Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis

Paul A. Reyfman, James M. Walter, Nikita Joshi, Kishore R. Anekalla, Alexandra C. McQuattie-Pimentel, Stephen Chiu, Ramiro Fernandez, Mahzad Akbarpour, Ching-I Chen, Ziyou Ren, Rohan Verma, Hiam Abdala-Valencia, Kiwon Nam, Monica Chi, SeungHye Han, Francisco J. Gonzalez-Gonzalez, Saul Soberanes, Satoshi Watanabe, Kinola J.N. Williams, Annette S Flozak, Trevor T. Nicholson, Vince K. Morgan, Deborah R. Winter, Monique Hinchcliff, Cara L. Hrusch, Robert D. Guzy, Catherine A. Bonham, Anne I. Sperling, Remzi Bag, Robert B. Hamanaka, Gökhan M. Mutlu, Anjana V. Yeldandi, Stacy A. Marshall, Ali Shilatifard, Luis A.N. Amaral, Harris Perlman, Jacob I. Sznajder, A. Christine Argento, Colin T. Gillespie, Jane Dematte, Manu Jain, Benjamin D. Singer, Karen M. Ridge, Anna P. Lam, Ankit Bharat, Sangeeta M. Bhorade, Cara J. Gottardi, G.R. Scott Budinger, and Alexander V. Misharin

ONLINE DATA SUPPLEMENT

Supplementary Methods

Patients: All procedures using human tissues were approved by the Northwestern University Institutional Review Board (STU00056197, STU00201137, and STU00202458) and by the University of Chicago Institutional Review Board (NCT00515567 and NCT00470327) and received further approval from The United States Army Medical Research and Materiel Command`s Office of Research Protections, Human Research Protection Office. All patients provided written informed consent.

Mice: All procedures using mice were approved by the Northwestern Institutional Animal Care and Use Committee. Twelve week old male C57Bl/6 mice obtained from Jackson Laboratories were used for these experiments.

Tissue processing and cell sorting: All samples were processed within 5 hours after procurement, except for the sample IPF 2, which was processed within 36 hours. Upon arrival to the laboratory, a representative piece of whole lung tissue was collected for histology followed by fixation in 4% PFA for 24 hours. A small piece (5x5x5 mm) from an adjacent region was collected and immediately frozen in liquid nitrogen for future processing. The remainder of the tissue (up to 10x10x10 mm) was processed as described previously to prepare a single-cell suspension for flow cytometry and cell sorting (61). Briefly, lung tissue was infiltrated with 2 mg/ml collagenase D (Roche) and 0.2 mg/ml DNase I (Roche) in HBSS with Ca^{2+} and Mg^{2+} using syringe with 30G needle followed by incubation for 30 minutes at 37°C with mild agitation, followed by chopping with scissors and subsequent incubation for 15 minutes at 37°C

in the same digestion buffer. The resulting single-cell suspension was filtered through a 40 μm nylon cell strainer and erythrocytes were lysed using lysing buffer (Pharm Lyse, BD Biosciences). Automated cell counting was performed (Nexcelom K2 Cellometer with AO/PI reagent). Following live/dead staining with viability dye (Fixable Viability Dye eFluor 506, eBioscience), cells were incubated with a human Fc-blocking reagent (Miltenyi) for 5 minutes at 4°C and a mixture of fluorochrome-labeled antibodies for 30 minutes at 4°C. Cells were sorted at the Northwestern University Robert H. Lurie Comprehensive Cancer Center Flow Cytometry Core facility on FACSAria III instrument (BD Biosciences) using a 100 μm nozzle and 40 psi pressure. Alveolar macrophages were identified using sequential gating as singlets/CD45^{+}/live/CD15^{-}/CD206^{++}/CD169^{+}/HLA-DR^{+}/high autofluorescence; alveolar epithelial type II cells were identified as singlets/CD45^{-}/CD31^{-}/EpCam^{+}/HLA-DR^{+}. Between 50,000–300,000 cells were sorted into MACS buffer (Miltenyi), immediately pelleted, and lysed in 350 μl of lysis buffer (RLT Plus, Qiagen) supplemented with beta-mercaptoethanol. Lysates were stored at -80°C until RNA extraction.

RNA isolation: Whole lung tissue was homogenized prior to RNA extraction (BeadBlaster 24, Benchmark Scientific) in the presence of lysis buffer (RLT Plus, Qiagen). Total RNA from whole lung tissue and sorted alveolar macrophages and alveolar type II cells was extracted using AllPrep DNA/RNA Mini Kit (Qiagen) and quality and quantity were assessed using TapeStation 4200 (Agilent).

Single-cell RNA-Seq: A single-cell suspension was prepared as described above. Viability exceeded 80% as determined by AO/PI reagent staining. Single-cell 3’ RNA-Seq libraries were

prepared using Chromium Single Cell V2 Reagent Kit and Controller (10x Genomics). Libraries were assessed for quality (TapeStation 4200, Agilent) and then sequenced on NextSeq 500 or HiSeq 4000 instruments (Illumina), for mouse and human single- cell RNA-Seq libraries, correspondingly.

Initial data processing was performed using the Cell Ranger version 2.0 pipeline (10x Genomics). Post-processing, including filtering by number of genes expressed per cell, was performed using the Seurat package V2.2 and R 3.4 (62, 63). Briefly, gene counts from donor cells and counts from cells of patients with pulmonary fibrosis were aggregated into two separate Seurat objects, which were then aligned to each other using the canonical correlation analysis procedure (64). Following clustering and visualization with t-Distributed Stochastic Neighbor Embedding (tSNE), initial clusters were subjected to inspection and merging based on similarity of marker genes and a function for measuring phylogenetic identity using BuildClusterTree in Seurat. Identification of cell clusters was performed on the final aligned object guided by marker genes. Differential gene expression analysis was performed for each cluster between cells from donors and from patients with pulmonary fibrosis. Heatmaps, tSNE plots, violin plots, and dot plots were generated using Seurat. Functional enrichment with Gene Ontology (GO) Biological Processes was performed using GOrilla (database updated from Gene Ontology Consortium on December 12, 2017) with the top 1,000 upregulated genes with pulmonary fibrosis ordered by FDR q-value, unless there were fewer than 1,000 significantly upregulated genes (q-value < 0.05), in which case all significant genes were used, compared against all remaining detected genes as a background (22, 23). Gene Set Enrichment Analysis (GSEA) was performed using the GSEA 3.0 GSEA Preranked tool for enrichment of the Pulmonary Fibrosis gene set (obtained from Comparative Toxicogenomics Database using MESH term D011658 in November, 2017

via Harmonizome portal), with genes ordered by log difference in average expression (24-26, 65). Counts tables for human single-cell RNA-seq data are available at GEO (GSE122960) and raw data have been deposited to dbGaP/SRA under accession number phs001750.v1.p1. Mouse single cell RNA-seq data is available at GEO: GSE121611. The processed data for figures 1, 5, 6, 10 and supplementary figure S2 is available for exploration via web-tool at nupulmonary.org/resources/. R code used for data analysis is available at https://github.com/NUPulmonary/Reyfman2018.

Bulk RNA-Seq: Libraries for RNA-Seq from whole lung tissue were prepared from 100 ng of total RNA using a 3' mRNA-Sequencing library prep kit (QuantSeq FWD, Lexogen) on an automated liquid handling platform (Bravo, Agilent). Libraries from flow cytometry sorted alveolar macrophages and alveolar type II cells were prepared from 50–100 ng of total RNA using a poly(A) enrichment module (NEBNext Ultra RNA, New England BioLabs). The resulting libraries were assessed (TapeStation 4200, Agilent), multiplexed, and sequenced as single-end 75 base pair reads (NextSeq 500, Illumina). For whole lung tissue RNA-Seq analysis, demultiplexing was performed using bcl2fastq (version 2.17.1.14), trimming was performed using BBDuk, alignment was performed using STAR (version 2.52 with the hg38 assembly) (66), and raw read counts were generated using HTSeq (67). For alveolar macrophage and alveolar type II cell analysis, demultiplexing was performed using bcl2fastq (version 2.17.1.14), trimming was performed using trimmomatic (version 0.33) (68), alignment was performed using TopHat2 (version 2.1.0 with the hg38 assembly) (69), and raw read counts were generated using HTSeq. For all analyses, normalized tables of counts were generated from raw counts using the DESeq2 R package and the rlog transformation function (70). Filtering for low counts was

performed by excluding all genes from which the minimum level of expression across all samples was lower than the average of the third quartile level of expression for all genes. Functional enrichment with GO Biological Processes was performed using GOrilla with the top 500 upregulated genes and top 500 downregulated genes with fibrosis ordered by FDR q-value compared against all remaining detected genes as a background (22, 23). Gene Set Enrichment Analysis (GSEA) was performed using the GSEA 3.0 GSEAPreranked tool testing for enrichment of the Pulmonary Fibrosis gene set (obtained from Comparative Toxicogenomics Database using MESH term D011658 in November, 2017 via Harmonizome portal) with genes ordered by negative log FDR q-value (24, 25, 65).

Fluorescence *in situ* hybridization: Multiplex fluorescent *in situ* hybridization was performed using RNAscope (Advanced Cell Diagnostics (ACD), Newark, CA). Mouse lungs were inflated to 15 cm of H_2O and fixed with 4% paraformaldehyde (EMS) for 24 h. Lungs were paraffin embedded, and 5 μm tissue sections were mounted on Superfrost Plus slides (Fisher). Slides were baked for 1 h at 60°C, deparaffinized in xylene and dehydrated in 100% ethanol. Sections were treated with hydrogen peroxide (ACD 322381) for 10 min at room temperature, then heated to mild boil (98-102˚C) in 1x target retrieval reagent buffer (ACD 322001) for 15 min. Protease Plus (ACD 322381) was applied to sections for 30 min at 40˚C in HybEZ Oven (ACD). Hybridization with target probes, preamplifier, amplifier, fluorescent labels and wash buffer (ACD 320058) was done according to ACD instructions for Multiplex Fluorescent Reagent Kit v2 (ACD 323100) and 4-Plex Ancillary Kit v2 (ACD 323120) when needed. Parallel sections were incubated with ACD positive (ACD 321811) and negative (ACD 321831) control probes. Sections were covered with ProLong Gold Antifade mountant (Thermo P36930). Probes used

were: human *WNT7B* (ACD 421561, NM_058238.2) and human *AXIN2* (ACD 400241-C2, NM_004655.3); human *SFTPC* (ACD 452561, NM_003018.3), human *CD68* (560591, NM_001040059.1), human *CHI3L1* (408121, NM_001276.2), human *SPP1* (420101, NM_001251829.1), mouse *Wnt7a* (ACD 401121, NM_009527.3), mouse *Axin2* (ACD 400331-C3, NM_015732.4) and mouse *Wnt7b* (ACD 401131-C4, NM_009528.3). Opal fluorophores (Opal 520 (FP1487001KT), Opal 620 (FP1495001KT), Opal 690 (FP1497001KT), Perkin Elmer, Shelton, CT) were used at 1:1500 dilution in Multiplex TSA buffer (ACD 322809). Images were captured on Nikon A1R confocal microscope with 40x objective (NU-Nikon Cell Imaging Facility). Images were uniformly processed and pseudocolored in Fiji.

Histopathology and immunohistochemistry: Lung tissue from the regions adjacent to the regions used for single-cell RNA-Seq was fixed in 4% paraformaldehyde for 48 hours, dehydrated and embedded in paraffin. Five micrometer thick sections were prepared, rehydrated and stained with primary and secondary antibodies according to protocols from Human Protein Atlas. The following primary antibodies were used: anti-SPP1 antibody produced in rabbit (Sigma-Aldrich HPA027541, polyclonal), anti-IL1RN antibody produced in rabbit (Sigma-Aldrich HPA001482, polyclonal), anti-MMP9 antibody produced in rabbit (Sigma-Aldrich HPA001238, polyclonal), anti-MARCKS antibody produced in rabbit (Sigma-Aldrich HPA054820, polyclonal), anti-CHI3L1 antibody (Abcam ab77528, polyclonal).

Statistics: Statistical tools for each analysis are explicitly described with the results or detailed in figure legends.

Supplementary Figures

Supplementary Figure S1 (Refers to Figure 1). Histopathology of the lungs used for single-cell RNA-Seq analysis. (**A**) Fibrotic lungs. Review was performed by an independent pathologist. (**B**) Donor lungs. All images are 10x magnification. Note that histopathology for Donor 3 was lost during processing.

Supplementary Figure S2 (Refers to Figure 1). Single-cell RNA-Seq identified diverse populations in human and murine lung. (**A**) Heatmap showing the top five positive marker genes within each cluster from the combined analysis of eight normal and eight fibrotic human lungs in Figure 1A. (**B**) Single-cell RNA-Seq identified 24 cell populations in the normal mouse lung, including several rare populations. tSNE plot shows clustering of 13,888 cells, data from 2 mice. (**C**) Heatmap shows average expression per cell type of the top 3 marker genes for each cell type. (**D**) Bar plot demonstrates that all identified populations, including rare populations were well represented by cells from both mice.

Supplementary Figure S3 (refers to Figure 1). Cell cycle phase does not affect clustering by cell type in the combined single-cell RNA-Seq analysis. Cell cycle score was calculated using CellCycleScoring function in Seurat package. Cells from donor and fibrosis demonstrate similar G2M phase (**A**) and S phase (**B**) score.

Supplementary Figure S4 (refers to Figures 3 and Figure 4). Comparison between bulk and single-cell RNA-Seq differential expression analyses and selected expression of Wnt-signalling genes in the murine lung. (A, B) Correlation between fibrotic versus normal log fold change (logFC) between average single-cell RNA-Seq expression and bulk RNA-Seq expression by gene for alveolar type II cells (**A**) and alveolar macrophages (**B**). Pearson's product-moment correlation $r=0.43$, $P<2.2$ E-16 and $r=0.59$, $P<2.2$ E-16, correspondingly. Genes with absolute(average logFC) <0.05 in the single-cell RNA-Seq dataset or absolute(logFC) <0.05 in the bulk RNA-Seq dataset were excluded (grey). (**C**) Violin plots demonstrating expression of the genes for Wnt ligands and related proteins reported by different groups to form niches responsible for maintenance of progenitor cell populations in mice in alveolar type I (AT1),

alveolar type II (AT2), club cells (Club), ciliated cells (Cil), fibroblasts (Fib) and smooth muscle cells (SM) (30, 31, 49). Expression of *Wnt1, Wnt3, Wnt8a, Wnt8b, Wnt9b, Wnt10a* and *Wnt16* was not detected and not shown.

Supplementary Figure S5 (refers to Figures 1, 5 and 6). Individual tSNE clustering of cells from eight donor lungs and eight subjects with pulmonary fibrosis. Individual single-cell RNA-Seq experiments were processed using Seurat R toolkit (see code in Supplementary methods), cell identities were assigned based on expression of canonical markers (Supplementary Table S5).

Supplementary Figure S6 (Refers to Figures 1, 5 and 6). Examples of cell populations identified in a minority of patients. (**A**) In subject IPF 2, *FOXP3* and *CD4* expressing regulatory T cells, and *ITGAE* and *CD8* expressing lung resident T cell subsets could be resolved. (**B**) In the subject with hypersensitivity pneumonitis, in addition to $CD1c^+$ conventional dendritic cells observed in several subjects, two additional distinct small populations of cells expressing CD45 were detected: one population expressed the plasmacytoid dendritic cell markers *CLEC4C*, *TCL1A*, *IRF8*, and *TLR7*, and the second expressed the conventional dendritic cell (DC1) marker *CLEC9A*. (**C**) In four subjects (IPF 2, HP, Donor 6, SSc-ILD 1), classical and non-classical monocytes could be resolved based on expression of *CD14* and *FCGR3A (CD16)*. These cells do not express the macrophage marker *MRC1*.

Supplementary Figure S7 (refers to Figure 5, 6 and 7). Alveolar macrophages or epithelial cells from patients with pulmonary fibrosis do not exhibit clustering according to clinical disease etiology. (A, B) Alveolar macrophages. **(C, D)** Epithelial cells. Clusters on panels B and D correspond to clusters on Figures 5 and 6, correspondingly. (**E**) Heterogeneity of pulmonary macrophages in the donor and fibrotic human lung. Donor 1 (first column) is representative of the homogeneity of alveolar macrophages observed in the majority of the donors. Donor alveolar macrophages express *FABP4* and do not express *CHI3L1* or *SPP1*. In contrast, macrophages from patients with pulmonary fibrosis (second, third, and fourth columns) demonstrate marked heterogeneity of alveolar macrophages and emergence of the clusters of *CHI3L1*- and *SPP1*-positive cells. (**F**) Macrophages in Donor 6 (fifth column) demonstrated heterogeneity. In

addition to the *MRC1*- and *FAPB4*-positive alveolar macrophages observed in the other donors, there was a distinct cluster of macrophages expressing the activation marker *CCL3* (MIP-1) as well as a cluster of *MRC1*-positive *FAPB4*-negative macrophage suggestive of recruited alveolar macrophages. A review of this donor's premortem labs showed respiratory cultures were positive for *Klebsiella pneumoniae*.

Supplementary Figure S8 (refers to Figures 7 and 8). *In situ* RNA hybridization with amplification in the human lung and expression of Wnt genes and *AXIN2* in the human lung. (A-D) Lung sections from Donor 1 (panels A, C), IPF 1 (panel B) and IPF 4 (panel D) were hybridized with indicated target probes. Low magnification (20x) images for differential interference contrast (DIC) phase are shown on the left, followed by single channel images for *SFTPC* (yellow), *CD68* (magenta) and either *CHI3L1* or *SPP1* (cyan). Overlay images with nuclei staining (Hoechst, blue) are shown on the right. (**A, B**) *CHI3L1*-positive and *CHI3L1*-negative alveolar macrophages coexist in the same niche in patients with pulmonary fibrosis. (**C, D**) *SPP1*-positive and *SPP1*-negative alveolar macrophages coexist in the same niche in patients with pulmonary fibrosis. Scale bar 50 μm. (**E**) Violin plots demonstrating that Wnt ligand expression is relatively homogenous in single-cell RNA-Seq data from epithelial cells in normal and fibrotic lungs.

Supplementary Figure S9 (refers to Figure 8). **Mutually exclusive Wnt-expressor and -responder cells in the adult murine lung.** (**A**) Alveolar type II cells were examined for expression of individual Wnt ligands. *Wnt3a* and *Wnt7b* were the most abundantly expressed ligands, but no cells were detected that expressed more than one Wnt ligand. (**B**) Cells expressing Wnt ligands were generally distinct from cells expressing the Wnt target gene *Axin2*. Clusters represent alveolar type II cells from two mice. (**C**) Representative RNAscope in situ hybridization images of adult murine lung sections incubated with the designated target probes, *Wnt7b* (magenta), *Wnt7a* (yellow) and *Axin2* (green). Nuclei are stained with Hoechst (blue) and overlay images are shown on the far right. Differential interference contrast (DIC) phase image shown on the left. As with human lung, *Wnt7b* is abundantly detected in airway epithelial/club cells (dotted line marks basement membrane of these cells). Boxed region corresponds to higher magnification view in Figure 8H. *Wnt7a* and *Axin2* are largely restricted to alveolar regions,

where they mark distinct cells. Scale = 50 μm. Asterisk (*) denotes airway lumen. (**D**) Graph shows quantification of single-positive cells, double-positive cells and rare triple positive cells from 5 fields of view. Only signals directly overlapping with Hoechst-stained nuclei were counted, as cytoplasmic signals were difficult to assign to a particular cell. Double- and triple-positive cells were significantly less abundant than single-positive cells by ANOVA with Bonferroni's multiple comparisons Test ($p < 0.01$).

Supplementary Figure S10 (refers to Figure 9). **Identification of rare cell populations using single-cell RNA-Seq data**. (**A**) Alveolar type II cells from the murine single-cell RNA-Seq data identified on Figure S1B were subjected to additional clustering. (**B**) All identified clusters contained cells from both replicates. (**C**) Violin plots show the top two most differentially expressed genes between clusters of alveolar type II cells (Supplementary Table S9). None of the clusters was enriched for expression of *Axin2* or *Tm4sf1*. (**D**) Violin plots show the enrichment score for the top 500 genes reported to be expressed in alveolar epithelial progenitor cells in clusters of mouse alveolar type II cells. (**E**) Non-overlapping expression of *Axin2* and *Tm4sf1* in murine alveolar type II cells. (**F**) Clustering of alveolar type II cells from the eight normal lungs does not reveal clusters enriched for *AXIN2* and *TM4SF1*. (**G**) Expression of canonical genes associated with replicative senescence (*CDKN2A, CDKN1A, GLB1, SERPINE1, IL6*) in the epithelial cells from normal and fibrotic lungs.

Supplementary Figure S11 (refers to Figure 10). **Single-cell analysis of cryobiopsy specimens.** (**A**) Gating strategy for flow cytometry performed on an ex vivo cryobiopsy. (**B-G)** Clustering of the 192 alveolar macrophages and 364 epithelial cells identified in the cryobiopsy specimen together with those from the eight normal and eight fibrotic lungs. **(B, E)** With the addition of cryobiopsy cells, clustering of alveolar macrophages and epithelial cells appeared similar to results shown in Figures 5 and 6. (**C, D**) The majority of the cryobiopsy alveolar macrophages were in a cluster containing cells from both normal and fibrotic lungs, while only a minority were in a cluster comprised almost exclusively of cells from fibrotic lungs. (**F, G**) The cryobiopsy epithelial cells appeared to be distributed more among clusters primarily containing cells from fibrotic lungs than clusters primarily containing cells from donor lungs.

Supplementary Tables

Supplementary Table 1. Full list of identified marker genes in the human canonical correlation analysis. Refers to Figure 1.

Supplementary Table 2. Marker genes used to assign clusters from murine lung single-cell RNA-Seq. Refers to Supplementary Figure S2.

Supplementary Table 3. Differentially expressed genes between donor and fibrosis lung samples in macrophages, alveolar type II cells, and fibroblasts.
Refers to Figure 2.

Supplementary Table 4. Differentially expressed genes between donor and fibrosis lung samples in flow sorted alveolar macrophages, alveolar type II cells, and whole lung.
Refers to Figure 3.

Supplementary Table 5. Marker genes used to identify cell populations from individual tSNE plots for each patient.
Refers to Supplementary Figure S5.

Supplementary Table 6. Marker genes identified in human macrophage subclustering analysis.
Refers to Figure 5.

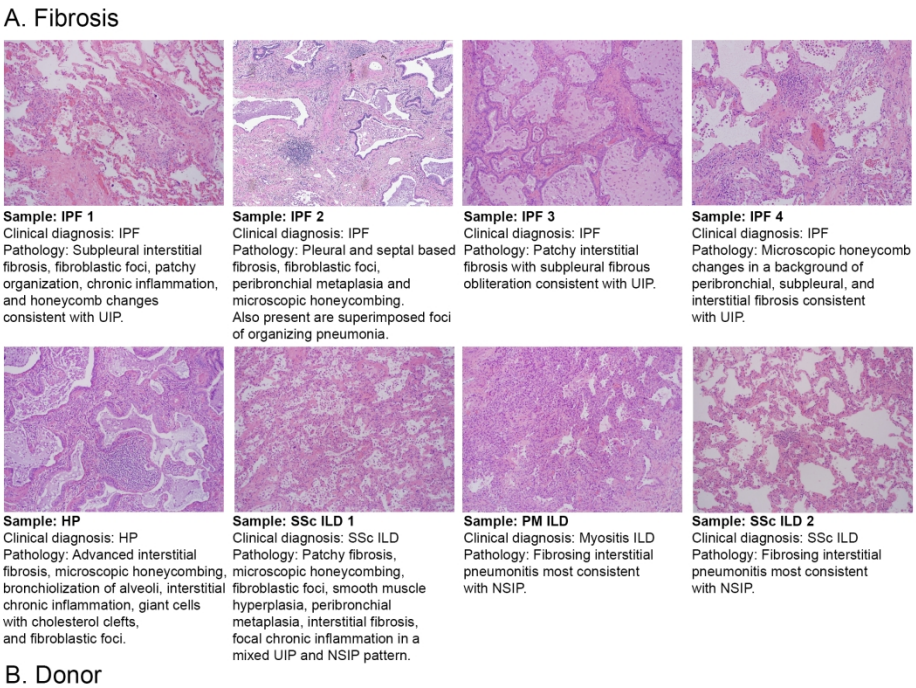
Supplementary Table 7. Full list of GO processes comparing gene expression in each macrophage cluster with average gene expression in alveolar macrophages.
Refers to Figure 5.

Supplementary Table 8. Marker genes identified in human epithelial cell subclustering analysis.

Refers to Figure 6.

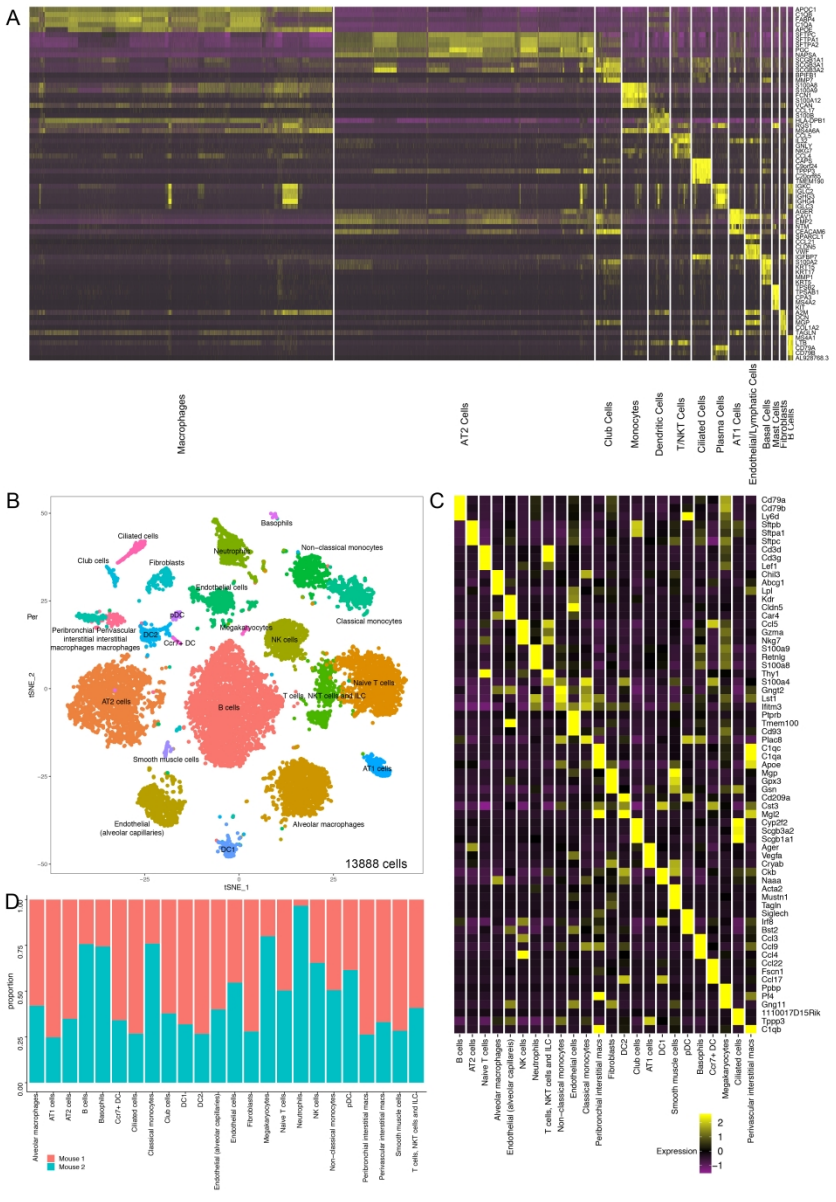
Supplementary Table 9. Marker genes identified in murine epithelial cell subclustering analysis.

Refers to Supplementary Figure S10.

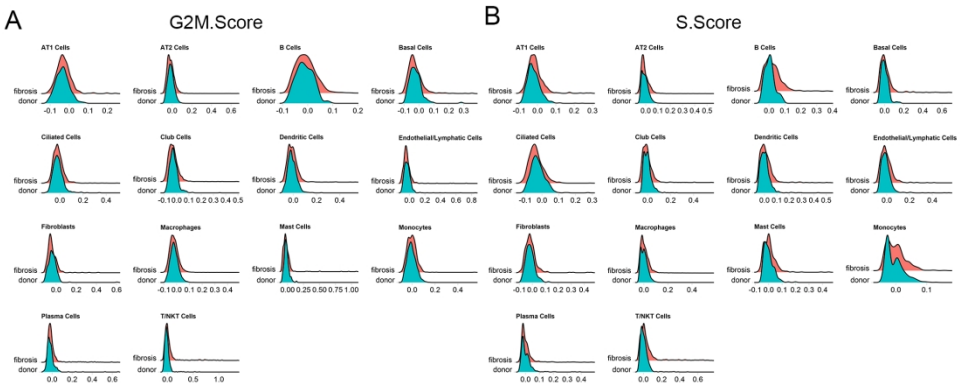


Supplementary Figure S1 (Refers to Figure 1). Histopathology of the lungs used for single-cell RNA-Seq analysis. (A) Fibrotic lungs. Review was performed by an independent pathologist. (B) Donor lungs. All images are 10x magnification. Note that histopathology for Donor 3 was lost during processing.

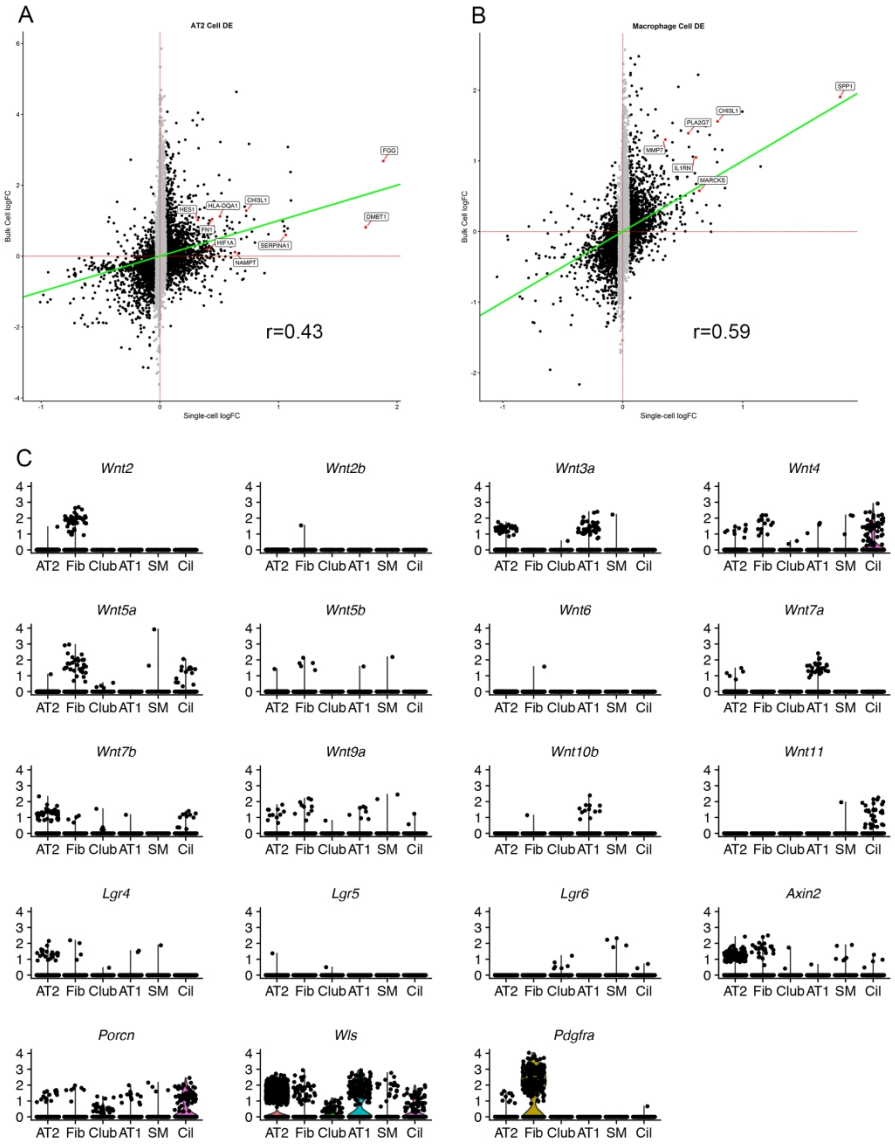
190x242mm (300 x 300 DPI)



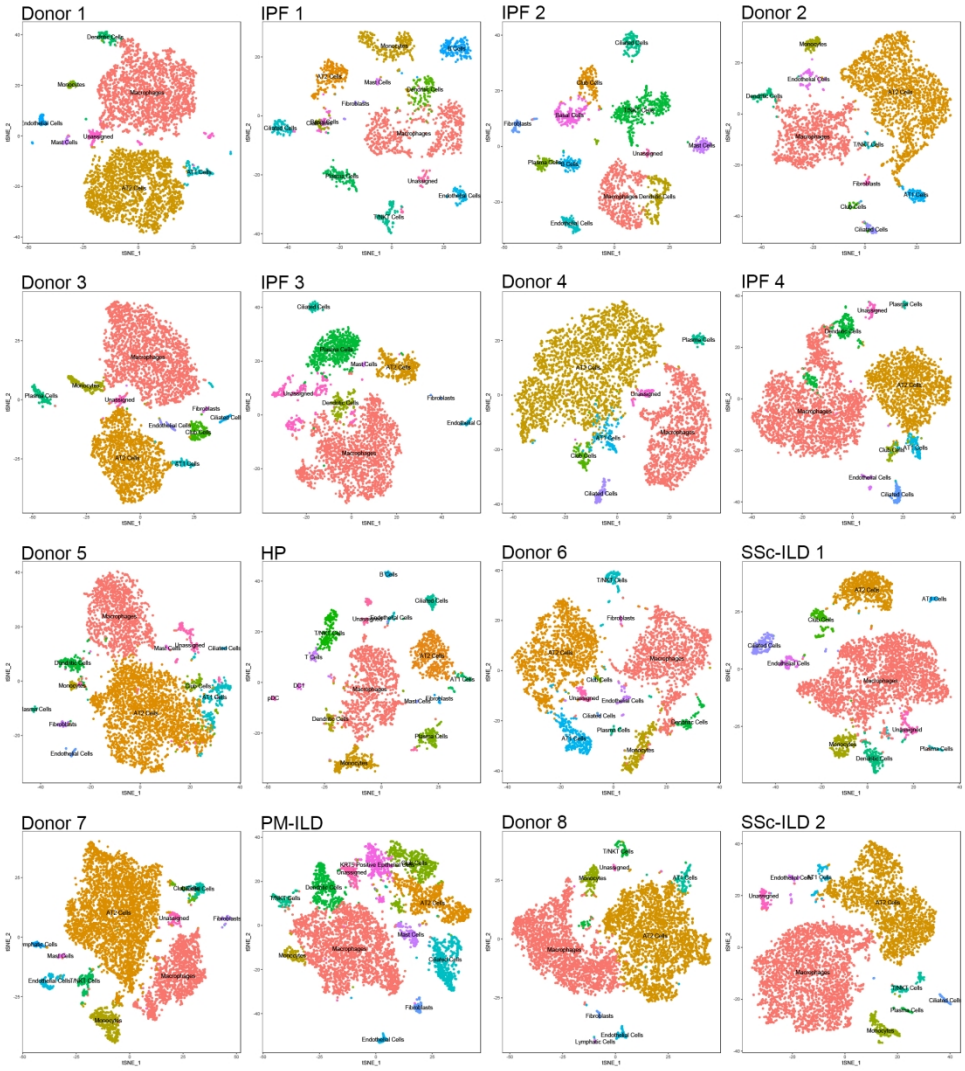
Supplementary Figure S2 (Refers to Figure 1). Single-cell RNA-Seq identified diverse populations in human and murine lung. (A) Heatmap showing the top five positive marker genes within each cluster from the combined analysis of eight normal and eight fibrotic human lungs in Figure 1A. (B) Single-cell RNA-Seq identified 23 cell populations in the normal mouse lung, including several rare populations. tSNE plot shows clustering of 13,888 cells, data from 2 mice. (C) Heatmap shows average expression per cell type of the top 3 marker genes for each cell type. (D) Bar plot demonstrates that all identified populations, including rare populations were well represented by cells from both mice.



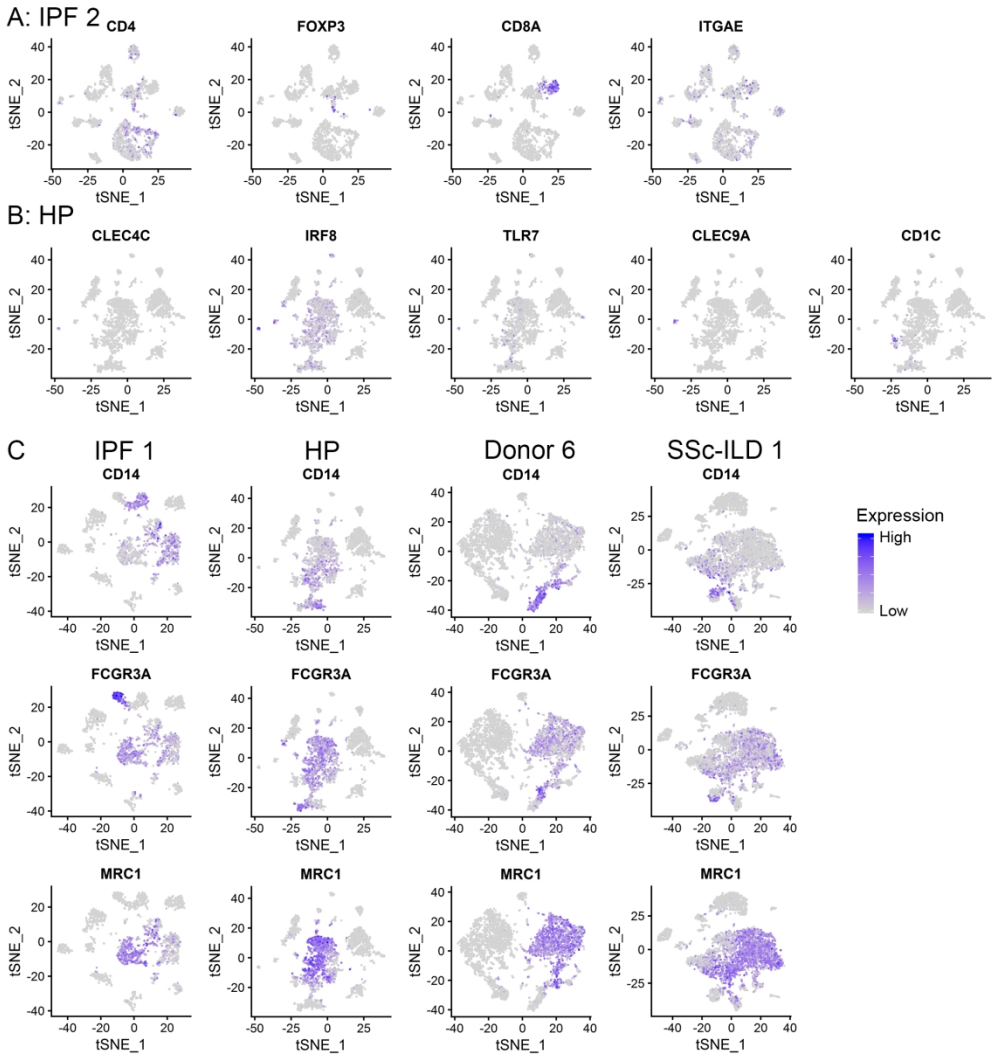
Supplementary Figure S3 (refers to Figure 1). Cell cycle phase does not affect clustering by cell type in the combined single-cell RNA-Seq analysis. Cell cycle score was calculated using CellCycleScoring function in Seurat package. Cells from donor and fibrosis demonstrate similar G2M phase (A) and S phase (B) score.



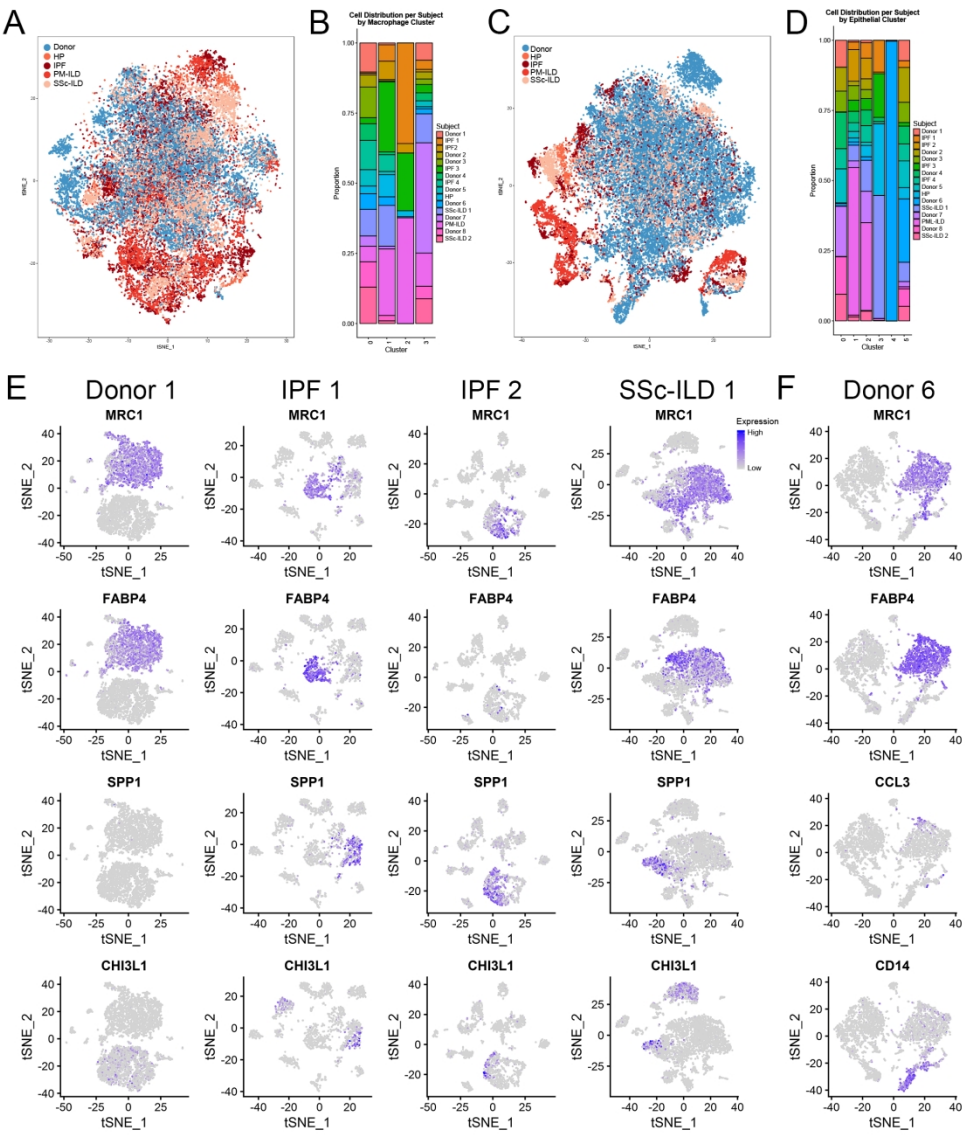
Supplementary Figure S4 (refers to Figures 3 and Figure 4). Comparison between bulk and single-cell RNA-Seq differential expression analyses and selected expression of Wnt-signalling genes in the murine lung. (A, B) Correlation between fibrotic versus normal log fold change (logFC) between average single-cell RNA-Seq expression and bulk RNA-Seq expression by gene for alveolar type II cells (A) and alveolar macrophages (B). Pearson's product-moment correlation r=0.43, P<2.2 E-16 and r=0.59, P<2.2 E-16, correspondingly. Genes with absolute(average logFC) <0.05 in the single-cell RNA-Seq dataset or absolute(logFC) <0.05 in the bulk RNA-Seq dataset were excluded (grey). (C) Violin plots demonstrating expression of the genes for Wnt ligands and related proteins reported by different groups to form niches responsible for maintenance of progenitor cell populations in mice in alveolar type I (AT1), alveolar type II (AT2), club cells (Club), ciliated cells (Cil), fibroblasts (Fib) and smooth muscle cells (SM) (31, 32, 50). Expression of Wnt1, Wnt3, Wnt8a, Wnt8b, Wnt9b, Wnt10a and Wnt16 was not detected and not shown.



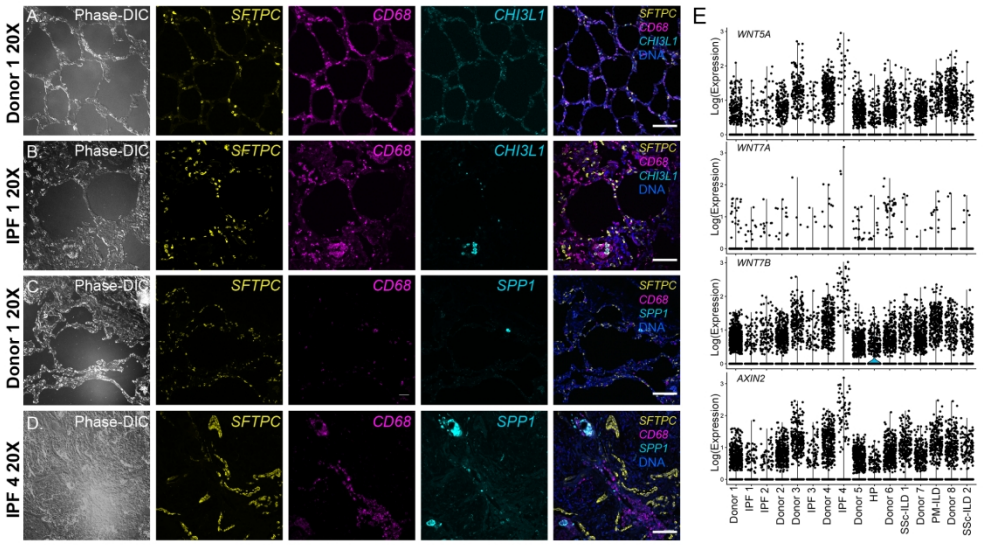
Supplementary Figure S5 (refers to Figures 1, 5 and 6). Individual tSNE clustering of cells from eight donor lungs and eight subjects with pulmonary fibrosis. Individual single-cell RNA-Seq experiments were processed using Seurat R toolkit (see code in Supplementary methods), cell identities were assigned based on expression of canonical markers (Supplementary Table S5).



Supplementary Figure S6 (Refers to Figures 1, 5 and 6). Examples of cell populations identified in a minority of patients. (A) In subject IPF 2, FOXP3 and CD4 expressing regulatory T cells, and ITGAE and CD8 expressing lung resident T cell subsets could be resolved. (B) In the subject with hypersensitivity pneumonitis, in addition to CD1c+ conventional dendritic cells observed in several subjects, two additional distinct small populations of cells expressing CD45 were detected: one population expressed the plasmacytoid dendritic cell markers CLEC4C, TCL1A, IRF8, and TLR7, and the second expressed the conventional dendritic cell (DC1) marker CLEC9A. (C) In four subjects (IPF 2, HP, Donor 6, SSc-ILD 1), classical and non-classical monocytes could be resolved based on expression of CD14 and FCGR3A (CD16). These cells do not express the macrophage marker MRC1.

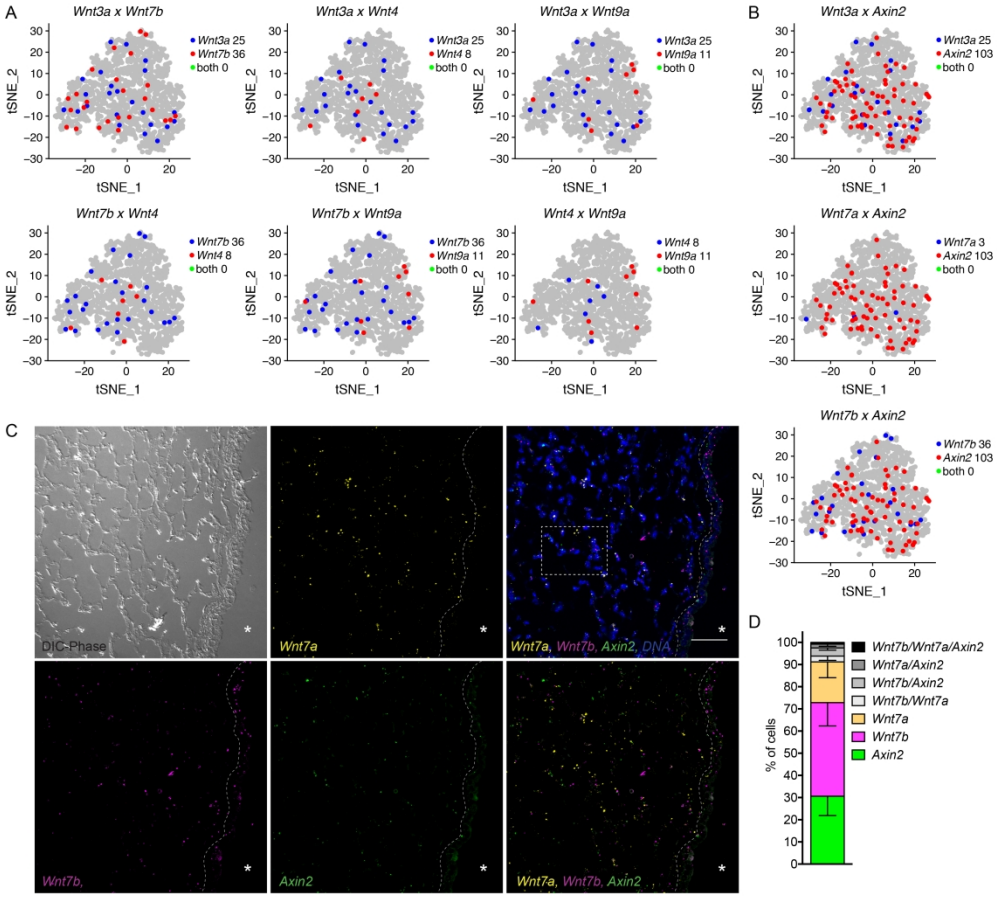


Supplementary Figure S7 (refers to Figure 5, 6 and 7). Alveolar macrophages or epithelial cells from patients with pulmonary fibrosis do not exhibit clustering according to clinical disease etiology. (A, B) Alveolar macrophages. (C, D) Epithelial cells. Clusters on panels B and D correspond to clusters on Figures 5 and 6, correspondingly. (E) Heterogeneity of pulmonary macrophages in the donor and fibrotic human lung. Donor 1 (first column) is representative of the homogeneity of alveolar macrophages observed in the majority of the donors. Donor alveolar macrophages express FABP4 and do not express CHI3L1 or SPP1. In contrast, macrophages from patients with pulmonary fibrosis (second, third, and fourth columns) demonstrate marked heterogeneity of alveolar macrophages and emergence of the clusters of CHI3L1- and SPP1-positive cells. (F) Macrophages in Donor 6 (fifth column) demonstrated heterogeneity. In addition to the MRC1- and FAPB4-positive alveolar macrophages observed in the other donors, there was a distinct cluster of macrophages expressing the activation marker CCL3 (MIP-1) as well as a cluster of MRC1-positive FAPB4-negative macrophage suggestive of recruited alveolar macrophages. A review of this donor's premortem labs showed respiratory cultures were positive for Klebsiella pneumoniae.

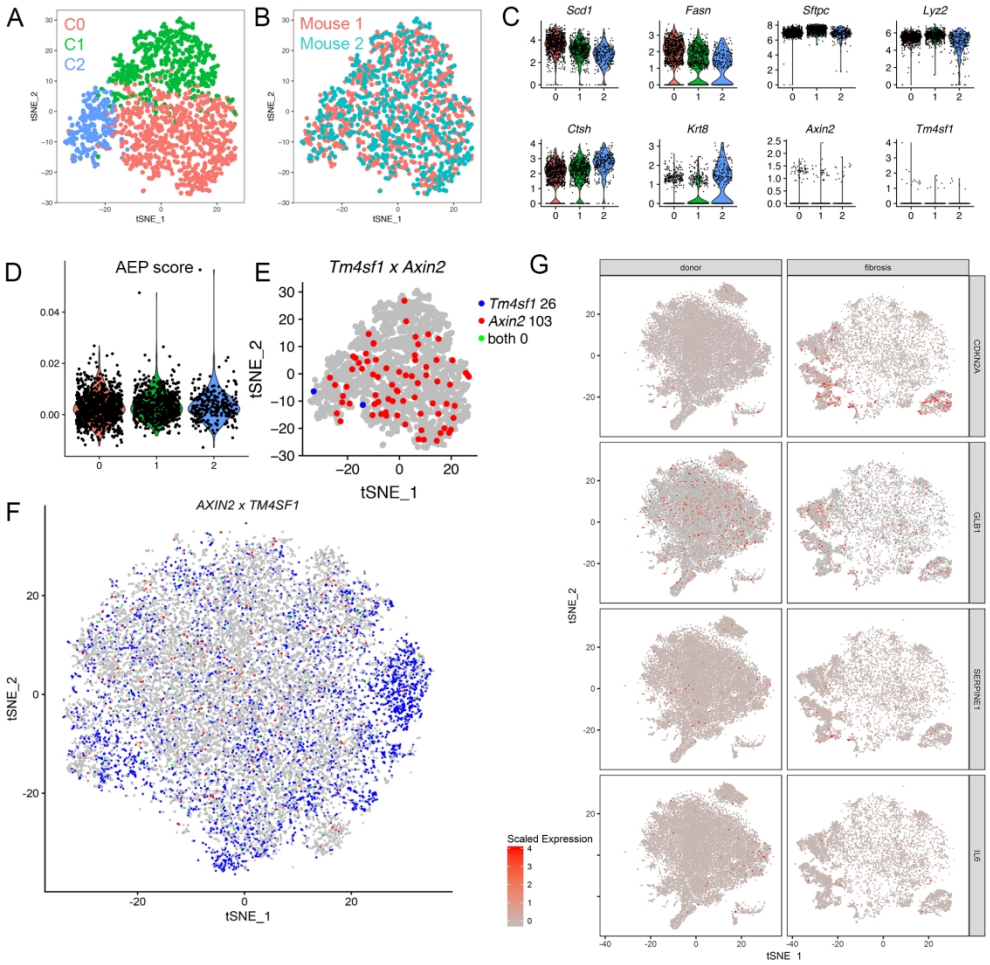


Supplementary Figure S8 (refers to Figures 7 and 8). In situ RNA hybridization with amplification in the human lung and expression of Wnt genes and AXIN2 in the human lung. (A-D) Lung sections from Donor 1 (panels A, C), IPF 1 (panel B) and IPF 4 (panel D) were hybridized with indicated target probes. Low magnification (20x) images for differential interference contrast (DIC) phase are shown on the left, followed by single channel images for SFTPC (yellow), CD68 (magenta) and either CHI3L1 or SPP1 (cyan). Overlay images with nuclei staining (Hoechst, blue) are shown on the right. (A, B) CHI3L1-positive and CHI3L1-negative alveolar macrophages coexist in the same niche in patients with pulmonary fibrosis. (C, D) SPP1-positive and SPP1-negative alveolar macrophages coexist in the same niche in patients with pulmonary fibrosis. Scale bar 50 μm. (E) Violin plots demonstrating that Wnt ligand expression is relatively homogenous in single-cell RNA-Seq data from epithelial cells in normal and fibrotic lungs.

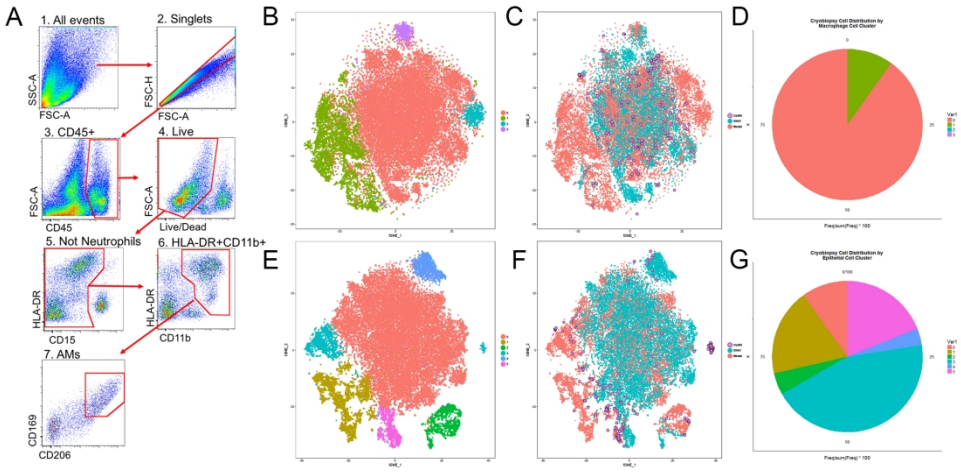
190x108mm (300 x 300 DPI)



Supplementary Figure S9 (refers to Figure 8). Mutually exclusive Wnt-expressor and -responder cells in the adult murine lung. (A) Alveolar type II cells were examined for expression of individual Wnt ligands. Wnt3a and Wnt7b were the most abundantly expressed ligands, but no cells were detected that expressed more than one Wnt ligand. (B) Cells expressing Wnt ligands were generally distinct from cells expressing the Wnt target gene Axin2. Clusters represent alveolar type II cells from two mice. (C) Representative RNAscope in situ hybridization images of adult murine lung sections incubated with the designated target probes, Wnt7b (magenta), Wnt7a (yellow) and Axin2 (green). Nuclei are stained with Hoechst (blue) and overlay images are shown on the far right. Differential interference contrast (DIC) phase image shown on the left. As with human lung, Wnt7b is abundantly detected in airway epithelial/club cells (dotted line marks basement membrane of these cells). Boxed region corresponds to higher magnification view in Figure 8H. Wnt7a and Axin2 are largely restricted to alveolar regions, where they mark distinct cells. Scale = 50 □m. Asterisk (*) denotes airway lumen. (D) Graph shows quantification of single-positive cells, double-positive cells and rare triple positive cells from 5 fields of view. Only signals directly overlapping with Hoechst-stained nuclei were counted, as cytoplasmic signals were difficult to assign to a particular cell. Double- and triple-positive cells were significantly less abundant than single-positive cells by ANOVA with Bonferroni's multiple comparisons Test ($p < 0.01$).



Supplementary Figure S10 (refers to Figure 9). Identification of rare cell populations using single-cell RNA-Seq data. (A) Alveolar type II cells from the murine single-cell RNA-Seq data identified on Figure S1B were subjected to additional clustering. (B) All identified clusters contained cells from both replicates. (C) Violin plots show the top two most differentially expressed genes between clusters of alveolar type II cells (Supplementary Table S9). None of the clusters was enriched for expression of Axin2 or Tm4sf1. (D) Violin plots show the enrichment score for the top 500 genes reported to be expressed in alveolar epithelial progenitor cells in clusters of mouse alveolar type II cells. (E) Non-overlapping expression of Axin2 and Tm4sf1 in murine alveolar type II cells. (F) Clustering of alveolar type II cells from the eight normal lungs does not reveal clusters enriched for AXIN2 and TM4SF1. (G) Expression of canonical genes associated with replicative senescence (CDKN2A, CDKN1A, GLB1, SERPINE1, IL6) in the epithelial cells from normal and fibrotic lungs.



Supplementary Figure S11 (refers to Figure 10). Single-cell analysis of cryobiopsy specimens. (A) Gating strategy for flow cytometry performed on an ex vivo cryobiopsy. (B-G) Clustering of the 192 alveolar macrophages and 364 epithelial cells identified in the cryobiopsy specimen together with those from the eight normal and eight fibrotic lungs. (B, E) With the addition of cryobiopsy cells, clustering of alveolar macrophages and epithelial cells appeared similar to results shown in Figures 5 and 6. (C, D) The majority of the cryobiopsy alveolar macrophages were in a cluster containing cells from both normal and fibrotic lungs, while only a minority were in a cluster comprised almost exclusively of cells from fibrotic lungs. (F, G) The cryobiopsy epithelial cells appeared to be distributed more among clusters primarily containing cells from fibrotic lungs than clusters primarily containing cells from donor lungs.