchymal subtypes clearly in development, health, and disease. Further, these data sets provide a basis for detailed investigation and fractionation of lipofibroblasts, a controversial subtype poorly described in the literature. Finally, in these data sets we have identified a novel fibroblast subtype, defined by *Ebf1*/*EBF1* expression and provided a comprehensive resource to researchers with which to study similarities and discrepancies in each subtype across development and disease states. Our findings open new avenues of research with respect to mesenchymal fate-specification and function of specific subtypes.

Our analysis also identified major issues with the use of commonly reported markers for known mesenchymal subtypes. These data suggest all mesenchymal subtypes contribute to ECM production in fibrosis in both mouse and human and highlighted differential expression of genes linked to IPF in different subtypes. Further,

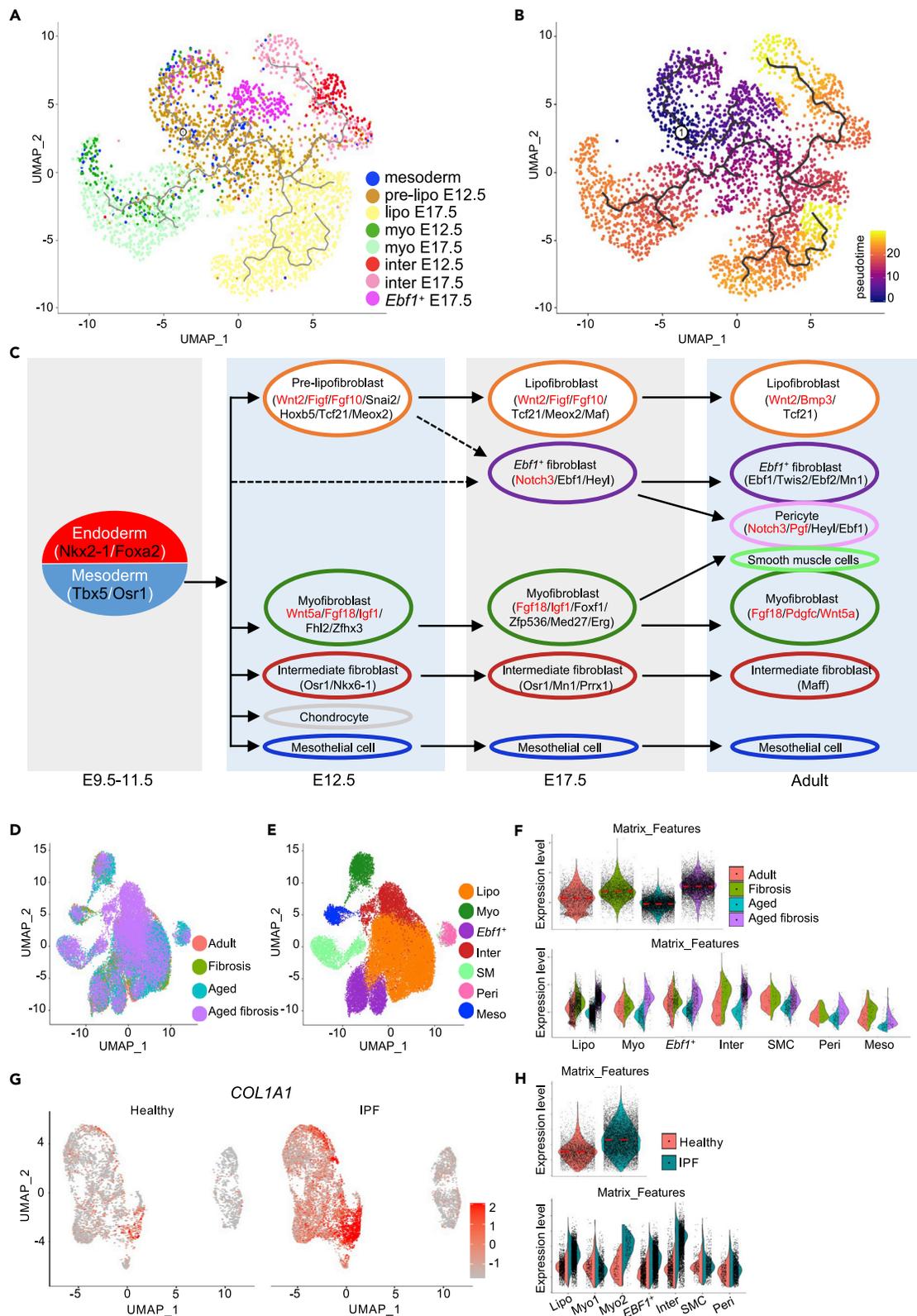


Figure 7. Differential potential of mesenchymal subtypes in development and their contribution to matrix in fibrosis

(A and B) Lineage bifurcation and differentiation potentials of mesenchymal cell subtypes in embryonic lungs.

(C) Lineage graph of mouse lung mesenchymal cell subtypes labeled by specific transcription factors and growth factors.

Figure 7. Continued

(D and E) Cell integration of adult and aged normal and fibrotic lung mesenchymal cells (D) and subtype definition (E).
 (F) Comparison of mouse Matrix_Features (average expression of *Col1a1*, *Col1a2*, *Col3a1*, *Fn1*, *Acta2*) in mouse lung total mesenchymal cells and subtypes.
 (G) UMAP visualization of *COL1A1* expression in healthy and IPF mesenchymal cells and subtypes.
 (H) Comparison of human Matrix_Features (average expression of *COL1A1*, *COL1A2*, *COL3A1*, *FN1*, *ACTA2*) in human healthy and IPF total mesenchymal cells and subtypes. Wilcoxon, $p < 2.2 \times 10^{-16}$ per comparison (F, H: upper panels).
 Pre-lipo/Lipo, Pre-lipofibroblasts/lipofibroblasts; Myo, myofibroblasts; *Ebf1*⁺/*EBF1*⁺, *Ebf1*⁺/*EBF1*⁺ fibroblasts; Inter, intermediate fibroblasts; SMC, smooth muscle cells; Peri, Pericytes; Meso, Mesothelial cells.
 See also [Figure S13](#).

these data do not lend support to the hypothesis that fibroblast subtypes trans-differentiate in the fibrotic lung into myofibroblasts. The identified fibroblast subtypes increased the expression of markers, like *Acta2*/*ACTA2*, most frequently associated with myofibroblasts, in fibrosis while remaining molecularly distinct and readily distinguishable from this subtype. However, a definitive resolution to this question will require lineage tracing of each subtype, work which is beyond the scope of the present study.

Pulmonary lipofibroblasts

Lipofibroblasts are a poorly described fibroblast subtype, frequently reported in the rodent lung but rarely in the human lung leading to controversy in the literature regarding their existence, identification, and relevance to human disease ([Tahedl, et al., 2014](#); [Rehan, et al., 2006](#)). Traditionally, lipofibroblasts have been identified histologically by the presence of intracellular lipid droplets, markers of an adipose like phenotype, enzymatic properties, characteristic cytokines, and canonical marker genes like *Plin2*, *Lpl*, and *Fgf10* among others ([El Agha, et al., 2014](#); [Rehan, et al., 2006](#)). The reliance on lipid dyes, and/or associated genes like *PLIN2*, to distinguish and quantitate lipofibroblasts in the lung is not ideal given that lipid droplets, and associated genes, exist and are expressed by a variety of pulmonary cell types ([Ntokou, et al., 2017](#); [Mochizuki, et al., 2011](#); [Besnard, et al., 2009](#); [Ochs, et al., 2004](#); [Zhang and Chawla, 2004](#); [Dvorak, et al., 1992](#)). More recently, lineage tracing studies have demonstrated *Tcf21* to be preferentially expressed in adult murine lung lipofibroblasts while previous reports suggested both *Pdgfra*⁺ and *Fgf10*⁺ lineage lung stromal cell populations included lipofibroblasts ([Park, et al., 2019](#); [Al Alam, et al., 2015](#); [Barkauskas, et al., 2013](#); [Chen, et al., 2012](#)). We found *Fgf10* expression was specific for, but lowly expressed by, murine lipofibroblasts while in agreement with Barkauskas et al. we noted that *Pdgfra*⁺ cells were lipofibroblasts in the adult murine lung ([Al Alam, et al., 2015](#); [Barkauskas, et al., 2013](#)). In humans only *TCF21* was consistently discriminative for lipofibroblasts.

We found that canonical lipofibroblast marker genes were prominent in cultured lipofibroblast-like cells and identified a population of fibroblasts clearly in the rodent lung between E16.5-E17.5, when lipofibroblasts emerge and are readily detectable, and P15, reportedly when the prevalence of lipofibroblasts in the rodent lung peaks ([Kaplan, et al., 1985](#); [Vaccaro and Brody, 1978](#)). Commonly reported genes were less effective at later developmental stages in the rodent lung and, other than *TCF21*, ineffective at identifying the lipofibroblast cluster in humans. Our novel lipofibroblast signature, in keeping with the recent lineage tracing study ([Park, et al., 2019](#)), included *Tcf21*/*TCF21* and was consistently discriminative for the associated, transcriptomically distinct, cluster of cells in all data sets. When we bulk-sequenced fibroblasts sorted using a novel lipofibroblast cell surface marker (CD249), the top DE genes overlapped substantially with the transcriptomic signature of lipofibroblasts in the scRNA-seq data. These data do suggest however that care should be taken with the use of common markers for lipofibroblasts reported in the literature. While these genes are typical of these cells pushed toward a lipofibroblast-like phenotype *in vitro*, they do not appear to be defining features of lipofibroblasts at most developmental stages.

Cultured lipofibroblasts were more supportive of alveolar type 2 cell colony formation in organoid assays than normal fibroblasts and expressed significantly greater levels of *Fgf10*, supporting the hypothesis that this population may play a role in lung repair and regeneration ([Yuan, et al., 2018](#)).

Myofibroblasts and SMCs

Myofibroblasts have long been considered the primary contributors of ECM deposition in fibrosis, and the key effector cells in IPF. A most recent study also identified a myofibroblast lineage acted as the drivers of the alveolar remodeling during the emergence of the alveolus ([Zepp, et al., 2021](#)). The definition of myofibroblasts has long relied on α SMA (*Acta2*) expression with many equating increased α SMA⁺ cells in fibrosis with contractile myofibroblasts and an expansion of this population in the fibrotic lung ([Sun, et al., 2016](#); [Rock, et al., 2011](#)). In our analysis, and as noted frequently in the literature, myofibroblasts and SMCs expressed a number of common

markers like α SMA (*Acta2*), and even SMC “specific” markers like *Myh11/MYH11* at similar levels (Rock, et al., 2011; Gan, et al., 2007; Hinz, et al., 2007; Sanders, et al., 2007; Yoshida and Owens, 2005). In fibrotic murine and human lungs, we noted increased *Acta2/ACTA2* expression in multiple mesenchymal subtypes without an associated increase in myofibroblasts. Many cells in non-myofibroblast sub-clusters from fibrotic lungs had increased expression of α SMA and other collagen genes, but the signatures of these cells still remained unchanged significantly. Gene signature-based clustering in packages like Seurat still gave similar cell clusters to non-fibrotic lungs but did not incorporate these cells into myofibroblast clusters, which denied the trans-differentiation of non-myofibroblasts into myofibroblasts. These data do not support the hypothesis that this increase can be attributed to an expansion of the myofibroblast population which was not noted in either murine or human fibrotic lungs (Rock, et al., 2011).

We successfully identified discriminative marker genes for myofibroblasts in the murine lung but not the human lung where the transcriptomic differences between these populations were either very subtle or non-existent as suggested by others (Yoshida and Owens, 2005). In the fibrotic lungs of both species the transcriptome of myofibroblasts and SMC became highly homologous, and the cells were closely associated. Myofibroblasts were *Thy1/THY1*⁺, as reported in the literature but other suggested myofibroblast delineating markers, like *S100A4* (Rock, et al., 2011; Sanders, et al., 2007; Niessen, et al., 2004), were not discriminative. Neither *Pdgfra/PDGFR*A nor *Pdgfrb/PDGFR*B expression were discriminative for myofibroblasts in either species in keeping with the previous observations (Crnkovic, et al., 2018; Hsia, et al., 2016; Rock, et al., 2011).

We identified a number of SMC-associated genes that displayed discrete expression in the SMC clusters, like *Actg2/ATCG2* in both species, *Actc1* in mice, or *MEF2C* in humans (Danopoulos, et al., 2020; Moiseenko, et al., 2017; Hinz, et al., 2007). Therefore, it was possible, using a select number of reported SMC markers, to distinguish SMCs from myofibroblasts particularly in non-fibrotic cells where they clustered distinctly. It should be noted that some genes, for instance *Hhip/HHIP* appear to be species specific. *Hhip* was specifically expressed by murine myofibroblasts. In humans, *HHIP* was expressed alongside SMC markers as recently reported (Danopoulos, et al., 2020).

Genetic lineage tracing studies of myofibroblasts, using *Fgf10*, *Axin2*, *Gli1*, *Wt1*, and SMCs have been limited by the dependence on α SMA or *Acta2* as the marker for myofibroblasts and/or SMCs (El Agha, et al., 2017; Moiseenko, et al., 2017; Zepp, et al., 2017; Xie, et al., 2016b). The contribution of these lineages to the distinct subsets is yet to be definitively resolved with reports leaning toward *Fgf10/Wt1*⁺ cells as predominantly fibroblast/mesothelial and *Gli1/Axin2* as the predominantly giving rise to myofibroblast/SMCs (Moiseenko, et al., 2017; Zepp, et al., 2017; Al Alam, et al., 2015).

Ebf1⁺ mesenchymal cells and pericytes

We identified a novel mesenchymal subpopulation characterized by *Ebf1* with a transcriptomic signature that could not be attributed to any known mesenchymal subtype. In the embryonic lung, this population co-expressed markers for pericytes. In the adult and fibrotic lungs, the *Ebf1*⁺ populations diverged and became distinct, one displayed discrete expression of known pericyte markers and the other had a unique transcriptomic signature and could be identified in most data sets. These data suggest that the novel *Ebf1*⁺ fibroblast population and pericytes may share a common developmental lineage. In the human postnatal lung, a mesenchymal population with a highly orthologous signature to the murine *Ebf1*⁺ fibroblasts were identified along with a distinct pericyte cluster.

There is little in the literature on the role of *Ebf1* in fibroblasts. However, in a recent study an *Ebf1*^{high} fibroblast population was identified as a distinct cluster in a scRNA-seq analysis of wound fibroblasts (Guerrero-Juarez, et al., 2019). Recent pre-print publications identified an “adventitial fibroblast” subtype (Mayr et al., 2020; Travaglini, et al., 2020) with a similar transcriptomic signature to the *Ebf1/EBF1* population in our study. The *in situ* hybridization localization of *SFRP2*, *SERPINF1*, *PI16* prominent genes in the *Ebf1/EBF1* cluster we identify in a recent study (Travaglini, et al., 2020) are compatible with the results of our *Ebf1* immunofluorescence localizing a proportion of *Ebf1*⁺ fibroblasts to the adventitia. *Ebf1* deletion was demonstrated to have critical effects on *Foxd1*⁺ stromal progenitors, a lineage that includes pericytes (Nelson, et al., 2019; Humphreys, et al., 2010). Further reports document that cells expressing the pericyte marker *Ng2*⁺ (*Cspg4*) require *Ebf1* for their function and a recent study reported an *Rgs5*⁺ subgroup of *PDGFR* β pericytes with a transcriptomic signature characterized by *Ebf1*, as well as *Ndufla4l2*, *Cox4i2* and *Higd1b* (Derecka, et al., 2020; Duan, et al., 2018) all genes we identify as discrete pericyte markers.

These reports are supportive of our identification of an *Ebf1*⁺/*EBF1* fibroblast population as a distinct subtype and our hypothesis that this subtype and pericytes may share a common lineage.

Commonly reported pericyte markers identified a distinct cluster of cells in the adult murine and human lungs. However, transcript expression of *Cspg4/CSPG4* and *Rgs5/RGS5*, prototypical pericyte marker genes, were low in both murine and human lung mesenchymal cells while *Pdgfrb/PDGFRB* had high background expression in almost all other mesenchymal subtypes. More novel markers we identified were expressed at greater levels and were more discriminative for pericytes.

ECM and differential gene expression in fibrotic lungs

The present study demonstrates that all identified fibroblast subpopulations, not just myofibroblasts increase their expression of transcripts for ECM components (Peysers, et al., 2019; Noble, et al., 2012; Rock, et al., 2011). Furthermore, we did not detect evidence of trans-differentiation of other mesenchymal populations into myofibroblasts, in either the fibrotic mouse or IPF donor lungs. These data are supportive of the work of a previous study, which reported a dramatic expansion of Col-EGFP⁺ cells in the bleomycin injured lung, with only a minority of cells expressing both Col-EGFP and Acta2-RFP (Sun, et al., 2016). They are also in keeping with a growing body of research challenging the assumption that α SMA is a consistent marker of collagen producing cells, and the focus on the myofibroblast as the major pathological cell type in IPF (Sun, et al., 2016; Xie, et al., 2016b; Rock, et al., 2011). Our analysis highlighted persistent downregulation of specific genes in multiple fibroblast subtypes, like *RGCC*, *WISP2* and *GPX3* and upregulation of genes like *POSTN* that are implicated in IPF suggesting specific subtypes may contribute more directly to IPF pathogenesis than others (Zhang, et al., 2014; Naik, et al., 2012).

CONCLUSION

This comprehensive resource, and definitive description of the transcriptome of all mesenchymal subtypes, will facilitate the investigation of mesenchymal cell types in development, health and disease by the academic and clinical research communities. For the first time we have provided a clear description of the fibroblast subtypes in the murine and human lung, the transcriptome of poorly described subtypes like lipofibroblasts and provided a comprehensive analysis of the efficacy of commonly reported subtype markers. In addition, these data highlight key areas for future study: the role of the novel *Ebf1/EBF1* subtype in the lung and the importance of each unique subtype in development and disease states. This comprehensive investigation provides a wealth of new markers and transcriptomic information with which to study these cell types and will enhance the research community's ability to study mesenchymal cells throughout development, health and disease.

Limitations of the study

The current analysis, although comprehensive, and the first longitudinal study of its kind are not without limitations. Given the magnitude of this analysis *in situ* localization and validation of the identified novel marker genes is ongoing. These data were not completed in time for inclusion in this manuscript. Lineage tracing using the identified novel transcription factor markers for each population is also underway and will enable fractionation of each subtype for validation and further analysis. We could not identify distinct marker genes for myofibroblasts in the human lung, and a definitive description of this population will likely require spatial RNA-sequencing analysis in order to provide this information. This analysis was focused on the mRNA level. We are yet to validate that the expression of the identified genes translates to prominent protein expression in each mesenchymal population.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Bleomycin instillation

- Mouse lung tissue isolation
 - Murine lung single cell isolation
 - *In vitro* culture of murine lung fibroblasts
 - *In vitro* 3D organoid culture with cultured lipofibroblast like cells
 - Fluorescence-activated cell sorting (FACS)
 - Human lung dissociation and cell isolation
 - Bulk RNA-seq analysis
 - Histology and immunofluorescence staining
 - scRNA-sequencing
 - scRNA-sequencing data analysis
 - Nuclei isolation for scATAC-seq
 - scATAC-seq transposition, library construction, and sequencing
 - scATAC-sequencing data analysis
 - Bioinformatics methods
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102551>.

ACKNOWLEDGMENTS

The authors thank the members of Noble and Jiang laboratory for support and helpful discussion during the course of the study. This work was supported by National Institutes of Health grants R35-HL150829, R01-HL060539, R01-AI052201, R01-HL077291 (PWN), and R01-HL122068 (D.J. and P.W.N.), and P01-HL108793 (P.W.N. and D.J.). S.C.R. was supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 797209. We thank the lung research community for the open sharing of data sets available.

AUTHOR CONTRIBUTIONS

J.L., P.W.N., and D.J. conceived the study. D.J., J.L., P.W.N., X.L., and S.C.R. designed the study. X.L. and S.C.R. performed most of the experiments and analyzed the data. J.L., G.H., X.L., C.Y., N.D., Y.W., and D.J. analyzed single-cell RNA transcriptome data. G.H. analyzed single cell RNA transcriptome data, performed flow cytometry analysis, and prepared figures. G.H., X.L., F.T., A.B., N.L., T.X., T.P., and S.C.R. took part in mouse, cell culture, and biological experiments. P.C., C.H., W.C.P., and B.S. analyzed and interpreted data. S.S.W. and J.B. provided human samples and interpreted data. X.L., S.C.R., P.W.N., and D.J. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 11, 2020

Revised: March 30, 2021

Accepted: May 14, 2021

Published: June 25, 2021

REFERENCES

- Adams, T.S., Schupp, J.C., Poli, S., Ayaub, E.A., Neumark, N., Ahangari, F., Chu, S.G., Raby, B.A., Delulius, G., Januszky, M., et al. (2020). Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* 6, eaba1983.
- Al Alam, D., El Agha, E., Sakurai, R., Kheirollahi, V., Moiseenko, A., Danopoulos, S., Shrestha, A., Schmoldt, C., Quantius, J., Herold, S., et al. (2015). Evidence for the involvement of fibroblast growth factor 10 in lipofibroblast formation during embryonic lung development. *Development* 142, 4139–4150.
- Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20, 163–172.
- Barkauskas, C.E., Counce, M.J., Rackley, C.R., Bowie, E.J., Keene, D.R., Stripp, B.R., Randell, S.H., Noble, P.W., and Hogan, B.L. (2013). Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* 123, 3025–3036.
- Besnard, V., Wert, S.E., Stahlman, M.T., Postle, A.D., Xu, Y., Ikegami, M., and Whitsett, J.A. (2009). Deletion of Scap in alveolar type II cells influences lung lipid homeostasis and identifies a compensatory role for pulmonary lipofibroblasts. *J. Biol. Chem.* 284, 4018–4030.
- Bostrom, H., Willetts, K., Pekny, M., Leveen, P., Lindahl, P., Hedstrand, H., Pekna, M., Hellstrom, M., Gebre-Medhin, S., Schalling, M., et al. (1996). PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis. *Cell* 85, 863–873.
- Chao, C.M., El Agha, E., Tiozzo, C., Minoo, P., and Bellusci, S. (2015). A breath of fresh air on the

- mesenchyme: impact of impaired mesenchymal development on the pathogenesis of bronchopulmonary dysplasia. *Front. Med. (Lausanne)* 2, 27.
- Chen, F., and Fine, A. (2016). Stem cells in lung injury and repair. *Am. J. Pathol.* 186, 2544–2550.
- Chen, L., Acciani, T., Le Cras, T., Lutzko, C., and Perl, A.K. (2012). Dynamic regulation of platelet-derived growth factor receptor alpha expression in alveolar fibroblasts during re-alveolarization. *Am. J. Respir. Cell Mol. Biol.* 47, 517–527.
- Cohen, M., Giladi, A., Gorki, A.D., Solodkin, D.G., Zada, M., Hladik, A., Miklosi, A., Salame, T.M., Halpern, K.B., David, E., et al. (2018). Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting. *Cell* 175, 1031–1044 e18.
- Crnkovic, S., Marsh, L.M., El Agha, E., Voswinckel, R., Ghanim, B., Klepetko, W., Stacher-Priehse, E., Olschewski, H., Bloch, W., Bellusci, S., et al. (2018). Resident cell lineages are preserved in pulmonary vascular remodeling. *J. Pathol.* 244, 485–498.
- Danopoulos, S., Bhattacharya, S., Mariani, T.J., and Al Alam, D. (2020). Transcriptional characterisation of human lung cells identifies novel mesenchymal lineage markers. *Eur. Respir. J.* 55, 1900746.
- Derecka, M., Herman, J.S., Cauchy, P., Ramamoorthy, S., Lupar, E., Grun, D., and Grosschedl, R. (2020). EBF1-deficient bone marrow stroma elicits persistent changes in HSC potential. *Nat. Immunol.* 21, 261–273.
- Duan, L., Zhang, X.D., Miao, W.Y., Sun, Y.J., Xiong, G., Wu, Q., Li, G., Yang, P., Yu, H., Li, H., et al. (2018). PDGFRbeta cells rapidly relay inflammatory signal from the circulatory system to neurons via chemokine CCL2. *Neuron* 100, 183–200 e8.
- Dvorak, A.M., Morgan, E., Schleimer, R.P., Ryeom, S.W., Lichtenstein, L.M., and Weller, P.F. (1992). Ultrastructural immunogold localization of prostaglandin endoperoxide synthase (cyclooxygenase) to non-membrane-bound cytoplasmic lipid bodies in human lung mast cells, alveolar macrophages, type II pneumocytes, and neutrophils. *J. Histochem. Cytochem.* 40, 759–769.
- El Agha, E., Herold, S., Al Alam, D., Quantius, J., MacKenzie, B., Carraro, G., Moiseenko, A., Chao, C.M., Mino, P., Seeger, W., et al. (2014). Fgf10-positive cells represent a progenitor cell population during lung development and postnatally. *Development* 141, 296–306.
- El Agha, E., Moiseenko, A., Kheirollahi, V., De Langhe, S., Crnkovic, S., Kwapiszewska, G., Szibor, M., Kosanovic, D., Schwind, F., Schemmuly, R.T., et al. (2017). Two-way conversion between lipogenic and myogenic fibroblastic phenotypes marks the progression and resolution of lung fibrosis. *Cell Stem Cell* 20, 261–273 e3.
- Gan, Q., Yoshida, T., Li, J., and Owens, G.K. (2007). Smooth muscle cells and myofibroblasts use distinct transcriptional mechanisms for smooth muscle alpha-actin expression. *Circ. Res.* 101, 883–892.
- Guerrero-Juarez, C.F., Dedhia, P.H., Jin, S., Ruiz-Vega, R., Ma, D., Liu, Y., Yamaga, K., Shestova, O., Gay, D.L., Yang, Z., et al. (2019). Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat. Commun.* 10, 650.
- Guo, M., Du, Y., Gokey, J.J., Ray, S., Bell, S.M., Adam, M., Sudha, P., Perl, A.K., Deshmukh, H., Potter, S.S., et al. (2019). Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat. Commun.* 10, 37.
- Habermann, A.C., Gutierrez, A.J., Bui, L.T., Yahn, S.L., Winters, N.I., Calvi, C.L., Peter, L., Chung, M.-I., Taylor, C.J., Jetter, C., et al. (2020). Single-cell RNA-sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* 6, eaba1972.
- Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the mouse cell atlas by microwell-seq. *Cell* 172, 1091–1107.e17.
- Herriges, M., and Morrisey, E.E. (2014). Lung development: orchestrating the generation and regeneration of a complex organ. *Development* 141, 502–513.
- Hinz, B., Phan, S.H., Thannickal, V.J., Galli, A., Bochaton-Piallat, M.L., and Gabbiani, G. (2007). The myofibroblast: one function, multiple origins. *Am. J. Pathol.* 170, 1807–1816.
- Hsia, L.T., Ashley, N., Oualet, D., Wang, L.M., Wilding, J., and Bodmer, W.F. (2016). Myofibroblasts are distinguished from activated skin fibroblasts by the expression of AOC3 and other associated markers. *Proc. Natl. Acad. Sci. U S A* 113, E2162–E2171.
- Humphreys, B.D., Lin, S.L., Kobayashi, A., Hudson, T.E., Nowlin, B.T., Bonventre, J.V., Valerius, M.T., McMahon, A.P., and Duffield, J.S. (2010). Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis. *Am. J. Pathol.* 176, 85–97.
- Kaplan, N.B., Grant, M.M., and Brody, J.S. (1985). The lipid interstitial cell of the pulmonary alveolus. Age and species differences. *Am. Rev. Respir. Dis.* 132, 1307–1312.
- Kugler, M.C., Loomis, C.A., Zhao, Z., Cushman, J.C., Liu, L., and Munger, J.S. (2017). Sonic hedgehog signaling regulates myofibroblast function during alveolar septum formation in murine postnatal lung. *Am. J. Respir. Cell Mol. Biol.* 57, 280–293.
- Kumar, M.E., Bogard, P.E., Espinoza, F.H., Menke, D.B., Kingsley, D.M., and Krasnow, M.A. (2014). Mesenchymal cells. Defining a mesenchymal progenitor niche at single-cell resolution. *Science* 346, 1258810.
- Lee, J.H., Tammela, T., Hofree, M., Choi, J., Marjanovic, N.D., Han, S., Canner, D., Wu, K., Paschini, M., Bhang, D.H., et al. (2017). Anatomically and functionally distinct lung mesenchymal populations marked by Lgr5 and Lgr6. *Cell* 170, 1149–1163 e12.
- Li, R., Bernau, K., Sandbo, N., Gu, J., Preissl, S., and Sun, X. (2018). Pdgfra marks a cellular lineage with distinct contributions to myofibroblasts in lung maturation and injury response. *Elife* 7, e36865.
- Mayr, C.H., Simon, L.M., Leuschner, G., Ansari, M., Geyer, P.E., Angelidis, I., Strunz, M., Singh, P., Kneidinger, N., Reichenberger, F., et al. (2020). Integrated single cell analysis of human lung fibrosis resolves cellular origins of predictive protein signatures in body fluids. *medRxiv*, 2020.01.21.20018358.
- McCulley, D., Wienhold, M., and Sun, X. (2015). The pulmonary mesenchyme directs lung development. *Curr. Opin. Genet. Dev.* 32, 98–105.
- McGowan, S.E., and McCoy, D.M. (2014). Regulation of fibroblast lipid storage and myofibroblast phenotypes during alveolar septation in mice. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 307, L618–L631.
- Mochizuki, I., Kubo, K., and Honda, T. (2011). Relationship between mitochondria and the development of specific lipid droplets in capillary endothelial cells of the respiratory tract in patients with sarcoidosis. *Mitochondrion* 11, 601–606.
- Moiseenko, A., Kheirollahi, V., Chao, C.M., Ahmadvand, N., Quantius, J., Wilhelm, J., Herold, S., Ahlbrecht, K., Morty, R.E., Rizvanov, A.A., et al. (2017). Origin and characterization of alpha smooth muscle actin-positive cells during murine lung development. *Stem Cells* 35, 1566–1578.
- Morrisey, E.E., and Hogan, B.L. (2010). Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev. Cell* 18, 8–23.
- Morse, C., Tabib, T., Sembrat, J., Buschur, K., Bittar, H.T., Valenzi, E., Jiang, Y., Kass, D.J., Gibson, K., Chen, W., et al. (2019). Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur. Respir. J.* 54, 1802441.
- Murgai, M., Ju, W., Eason, M., Kline, J., Beury, D.W., Kaczanowska, S., Miettinen, M.M., Kruhlak, M., Lei, H., Shern, J.F., et al. (2017). KLF4-dependent perivascular cell plasticity mediates pre-metastatic niche formation and metastasis. *Nat. Med.* 23, 1176–1190.
- Naik, P.K., Bozyk, P.D., Bentley, J.K., Popova, A.P., Birch, C.M., Wilke, C.A., Fry, C.D., White, E.S., Sisson, T.H., Tayob, N., et al. (2012). Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 303, L1046–L1056.
- Nelson, T., Velazquez, H., Troiano, N., and Fretz, J.A. (2019). Early B cell factor 1 (EBF1) regulates glomerular development by controlling mesangial maturation and consequently COX-2 expression. *J. Am. Soc. Nephrol.* 30, 1559–1572.
- Niessen, P., Clement, S., Fontao, L., Chaponnier, C., Teunissen, B., Rensen, S., van Eys, G., and Gabbiani, G. (2004). Biochemical evidence for interaction between smoothelin and filamentous actin. *Exp. Cell Res.* 292, 170–178.
- Noble, P.W., Barkauskas, C.E., and Jiang, D. (2012). Pulmonary fibrosis: patterns and perpetrators. *J. Clin. Invest.* 122, 2756–2762.

- Ntokou, A., Szibor, M., Rodriguez-Castillo, J.A., Quantius, J., Herold, S., El Agha, E., Bellusci, S., Salwig, I., Braun, T., Voswinckel, R., et al. (2017). A novel mouse Cre-driver line targeting Perilipin 2-expressing cells in the neonatal lung. *Genesis* 55.
- Ochs, M., Fehrenbach, H., and Richter, J. (2004). Occurrence of lipid bodies in canine type II pneumocytes during hypothermic lung ischemia. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 277, 287–297.
- Parimon, T., Yao, C., Habel, D.M., Ge, L., Bora, S.A., Brauer, R., Evans, C.M., Xie, T., Alonso-Valenteen, F., Medina-Kauwe, L.K., et al. (2019). Syndecan-1 promotes lung fibrosis by regulating epithelial reprogramming through extracellular vesicles. *JCI Insight* 5, e129359.
- Park, J., Ivey, M.J., Deana, Y., Riggsbee, K.L., Sorensen, E., Schwabl, V., Sjoberg, C., Hjertberg, T., Park, G.Y., Swonger, J.M., et al. (2019). The Tcf21 lineage constitutes the lung lipofibroblast population. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 316, L872–L885.
- Peysner, R., MacDonnell, S., Gao, Y., Cheng, L., Kim, Y., Kaplan, T., Ruan, Q., Wei, Y., Ni, M., Adler, C., et al. (2019). Defining the activated fibroblast population in lung fibrosis using single-cell sequencing. *Am. J. Respir. Cell Mol. Biol.* 61, 74–85.
- Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 490–495.
- Raredon, M.S.B., Adams, T.S., Suhail, Y., Schupp, J.C., Poli, S., Neumark, N., Leiby, K.L., Greaney, A.M., Yuan, Y., Hori, C., et al. (2019). Single-cell connectomic analysis of adult mammalian lungs. *Sci. Adv.* 5, eaaw3851.
- Rehan, V.K., Sugano, S., Wang, Y., Santos, J., Romero, S., Dasgupta, C., Keane, M.P., Stahlman, M.T., and Torday, J.S. (2006). Evidence for the presence of lipofibroblasts in human lung. *Exp. Lung Res.* 32, 379–393.
- Reyffman, P.A., Walter, J.M., Joshi, N., Anekalla, K.R., McQuattie-Pimentel, A.C., Chiu, S., Fernandez, R., Akbarpour, M., Chen, C.-I., Ren, Z., et al. (2019). Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 199, 1517–1536.
- Robin, Y.M., Penel, N., Perot, G., Neuville, A., Velasco, V., Ranchere-Vince, D., Terrier, P., and Coindre, J.M. (2013). Transgelin is a novel marker of smooth muscle differentiation that improves diagnostic accuracy of leiomyosarcomas: a comparative immunohistochemical reappraisal of myogenic markers in 900 soft tissue tumors. *Mod. Pathol.* 26, 502–510.
- Rock, J.R., Barkauskas, C.E., Cronic, M.J., Xue, Y., Harris, J.R., Liang, J., Noble, P.W., and Hogan, B.L. (2011). Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc. Natl. Acad. Sci. U S A* 108, E1475–E1483.
- Sanders, Y.Y., Kumbala, P., and Hagood, J.S. (2007). Enhanced myofibroblastic differentiation and survival in Thy-1(-) lung fibroblasts. *Am. J. Respir. Cell Mol. Biol.* 36, 226–235.
- Sia, D., Torres-Martin, M., Wang, H., Haber, P., Puigvehi, M., Montironi, C., Maeda, M., Wei-Qiang, L., Harrod, E., Taik, P., et al. (2019). Decoding the immune tumor microenvironment of hepatocellular carcinoma. *Hepatology* 70, 1133a–1134a.
- Sun, K.H., Chang, Y., Reed, N.I., and Sheppard, D. (2016). alpha-Smooth muscle actin is an inconsistent marker of fibroblasts responsible for force-dependent TGFbeta activation or collagen production across multiple models of organ fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 310, L824–L836.
- Tahedi, D., Wirkes, A., Tschanz, S.A., Ochs, M., and Muhlfeld, C. (2014). How common is the lipid body-containing interstitial cell in the mammalian lung? *Am. J. Physiol. Lung Cell. Mol. Physiol.* 307, L386–L394.
- Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J., et al. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587, 619–625.
- Vaccaro, C., and Brody, J.S. (1978). Ultrastructure of developing alveoli. I. The role of the interstitial fibroblast. *Anat. Rec.* 192, 467–479.
- Valenzi, E., Bulik, M., Tabib, T., Morse, C., Sembrat, J., Trejo Bittar, H., Rojas, M., and Lafyatis, R. (2019). Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann. Rheum. Dis.* 78, 1379–1387.
- Xie, T., Liang, J., Liu, N., Huan, C., Zhang, Y., Liu, W., Kumar, M., Xiao, R., D'Armiento, J., Metzger, D., et al. (2016a). Transcription factor TBX4 regulates myofibroblast accumulation and lung fibrosis. *J. Clin. Invest.* 126, 3063–3079.
- Xie, T., Liang, J., Liu, N., Huan, C., Zhang, Y., Liu, W., Kumar, M., Xiao, R., D'Armiento, J., Metzger, D., et al. (2016b). Transcription factor TBX4 regulates myofibroblast accumulation and lung fibrosis. *J. Clin. Invest.* 126, 3626.
- Xie, T., Wang, Y., Deng, N., Huang, G., Taghavifar, F., Geng, Y., Liu, N., Kulur, V., Yao, C., Chen, P., et al. (2018). Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep.* 22, 3625–3640.
- Yoshida, T., and Owens, G.K. (2005). Molecular determinants of vascular smooth muscle cell diversity. *Circ. Res.* 96, 280–291.
- Yuan, T., Volckaert, T., Chanda, D., Thannickal, V.J., and De Langhe, S.P. (2018). Fgf10 signaling in lung development, homeostasis, disease, and repair after injury. *Front. Genet.* 9, 418.
- Zepp, J.A., Morley, M.P., Loebel, C., Kremp, M.M., Chaudhry, F.N., Basil, M.C., Leach, J.P., Liberti, D.C., Niethamer, T.K., Ying, Y., et al. (2021). Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. *Science* 371, eaabc3172.
- Zepp, J.A., Zacharias, W.J., Frank, D.B., Cavanaugh, C.A., Zhou, S., Morley, M.P., and Morrissey, E.E. (2017). Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* 170, 1134–1148 e10.
- Zhang, L., and Chawla, A. (2004). Role of PPARgamma in macrophage biology and atherosclerosis. *Trends Endocrinol. Metab.* 15, 500–505.
- Zhang, L., Li, Y., Liang, C., and Yang, W. (2014). CCN5 overexpression inhibits profibrotic phenotypes via the PI3K/Akt signaling pathway in lung fibroblasts isolated from patients with idiopathic pulmonary fibrosis and in an in vivo model of lung fibrosis. *Int. J. Mol. Med.* 33, 478–486.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Biotin anti-mouse CD31	Biologend	Cat# 102404; RRID: AB_312899
Rat monoclonal anti-mouse CD31	BD Biosciences	Cat# 551262; RRID: AB_398497
APC rat anti-mouse CD45	BD Biosciences	Cat# 559864; RRID: AB_398672
FITC mouse anti-mouse CD45.2	BD Biosciences	Cat# 553772; RRID: AB_395041
FITC monoclonal anti-mouse CD326	Invitrogen	Cat# 11-5791-82; RRID: AB_11151709
PE/Cy7 monoclonal anti-mouse CD326	Biologend	Cat# 118216; RRID: AB_1236471
PE mouse anti-mouse CD249	BD Biosciences	Cat# 553735; RRID: AB_395018
PE Mouse IgG2a, κ Isotype Control	BD Biosciences	Cat# 553457; RRID: AB_394871
Streptavidin PE Conjugate	eBiosciences	Cat# 12-4317-87
Alexa Fluor® 647 anti-human CD326	Biologend	Cat# 324212; RRID: AB_756086
PE/Cy7 anti-human CD31	Biologend	Cat# 303118; RRID: AB_2247932
PE/Cy7 anti-human CD45	Biologend	Cat# 304016; RRID: AB_314404
Fixable Viability Dye eFluor™ 780	Invitrogen	Cat# 65-0865-14
Fixable Viability Dye eFluor™ 506	Invitrogen	Cat# 65-0866-14
Goat polyclonal anti-human/mouse EBF-1	R&D	Cat# AF5165; RRID: AB_2097398
Rabbit polyclonal anti-vWF	abcam	Cat# ab6994; RRID: AB_305689
Mouse monoclonal anti- α -Actin	Santa Cruz	Cat# sc-32251; RRID: AB_262054
AlexaFluor 488 donkey anti-goat IgG (H+L)	JacksonImmuno	Cat# 705-545-003; RRID: AB_2340428
AlexaFluor 647 donkey anti-rabbit IgG(H+L)	Thermo Fisher	Cat# A-31573; RRID: AB_2536183
AlexaFluor 555 donkey anti-mouse IgG(H+L)	Thermo Fisher	Cat# A-31570; RRID: AB_2536180
Chemicals, Peptides, and Recombinant Proteins		
Bleomycin	Hospira	Cat# NDC 61703-332-18
DPBS (1X)	Thermo fisher	Cat# 14190144
Deoxyribonuclease I	Sigma Aldrich	Cat# D4527-20KU
Collagenase, Type 4	Worthington	Cat# LS004209
DMEM (1X)	Thermo fisher	Cat# 11965092
HyClone™ Fetal Bovine Serum (FBS)	Cytiva	Cat# SH30071.03
HBSS	Thermo fisher	Cat# 14175103
HEPES	Thermo fisher	Cat# 15630106
RBC Lysis Buffer (10X)	Ebioscience	Cat# 420301
Dispase II (neutral protease, grade II)	Sigma Aldrich	Cat# 4942078001
DMEM/F12	Thermo fisher	Cat# 11330057
Insulin-Transferrin-Selenium (ITS -G) (100X)	Thermo fisher	Cat# 41400045
ITS+3 Liquid Media Supplement (100x)	Sigma Aldrich	Cat# I2771
Elastase, Suspension	Worthington	Cat# LS002279
Rosiglitazone	Sigma Aldrich	Cat# R2408
SB4315442	R&D	Cat# 1614
rhBMP4	R&D	Cat# 314-BP
Matrigel	Corning	Cat# 354230
DAPI	Thermo fisher	Cat# 62247

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BODIPY™ 493/503	Invitrogen	Cat# D3922
Nile Red	Thermo fisher	Cat# N1142
Oil Red O	Sigma Aldrich	Cat# O0625
Tissue-Tek® O.C.T. Compound	Sakura	Cat# 4583
RNeasy Micro Kit	Qiagen	Cat# 74004
Experimental Models: Organisms/Strains		
Adult and aged C57BL/6J mice	Jackson Labs	Stock No: 000664
Healthy and IPF lung tissues	Cedars-Sinai	N/A
Software and Algorithms		
R Studio Version 1.2.5033		https://rstudio.com/
Seurat v3.2		https://satijalab.org/seurat/v3.2/pbmc3k_tutorial.html
Cell Ranger v3.0	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/count
FlowJo Version 10.6.1	BD Biosciences	https://www.flowjo.com/solutions/flowjo/downloads
NextSeq500	Illumina	https://www.illumina.com/systems/sequencing-platforms/nextseq.html
Zeiss LSM 780 confocal microscope system	Zeiss	https://www.zeiss.com/microscopy/us/service-support/glossary/nlo.html
LSR Fortessa™ Flow Cytometer	BD Biosciences	https://www.bdbiosciences.com/en-in/instruments/research-instruments/research-cell-analyzers/lrsfortessa
FACS Aria™ III Cell Sorter	BD Biosciences	https://www.bdbiosciences.com/en-us/instruments/research-instruments/research-cell-sorters/facsaria-iii
Leica® CM1900 Cryostat	Leica	https://www2.leicabiosystems.com/

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dianhua Jiang (Dianhua.Jiang@CSHS.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The GEO accession numbers for mouse lung raw and processed scRNA-seq and scATAC-seq data generated for this paper are listed below: single cell RNA-seq on E16.5 mouse lung Tbx4-lineage⁺, a-SMA⁺ cells, and E17.5 mouse lungs, GSE156329; single cell RNA-seq on E17.5 mouse lungs and tracheas, GSE157654; single cell ATAC-seq on E17.5 mouse lungs, GSE157378; single cell RNA-seq on adult mouse normal lungs and fibrotic lungs, GSE131800 and GSE104154; single cell RNA-seq on sorted mesenchymal cells from aged mouse lungs and aged fibrotic mouse lungs, GSE157379; bulk RNA-seq on flow sorted CD249⁺ and CD249⁻ adult mouse lung mesenchymal cells, GSE157320; single cell RNA-seq on cultured control and stimulated mouse lung lipofibroblast-like fibroblasts, GSE157377. Published datasets include single cell RNA-seq on E9.5-E11.5 mouse lung, GSE87038; single cell RNA-seq on CD45⁻ cells from E12.5 mouse lungs, GSE119228; single cell RNA-seq on E14.5 mouse lungs, GSE108097; single cell RNA-seq on P1

mouse lungs, GSE122332; single cell RNA-seq on Pdgfra-GFP⁺ cells from P7 and P15 mouse lung, GSE118555; single cell RNA-seq on adult mouse lung, GSE111664, GSE133747, and GSE121611.

The GEO accession numbers for human lung raw and processed scRNA-seq data generated for this paper are listed below: single cell RNA-seq on adult healthy and IPF human lungs, GSE157376. Published datasets include single cell RNA-seq on P1 and M21 human lung, LungMAP: <https://lungmap.net/>; single cell RNA-seq on adult human lung, GSE135893, GSE128033, GSE122960 and GSE128169; single cell RNA-seq on IPF human lung, GSE135893, GSE128033 and GSE122960.

Codes for data processing and analysis are available upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All the human lung tissues were collected from Cedars-Sinai Medical Center and data on both male and female donors were used for analysis (Table S2). The use of human tissues for research were approved by the Institutional Review Board (IRB) of Cedars-Sinai Medical Center and were under the guidelines outlined by the IRB (Pro00032727). All animal experiments performed in this study were approved by Cedars-Sinai Medical Center Institutional Animal Care and Use Committee (IACUC008529). Both male and female adult (12-16 weeks old) and aged (82-95 weeks old) C57BL/6J mice were used for this study (Table S1).

METHOD DETAILS

Bleomycin instillation

Under anesthesia the trachea was surgically exposed. 1.25U/kg bleomycin in 25 μ l PBS was instilled into the mouse trachea with a 25-G needle inserted between the cartilaginous rings of the trachea. Control animals received saline alone. The tracheostomy site was sutured, and the animals monitored intensively until active. Animals were randomly allocated to control or treatment groups. Bleomycin treated mice were actively monitored by trained animal welfare staff until sacrificed. Mice were sacrificed at indicated and lung tissues were collected.

Mouse lung tissue isolation

Wild-type C57/Bl6J mice from an in-house colony were used in all experiments. Animals were randomly assigned to treatment groups. Animals of both genders were used without bias. Mice were considered adult at 8- to 12-weeks-old and aged at between 82- to 95-weeks-old. All mice had access to autoclaved water and pelleted mouse diet *ad libitum* were housed in a pathogen free facility at Cedars-Sinai Medical Center. For the isolation of embryonic murine lung tissues breeding cages; containing a male and two female mice, were monitored intensively following the addition of the male to the breeding cage. The presence of a female with a vaginal plug was considered embryonic day 0.5 (E0.5). Adult (12-16 weeks old) aged (82-95 weeks old), or pregnant mice were deeply anaesthetized by intraperitoneal injection (I.P.) of Ketamine (100mg/kg) and Xylazine (10 mg/kg) followed by exsanguination. Adequate depth of anesthesia was determined by lack of a withdrawal reflex to paw, followed by tail pinch, prior to the start of any surgical intervention. In adult mice the lungs were cleared of blood by PBS through the pulmonary artery via cardiac puncture prior to isolation. Pregnant mice were euthanized at the indicated time and the embryos were quickly isolated after removal of the uterus, and the lungs of the embryos were resected. The lungs of embryos, adult, and aged mice were transferred to a 15ml tube containing ice-cold PBS and processed immediately.

Murine lung single cell isolation

Murine lung tissues were dissociated using a standard protocol in our laboratory. Detailly, Isolated tissues were taken immediately to a sterile laminar flow tissue culture hood where they were, rinsed in fresh PBS and then minced finely using scissors in a 100mm² Petri dish. The minced lung tissue was then suspended in a digestion media containing 0.125% vol/vol Trypsin-EDTA, 1mg/ml Bovine Serum Albumin, 100 U/ml DNase 1, 1 mg/ml Collagenase IV and transferred a tissue culture incubator at 37°C for 30 minutes. At 10-minute intervals the lung digestion solution was titrated 10 times using a 10 ml glass pipette. Following the incubation period, the supernatant and remaining tissue was passed through a 100 μ m strainer into a 50ml tube. The strainer was washed with DMEM containing 10% vol/vol FBS. The tube was then centrifuged at 1600 rpm for 10 minutes at 4°C and the pellet resuspended in HBSS containing 0.2 mM EGTA, 10mM HEPES, 2% vol/vol FBS and 1% vol/vol antibiotic-antimycotic referred to hereafter as HBSS⁺. Red blood

cells were preferentially lysed by treating the isolated cells with 1X RBC lysis buffer for 45 seconds, followed by immediate dilution in 20 ml HBSS⁺. Cells were centrifuged again and resuspended in fresh HBSS⁺ prior to fluorescence-activated cell sorting (FACS).

In vitro culture of murine lung fibroblasts

Adult (10-12 weeks old) male and female murine lungs were dissociated as described above. The entire cell suspension was then plated on appropriately sized tissue culture plasticware. Fibroblasts were cultured in advanced DMEM/F12 (#12634010, Thermo Fisher Scientific), with 10% vol/vol FBS or to induce a lipofibroblast like phenotype with the addition of 1% vol/vol ITS+3 liquid media supplement (#12771, Merck Milipore) to the culture media. Media was changed every other day and the cells were sub-cultured when 80% confluent. Fibroblasts cultured to passage three (P3) were considered free of contaminating cells based on previous experience in the laboratory. At P3 the cultured fibroblasts were then stimulated with one or a combination (as indicated) of 10 μ M rosiglitazone (#72622, Stem Cell Technologies), 1 μ M SB431542 (#1614, R&D Systems), 4 μ M rhBMP4 (#314-BP, R&D Systems) for 14 days with the media changed every other day. These stimulations were previously reported by others to induce a lipofibroblast-like phenotype *in vitro* (Literatures cited in main text). Matched control cells isolated from the same lung were stimulated in an identical media, for an identical duration, with the appropriate vehicle.

In vitro 3D organoid culture with cultured lipofibroblast like cells

Fibroblasts in which a lipofibroblast like phenotype had been induced *in vitro*, and their associated control cells, were cultured in Matrigel/Medium (1:1) mixture in the presence of flow sorted type 2 epithelial cells (Cd45⁻ Cd31⁺ Cd24⁻ Cd34⁺ Sca1⁻ CD326⁺). 100 μ l Matrigel/medium mix containing 3×10^3 AEC2 cells and 2×10^5 control or lipofibroblast like cells were plated into each 24 well 0.4 μ m Transwell insert. 400 μ l of medium was added in the lower chambers. Half of the media in each well was changed every other day. Cultures were maintained in humidified 37°C and 5% CO₂ incubator. Colonies were visualized with a Zeiss Axiovert40 inverted fluorescent microscope. Number of colonies with a diameter of $\geq 50 \mu$ m from each insert was counted and colony forming efficiency (CFE) was determined by the number of colonies in each culture as a percentage of input epithelial cells at day 14 after plating.

Fluorescence-activated cell sorting (FACS)

For staining cells were counted and diluted to a maximum of 10^6 cells/100 μ l in HBSS⁺. The cell suspension was then divided into Eppendorf's for staining. Controls included unstained, single color controls and where appropriate isotype controls. The cells were pelleted in a pre-cooled (4°C) centrifuge at 1600 rpm for 5 minutes and resuspended in appropriately diluted primary antibody. The cell suspension was incubated with the antibody and live/dead maker (fixable viability) on ice in the dark for between 30 minutes to 1 hour. The cells were then washed by gentle pipetting with HBSS⁺. The cell suspension was centrifuged, the supernatant carefully removed, and the cells washed again. In total three wash steps were performed after each incubation. If the antibody was directly conjugated following the final wash the cells were resuspended in 500 μ l HBSS⁺ and passed through a cell strainer cap into a falcon test tube by centrifugation prior to FACS. If the antibody was not directly conjugated following the final wash the cells were incubated with appropriate secondary antibody for 30 minutes to 1 hour on ice protected from light. The cells were then washed as described previously and resuspended prior to analysis. The antibodies were used at optimized concentrations and are listed in Key resource table. Where DAPI was used as the live/dead marker it was added to the cell suspension prior 5-10 minutes prior to sorting. Total mesenchymal cells were analyzed using LSR Fortessa™ Flow Cytometer or sorted using a 13-color BD FACSAria™ III Cell Sorter from both mice and human single cell lung suspensions using negative selection as: live cells (fixable viability or DAPI negative) and EPCAM/PECAM1/PTPRC negative.

Human lung dissociation and cell isolation

Freshly isolated human lung tissues were obtained from Cedars-Sinai Medical Center and UCLA and were dissociated using a standard protocol in our laboratory. Lung tissue was taken to a sterile tissue culture hood, transferred to a Petri dish and rinsed in PBS. Airways >2 mm were resected from the surrounding tissues and discarded along with the visceral pleura. The remaining tissue was finely minced with a scissors and then a straight razor blade. The minced lung tissue was then washed in DMEM/F12 media at 4°C for 20 minutes to remove blood and then centrifuged at 600 rpm for 5 minutes in a pre-cooled centrifuge. The media was removed, and the tissue transferred to a 50 ml conical tube containing 2 mg/ml Dispase

II in DMEM/F12 overnight at 4°C with gentle agitation. The next day the suspension was heated to 37°C for 30 minutes, and the centrifuged for 5 minutes at 4°C. The supernatant was removed, and any large pieces of tissue finely minced again with a straight razor blade. The tissue was then titrated in a digestion media containing 10 U/ml elastase and incubated for 30 minutes at 37°C. An equal volume of HBSS⁺ was then added, the solution titrated and then centrifuged at (600 g, 5 minutes, 4°C). The supernatant was removed, and the tissue incubated at 37°C for 15 minutes with DNase I solution. The suspension was titrated and transferred to a 70 m cell strainer over a new 50 ml tube. The strainer was rinsed three times with 10 ml HBSS⁺. The suspension was centrifuged (600 g, 5 minutes, 4°C) and the cells resuspended in 1 X RBS lysis buffer for 2 minutes on ice, the solution diluted with HBSS then centrifuged (600 g, 5 minutes, 4°C). The supernatant was removed, and the cells resuspended in appropriate solution for further analysis.

Bulk RNA-seq analysis

Total RNA was extracted from CD249⁺ and CD249⁻ fibroblasts flow sorted from murine lungs using the RNeasy Micro Kit according to manufacturer's instructions. Total RNA was stored at -80°C until the day of analysis. RNA integrity of an aliquot from each sample was analyzed using a Bioanalyzer and only samples with a RIN ≥ 8 retained for analysis. A minimum of 50 ng and maximum of 400 ng total RNA in a maximum of 20 μ l was sequenced (1x75 bp single-end sequencing, average 25 million reads/sample) using a NextSeq 500. Library preparation, library QC and differential expression analysis was performed by the Cedars-Sinai Genomics Core facility.

Histology and immunofluorescence staining

To prepare the mouse lungs for histology after being deeply anaesthetized and sacrificed as described previously, the trachea was cannulated. The left lung cleared of blood by perfusion of the pulmonary artery with PBS via cardiac puncture. The lungs were then inflated with 0.5ml pf 10% neutral buffered formalin. The tissues were fixed overnight, and the following day embedded in Optimal Cutting Temperature Compound and flash frozen. Cryosections (5 μ m) were cut using a cryostat onto Superfrost Plus Microscope Slides. Immunofluorescence was performed using primary antibodies raised against the following antigens and used at the indicated dilutions to stain slides overnight at 4°C: α -Smooth Muscle - Cy3TM, Ebf1, Von Willebrand Factor, VWF.

To stain intracellular lipid droplets in lipofibroblast like cells and controls the media was removed and the cells washed with PBS. The cells incubated with 10 μ M Bodipy 493/505 or 1 mg/ml Nile Red, protected from light in a tissue culture incubator at 37°C. The Bodipy solution was removed, and the cells washed twice with PBS. The cells were fixed *in situ* with 4% vol/vol formaldehyde at room temperature for 15 minutes protected from light and stained with 10 μ g/ml DAPI for 10 minutes prior to imaging. For Oil Red O staining after fixation the cells were dehydrated with 100% 1, 2-Propanediol solution for 5 m minutes. This step was repeated and then 2 ml/ 10 cm² 0.5% Oil Red O solution diluted in 1, 2-Propanediol solution was added and incubated for 30 minutes at 37°C. The stain was aspirated and differentiated with 85% 1, 2-Propanediol solution for 1 minute. The cells were then rinsed with dH₂O 2-3 times, counterstained with Mayer's Hematoxylin for 10-15 minutes at room temperature. The cells were rinsed 4-5 times with dH₂O and then imaged. Stained sections were imaged using Zeiss 780 reverse Laser Scanning Confocal Microscope.

scRNA-sequencing

mRNA from single cells sorted from lung into lysis plates was reverse transcribed to complementary DNA (cDNA) and amplified. Library preparation and sequencing were performed. Sequencing libraries for cDNA from single cells were prepared as per the Single Cell 3' v2 Reagent Kits User Guide (10x Genomics, Pleasanton, CA, USA). Cellular suspensions were loaded on a Chromium Controller instrument (10x Genomics) to generate single-cell Gel Bead-In-EMulsions (GEMs). GEM-reverse transcription (RT) was performed in a Veriti 96-well thermal cycler (Thermo Fisher Scientific, Waltham, MA, USA). GEMs were collected and the cDNA was amplified and purified with SPRIselect Reagent Kit (Beckman Coulter, Brea, CA, USA). Indexed sequencing libraries were constructed using Chromium Single-Cell 3' Library Kit for enzymatic fragmentation, end-repair, A-tailing, adapter ligation, ligation cleanup, sample index PCR, and PCR cleanup. The bar-coded sequencing libraries were quantified by quantitative PCR using the KAPA Library Quantification Kit for Illumina platforms (KAPA Biosystems, Roche Holding AG, Basel, Switzerland). Sequencing libraries were loaded on a NextSeq500 with a custom sequencing setting (26bp for Read 1 and 98bp for Read 2) to obtain a sequencing depth of ~200K reads per cell.

scRNA-sequencing data analysis

Detailed scRNA-seq analysis could be found in below [Bioinformatics Methods](#). The demultiplexed raw reads were aligned to the transcriptome using STAR (version 2.5.1) with default parameters, using human GRCh38 (or mouse mm10) transcriptome reference from Ensembl version 84 annotation, containing all protein coding and long non-coding RNA genes. Expression counts for each gene in all samples were collapsed and normalized to unique molecular identifier counts using Cell Ranger software version 3.0 (10X Genomics). The result is a large digital expression matrix with cell barcodes as rows and gene identities as columns. Seurat suite version 3.0 was used for downstream analysis. Quality control before analysis on each individual sample were performed on "nFeature_RNA", "nCount_RNA" and "percent_mt" in each cell. For clustering, principal component analysis (PCA), T-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) were performed for dimension reduction. Batch correction was performed if sample integration was needed. Trajectory analysis was performed by package monocle3. The bioinformatics methodology is described in full in below [Bioinformatics Methods](#). Details on the cell numbers pre- and post-QC and the proportion of cells in each of the major factions (Immune, Endothelial, Epithelial, Mesenchymal) in murine and human lung datasets can be seen in Supplementary Table.

Nuclei isolation for scATAC-seq

Dissociated cells resuspended in PBS + 0.04% BSA were treated with DNase I prior to nuclei isolation. Specifically, cells were pelleted, the supernatant removed, and the cell pellet resuspended in DNase Solution, 0.1 U/ μ l DNase I. Cells were then incubated for 5 mins, pelleted, and washed twice with PBS + 0.04% BSA. The cell suspension was passed through a 70 μ m Flowmi Cell Strainer, and the cell concentration determined before proceeding to nuclei isolation. For nuclei isolation, up to 1×10^6 cells were pelleted, followed by removal of the supernatant, and resuspension in chilled 0.1X Lysis Buffer, 0.1% BSA for a pre-determined optimal cell lysis time. Wash Buffer was added to dilute the lysis buffer, then the nuclei were immediately pelleted. The supernatant was removed, and the isolated nuclei were resuspended in 1X Nuclei Buffer. Finally, nuclei concentration was determined and scATAC-Seq transposition and library construction was immediately performed.

scATAC-seq transposition, library construction, and sequencing

scATAC-Seq transposition and library construction was performed according to the manufacturer's protocol using the Chromium Next GEM Single Cell ATAC v 1.1 reagent kit (10x Genomics). Transposition of single nuclei obtained in the previous step was performed in bulk, followed by capture of the transposed, single nuclei into GEMs (Gel Bead-In-Emulsions) using the Chromium Controller (10x Genomics). GEM cleanup and library size selection was performed using Dynabeads MyOne Silane beads (10x Genomics) and SPRIselect reagent, respectively. Sample index PCR was performed for 10 cycles. Indexed sequencing libraries were quantified by qPCR using the Colibri Library Quantification Kit for Illumina platforms. Libraries were sequenced on the NovaSeq at 2x50bp at a sequencing depth of ~25K reads per nuclei.

scATAC-sequencing data analysis

Cells from E17.5 murine lung were isolated in the same way as cell isolation for scRNA-seq. Cell nuclei isolation and library preparation were described above. Raw sequencing data is demultiplexed and converted to fastq format by using bcl2fastq v2.20. Cell Ranger ATAC software v1.1.0 (10X Genomics) is used for barcodes identification, reads alignment, duplicate marking, peak calling and cell calling with default parameter. Briefly, each barcode sequence is checked against a 'whitelist' of correct barcode sequences, and the frequency of each whitelist barcode is counted. Raw reads are aligned to the human reference genome GRCm38 using BWA-MEM with default parameters, then duplicated reads that have identical mapping positions on the reference are marked. For peak calling, the number of transposition events at each base-pair along the genome is counted, then signal above a threshold are determined as peak signal after modeling. For cell calling, barcodes with high fraction of fragments overlapping called peak are selected, then odds ratio of 100000 is used to separate the barcodes that correspond to real cells from the non-cell barcodes. Finally, a count matrix is generated consisting of the counts of fragments ends within each peak region for each barcode. For further QC, clustering and gene accessibility visualization were performed following online vignette (https://satijalab.org/signac/articles/mouse_brain_vignette.html). Briefly, a Seurat object was generated on count matrix and fragments, and QC was performed by removing cells that are outliers for QC metrics: pct_reads_in_peaks, peak_region_fragments, blacklist_ratio, nucleosome_signal. After

normalization and linear dimensional reduction, non-linear dimension reduction and clustering, gene accessibilities were visualized by UMAP. Cell types and mesenchymal cell sub-clusters were defined by checking unknown cell type markers.

Bioinformatics methods

Read alignments. The demultiplexed raw reads were aligned to the transcriptome using STAR (version 2.5.1) with default parameters, using human GRCh38 (or mouse mm10) transcriptome reference from Ensembl version 84 annotation, containing all protein coding and long non-coding RNA genes. Expression counts for each gene in all samples were collapsed and normalized to unique molecular identifier counts using Cell Ranger software version 3.0 (10X Genomics). The result was a large digital expression matrix with cell barcodes as rows and gene identities as columns.

Quality control, cell clustering, doublet calling and annotation. Expression profiles of cells from different subjects and publicly available datasets were analyzed and clustered separately using the R software package Seurat (version 3.0). For each individual sample the total number of genes detected per cell ('nFeature_RNA'), number of transcripts per cell ('nCount_RNA') and percentage of transcripts mapping to mitochondrial genes ('percent.mt' or 'percent.MT') were visualized. Samples with less than 5 cells and/or less than 200 detected genes per cell were excluded from further analysis. Quality control based on these metrics included the exclusion of outliers, low quality cells with low gene and/or transcripts detected per cell and/or cells with a high number of transcripts mapping to mitochondrial genes. Doublets were identified as outliers with a dramatically higher number of detected genes per cell than the median \pm interquartile range genes detected per cell. After QC unique molecular identifiers (UMIs, 10x) were then normalized across cells, scaled per 10^4 and converted to log scale using the 'NormalizeData' function. These data were converted to z-scores using the 'ScaleData' command and highly variable genes were selected with the 'FindVariableGenes' function. To integrate multiple samples integration anchors were identified in the list of samples (individual Seurat Objects) with the 'FindIntegrationAnchors' command and the list of samples integrated with the 'IntegrateData' function. Principal components were calculated for these selected genes with the 'RunPCA'. The optimum dimensionality of the dataset for downstream clustering was determined using both the JackStraw and Elbow plot methods. Clusters of similar cells were detected using the Louvain method for community detection including only biologically meaningful principle components to construct the shared nearest neighbor map and an empirically set resolution, as implemented in the 'FindClusters' and 'FindNeighbors' functions.

Clusters were assigned an identity to a given cluster based on expression of tissue compartment markers. The mesenchymal fractions were identified by expression of known marker genes commonly reported in the literature and separated for further analysis using the 'SubsetData' command. RNA markers for each cluster were identified using the 'FindAllMarkers' command in Seurat and examining the top differentially expressed genes in each cluster for homology with the known marker genes. Where a cluster could not be identified using known marker genes, they were identified by a highly discriminative gene that was among the most differentially expressed in that cluster. Differentially expressed genes for each mesenchymal subpopulation relative to all other mesenchymal cells were identified using the 'MAST' statistical framework implemented in the 'FindMarkers' command. To obtain the most sensitive and specific differentially expressed genes for each subpopulation we identified genes with a p-value less than 10^{-5} and an average log fold-change greater than 1.

Comparative analysis of changes in gene expression between mesenchymal cells of the same subpopulation from healthy and fibrotic lungs was performed by calculating the \log_2 (average expression) values for each gene in Seurat and visualizing these on a scatterplot. The genes that changed most significantly between conditions were identified and annotated using the 'MAST' statistical framework implemented in the 'FindMarkers' command and the top genes annotated on the relevant figure.

Enriched genes were annotated as transcription factors from the differentially expressed genes of each cluster by imputing the top differentially expressed genes into the NCBI, EMBL-EBI, UniprotKB Gene Ontology database. Genes were identified as transcription factors if included under the "DNA binding, transcription factor activity" categorisation in the returned results.



QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical difference between groups in the bioinformatics analysis was calculated using the Wilcoxon Signed-rank test. For the scRNA-seq data the lowest p-value calculated in Seurat was $p < 2.2e-10^{-16}$. For all other data the statistical difference between groups was calculated using GraphPad and the exact value was shown.