Cell isolation

The human explant tissues were processed as described (1). Firstly, human lung tissue was finely minced, washed in Hank's balanced salt solution (HBSS) at 4°C for 5 minutes with rocking, and then centrifuged for 5 minutes at 500g and 4°C. This process was repeated as necessary to remove as much blood as possible. Subsequently, the minced cleaned tissue was incubated in HBSS containing 1X Liberase (0.25U/ml) and DNaseI (0.5mg/ml) and agitated at 1000 rpm at 37°C for 30 minutes using a Thermomixer (Eppendorf) to dissociate the tissue chunks. The single-cell preparations obtained were then further dissociated by passing them through 16 G syringe needles and agitated gently at 1000 rpm at 37°C on the Thermomixer for another 15 minutes.

For enriching epithelial cells, the dissociated single-cell preparations were stained with antibodies against EPCAM, CD235a, CD45, CD31 (Biolegend) and HTII-280 (Terrace Biotech) and sorted using fluorescence-activated cell sorting (FACS) with a BD Influx cell sorter (Becton Dickinson). The viability of the cells was determined by staining the cell preparations with DAPI (ThermoFisher Scientific) 15 minutes prior to cell sorting.

scRNA-Seq

$CD45^{-}CD31^{-}EPCAM^{+}HTII\text{-}280^{+}$ and $CD45^{-}CD31^{-}EPCAM^{+}HTII\text{-}280^{-}$ lung epithelial cells were FACS sorted from the healthy and IPF explant samples. Totalseq-A human hashing antibodies (Biolegend) were used to label FACS sorted cells. Hashed $CD45^{-}CD31^{-}EPCAM^{+}HTII\text{-}280^{+}$ and $CD45^{-}CD31^{-}EPCAM^{+}HTII\text{-}280^{-}$ lung epithelial cells were mixed 1:1 ratio and subjected to scRNA-Seq using the 10x Genomic Chromium system. Single cells were captured using a 10X Chromium device (10X Genomics), and the libraries were prepared according to the Chromium Next GEM Single Cell 3' v3.1Reagent Kits User Guide (10X Genomics). The barcoded sequencing libraries were quantified using the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA). The sequencing libraries were then sequenced using Novaseq 6000 (Illumina) with 150 bp pair-end to obtain a sequencing depth of ~$5x10^4$ reads per cell.

Data analysis

The CellRanger v6.0.0 software was used with the default settings for demultiplexing, aligning reads with STAR software to a human GRCh38-2020A transcriptome reference genome (provided by 10X Genomics) https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest.The single-cell analysis R package Seurat v5.0was used for further data analysis (2). For HTO tag demultiplex, we use "positive.quantile = 0.99" to cut off singlets. For quality control and filtering out low quality cells, only cells expressing more than 200 genes (defined as genes detected in at least 3 cells) and fewer than 10% mitochondrial genes were selected. To minimize doublet contamination, cells with a high number of detected genes were removed using a fit model generated from the suggested "multiplet rate"/ "number of cells recovered" ratio, as stated in the 10X Genomics user manual on top of the doublets removal at the HTO tag demultiplex step (3).

We employed default normalization and data scaling procedures from the Seurat package, which involve log normalization and linear model for data scaling. For data integration, we utilized the batch correction package Harmony (4). Principal Component Analysis (PCA) was performed using the 4000 most variable genes, and the first 30 independent components were used for downstream unbiased clustering with a resolution of 0.6. The Uniform Manifold Approximation and Projection (UMAP) method was utilized for unsupervised clustering visualization. The cell

cluster identities were determined using known gene markers of major epithelial and immune cell types (5).

To identify differentially expressed genes between different clusters and groups, we utilized the Mode-based Analysis of Single-cell transcriptomics (MAST) method (6).

Immunofluorescence staining, histology staining, imaging and quantification

Immunofluorescence staining, Hematoxylin and Eosin (H&E) staining was performed on fixed lung tissue embedded in paraffin and processed as previously described (1). Immunofluorescence images were taken using a Zeiss LSM780 confocal microscope and Nikon Ti2 microscope. Quantification is done either using QuPath-0.6.0.

Primary antibodies used for immunofluorescence staining include: mouse anti-proSftpc (1:100, Santa Cruz, sc-518029), mouse-HTII-280 (IgM) (1:500, Terrace Biotech, TB 27AHT2-280), mouse anti- ac-tubulin (1:500, Sigma T6793), mouse anti-FOXJ1 (1:300, ThermoFisher Scientific, 14-9965-82), rabbit anti-MUC5B (1:500, Sigma, HPA008246), mouse anti-MUC5B (1:200, Abcam, ab77995), Rat anti-Uteroglobin/SCGB1A1, 1:100, R&D system, MAB4218), Rabbit anti-LAMP3 (1:500, Cell Signaling Technology, 4778S), Rabbit anti-SCGB3A2 (1:100, Abcam, ab181853), Chicken anti-Keratin 5 (1:500, BioLegend, 905904)

Secondary antibodies were: Alexa Fluor 647-conjugated donkey anti-mouse IgG(H+L) (1:1000, Life Technologies, A31571), Alexa Fluor 594-conjugated donkey anti-rabbit IgG(H+L) (1:1000, Life Technologies, A21207), Alexa Fluor 488- conjugated donkey anti-rabbit IgG(H+L) (1:1000, Life Technologies, A21206), Alexa Fluor 488-conjugated donkey anti-mouse IgG(H+L) (1:1000, Life Technologies, A21202), Alexa Fluor 488-conjugated goat anti-chicken IgG(H+L) (1:1000, Life Technologies, A11039) and Alexa Fluor 488-conjugated donkey anti-chicken IgY (IgG) (H+L) (1:1000), FITC Goat anti-mouse IgM (1:1000, ThermoFisher Scientific, 31992), Alexa Fluor 568-conjugated goat anti-mouse IgG2b (1:1000, ThermoFisher Scientific, A-21144), Alexa Fluor 568-conjugated donkey anti-Rat (1:1000, ThermoFisher Scientific, A78946)

References:

1. Yao C, Guan X, Carraro G, Parimon T, Liu X, Huang G, Mulay A, Soukiasian HJ, David G, Weigt SS, Belperio JA, Chen P, Jiang D, Noble PW, Stripp BR. Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2021; 203: 707-717.
2. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell* 2019; 177: 1888-1902 e1821.
3. Carraro G, Mulay A, Yao C, Mizuno T, Konda B, Petrov M, Lafkas D, Arron JR, Hogaboam CM, Chen P, Jiang D, Noble PW, Randell SH, McQualter JL, Stripp BR. Single-Cell Reconstruction of Human Basal Cell Diversity in Normal and Idiopathic Pulmonary Fibrosis Lungs. *American Journal of Respiratory and Critical Care Medicine* 2020; 202: 1540-1550.
4. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019; 16: 1289-1296.

5. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, Peter L, Chung MI, Taylor CJ, Jetter C, Raju L, Roberson J, Ding G, Wood L, Sucre JMS, Richmond BW, Serezani AP, McDonnell WJ, Mallal SB, Bacchetta MJ, Loyd JE, Shaver CM, Ware LB, Bremner R, Walia R, Blackwell TS, Banovich NE, Kropski JA. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020; 6: eaba1972.
6. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS, Gottardo R. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015; 16: 278.