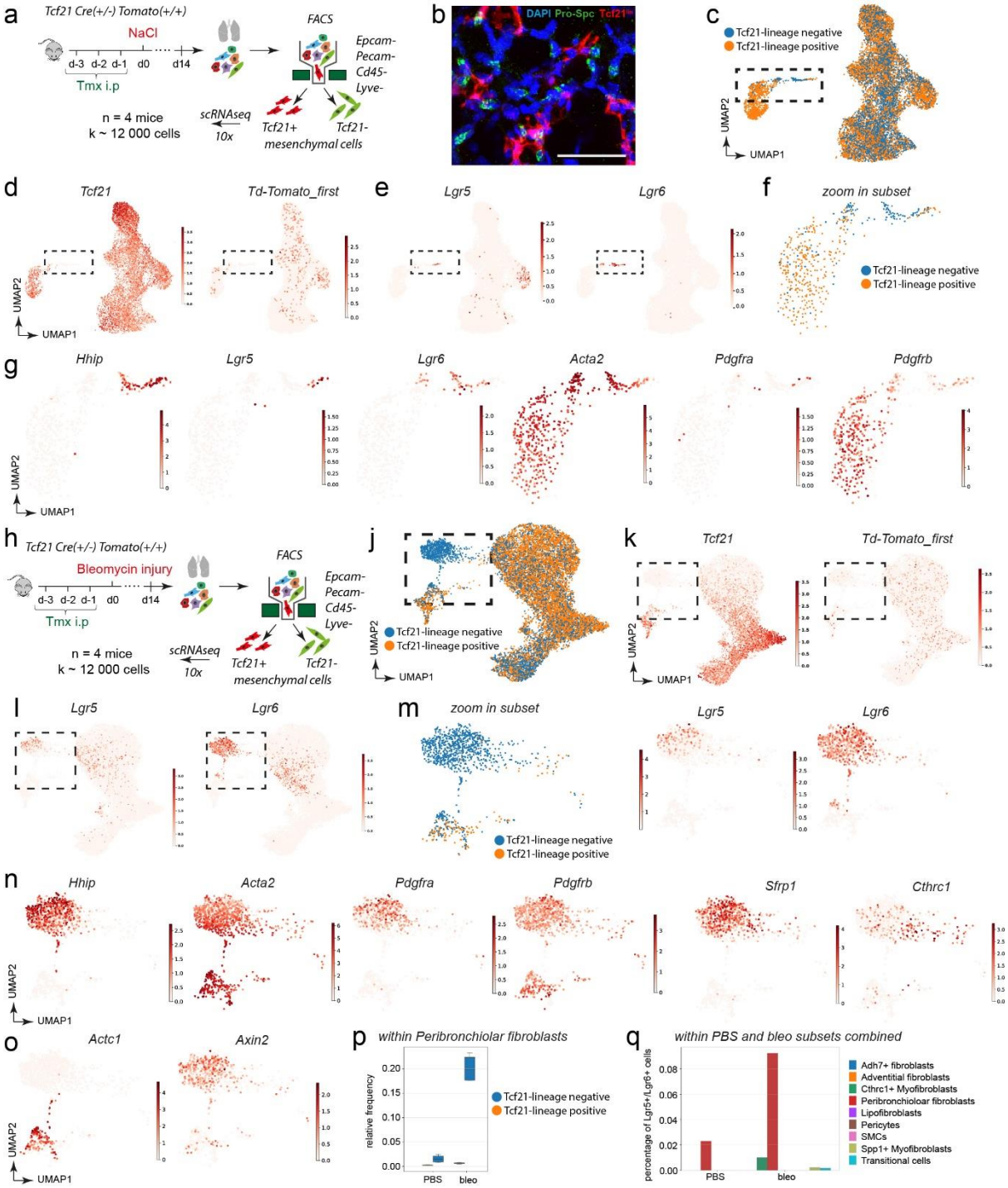


Figure S1. Complexity within Hhip+ peribronchiolar fibroblast population. a,h) Adult Tcf21-Cre(+/-)-Td-Tomato(+/-) were injected with Tamoxifen three days before applying **a)** NaCl or **h)** bleomycin and sacrificed at day 14. Isolated cells from mouse lungs were negative selected (Epcam-/Pecam-/Cd45/Lyve-) for mesenchymal cells using FACS, and Tcf21 lineage positive and negative cells were sorted and subjected for scRNAseq analysis using the 10x Genomics platform. **b)** Immunofluorescent analysis of Pro-SPC (green) shows AT-2 cells and Tcf21+Td-tomato+ (red) positive mesenchymal cells in the lung. **c, j)** UMAP embeddings display cells colored by Tcf21 lineage identity for c) PBS or j) bleomycin derived cells respectively, **d,k)** Tcf21 and Td-tomato gene expression, **e,i)** Lgr5 and Lgr6 expression. Rectangle indicates subset of cells used for the embedding panels in f-g**. f-g)** Zoom into PBS derived subset of Lgr5+/Lgr6+ positive cells show UMAP embeddings displaying Hhip, Lgr5, Lgr6, Acta2, Pdgfra and Pdgfrb gene expression (g)**. m-o)** Zoom into bleomycin day14 derived subset of Lgr5+/Lgr6 positive cells shows UMAP embeddings displaying **m)** Lgr5 and Lgr6, **n)** Hhip, Acta2, Pdgfra, Pdgfrb, Sfrp1, Cthrc1, and **o)** Actc1 and Axin2 gene expression. **p)** Frequency analysis of Tcf21 lineage positive and

negative cells within the population of Peribronchiolar fibroblasts. **q)** Frequency analysis of Lgr5+/Lgr6+ positive cells across all mesenchymal subtypes and between PBS and bleomycin condition.

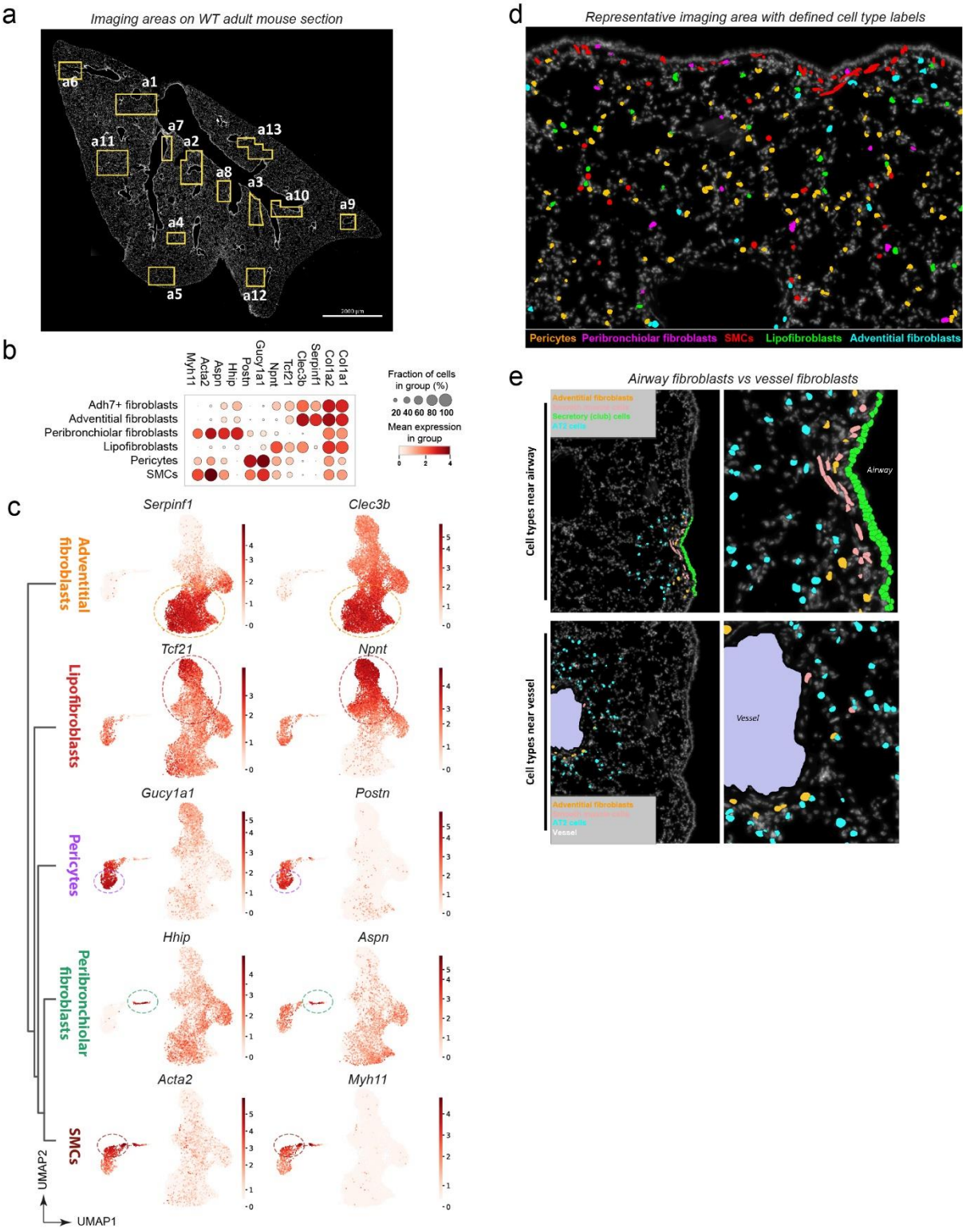
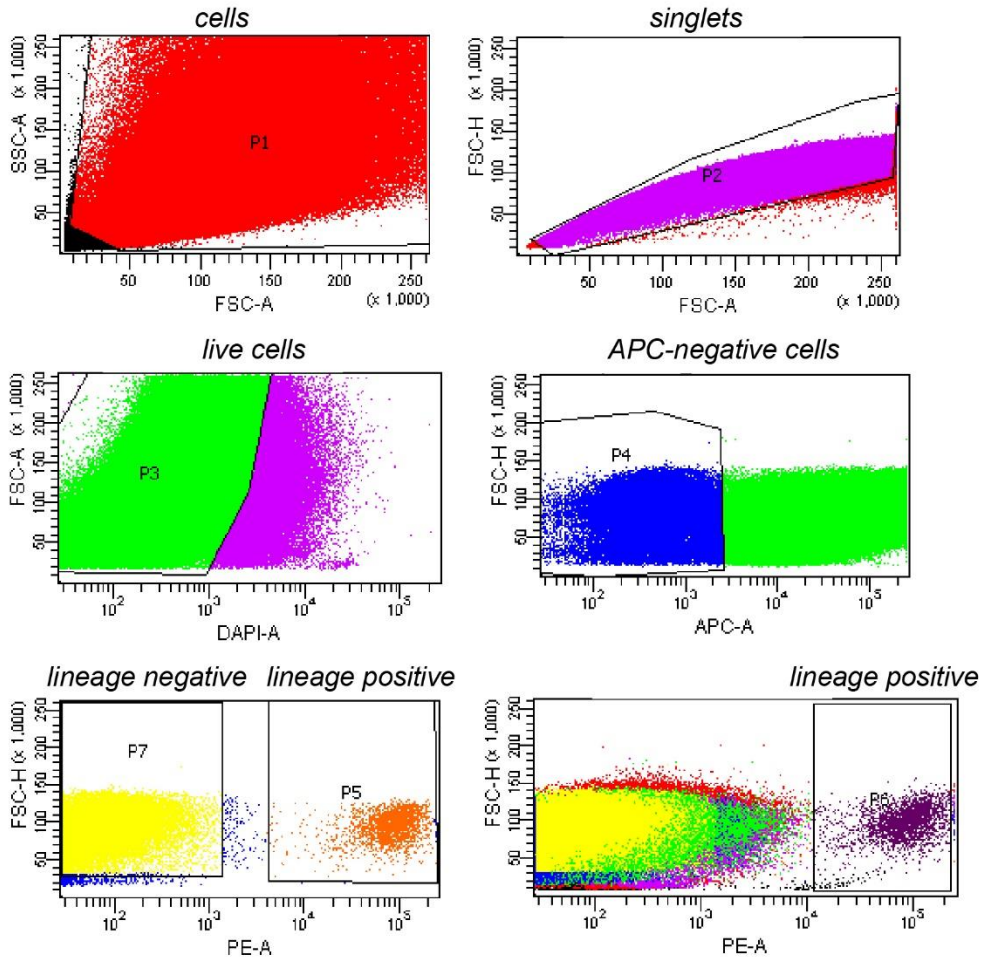


Figure S2. SCRINSHOT analysis of mesenchymal subtypes. **a)** Regions shown were used for SCRINSHOT on a healthy mouse lung cryo-section. **b)** Dotplot shows gene expression of the two meta cell type marker genes that were chosen for SCRINSHOT analysis. **c)** UMAP plots depict gene expression of the two meta cell type marker genes that were chosen for SCRINSHOT analysis for each mesenchymal cell type. **d)** Image depicts representative region of the mouse lung with cells labeled based on their SCRINSHOT marker expression. **e)** Images show

representative regions for airway and vessel domains. **f)** Raw images show interaction of mesenchymal cell types with AT-2 cells in alveolar regions.

a

Examplary sorting startegy for Tcf21-lineage positive and negative cells for scRNA-seq



Tube: SORT 2

Population	#Events	%Parent	%Total
All Events	961,960	####	100.0
P1 *cells*	894,728	93.0	93.0
P2 *singlets*	864,742	96.6	89.9
P3 *live cells*	803,364	92.9	83.5
P4 *mesenchymal cells*	42,819	5.3	4.5
P5 *lineage positive*	2,122	5.0	0.2
P7 *lineage negative*	27,211	63.5	2.8
P6 *lineage positive vs all cells*	2,418	0.3	0.3

b

channel	staining	color
DAPI	live/dead	Zombie Violet
APC	Epcam	APC
	CD45	APC
	CD31	APC
	Lyve1	APC
	Ter119	APC
PE	Tcf21-labeled cells	td-tomato

Figure S3. Gating strategy for Tcf21 lineage positive and negative mesenchymal cells.

a) Exemplary gating strategy for the analysis of Epcam-/CD45-/CD31-/Lyve-/Ter119- negative selected, living mesenchymal cells from Tcf21-Td-tomato mice, that were split into lineage positive (gate P5) and lineage negative cells (gate P7). **b)** List of channels, stained markers and colors used for the sorting.

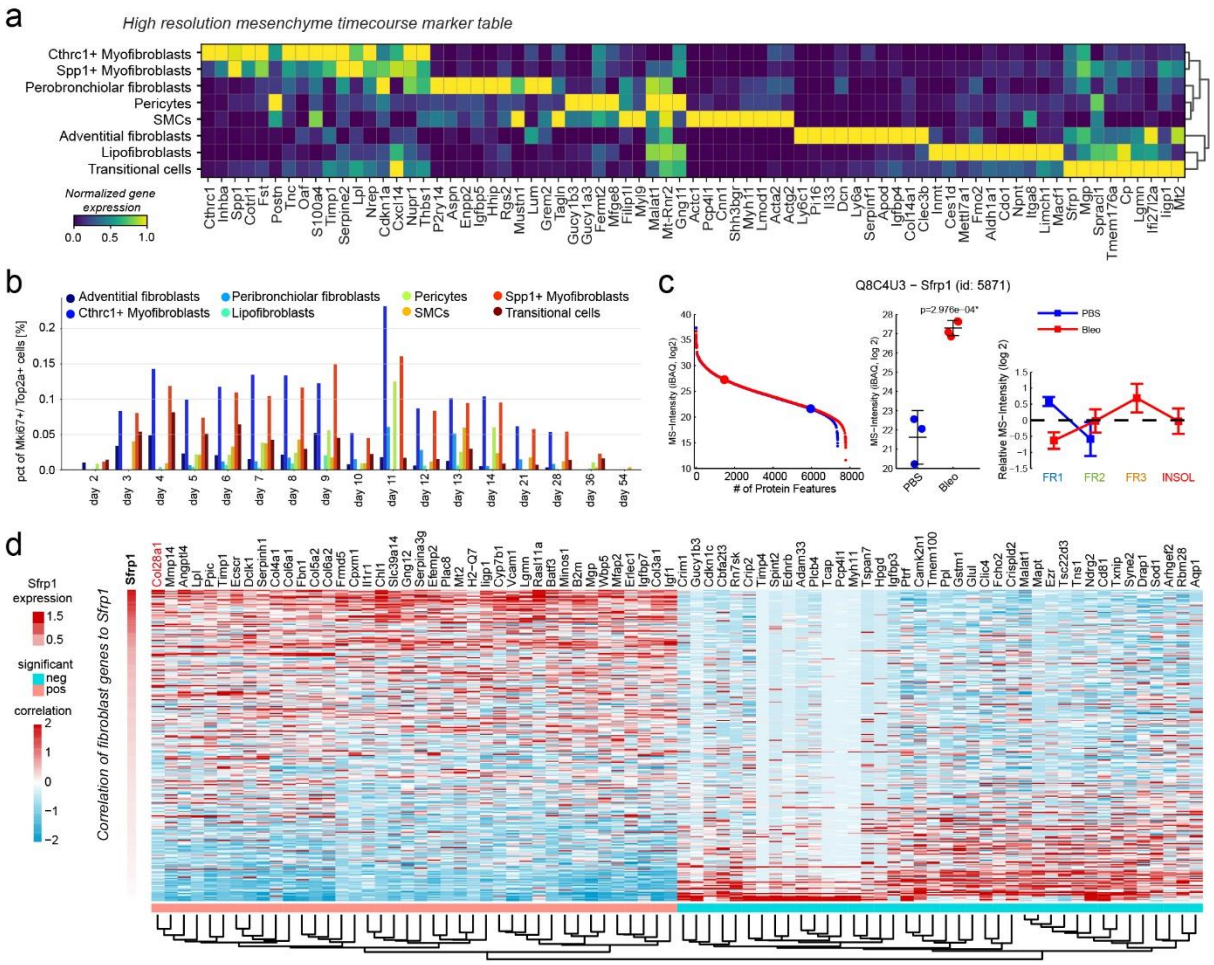


Figure S4. Bleomycin time course analysis. a) Matrixplot of marker genes for the high temporal resolution bleomycin time course of mouse lung mesenchymal cell types. **b)** Proliferation analysis shows the percentage of cells expressing Mki67 or Top2a per cell type and time point. **c)** Quantitative proteome data depict the protein abundance rank of Sfrp1 relative to all other quantified proteins, the iBAQ-normalized mean MS-intensity between healthy PBS treated control mice and Bleomycin treated mice from day 14, and the QDSP profiles indicative for the detergent solubility. The mean and standard error of the mean are shown (day 3, *n* = 3; day 14, *n* = 7; day 28, *n* = 4; day 56, *n* = 3) [25]. **d)** The heatmap shows the expression of genes most strongly associated with Sfrp1 expression within the mesenchymal cells (bottom colorbar indicates significant correlation or anti-correlation).

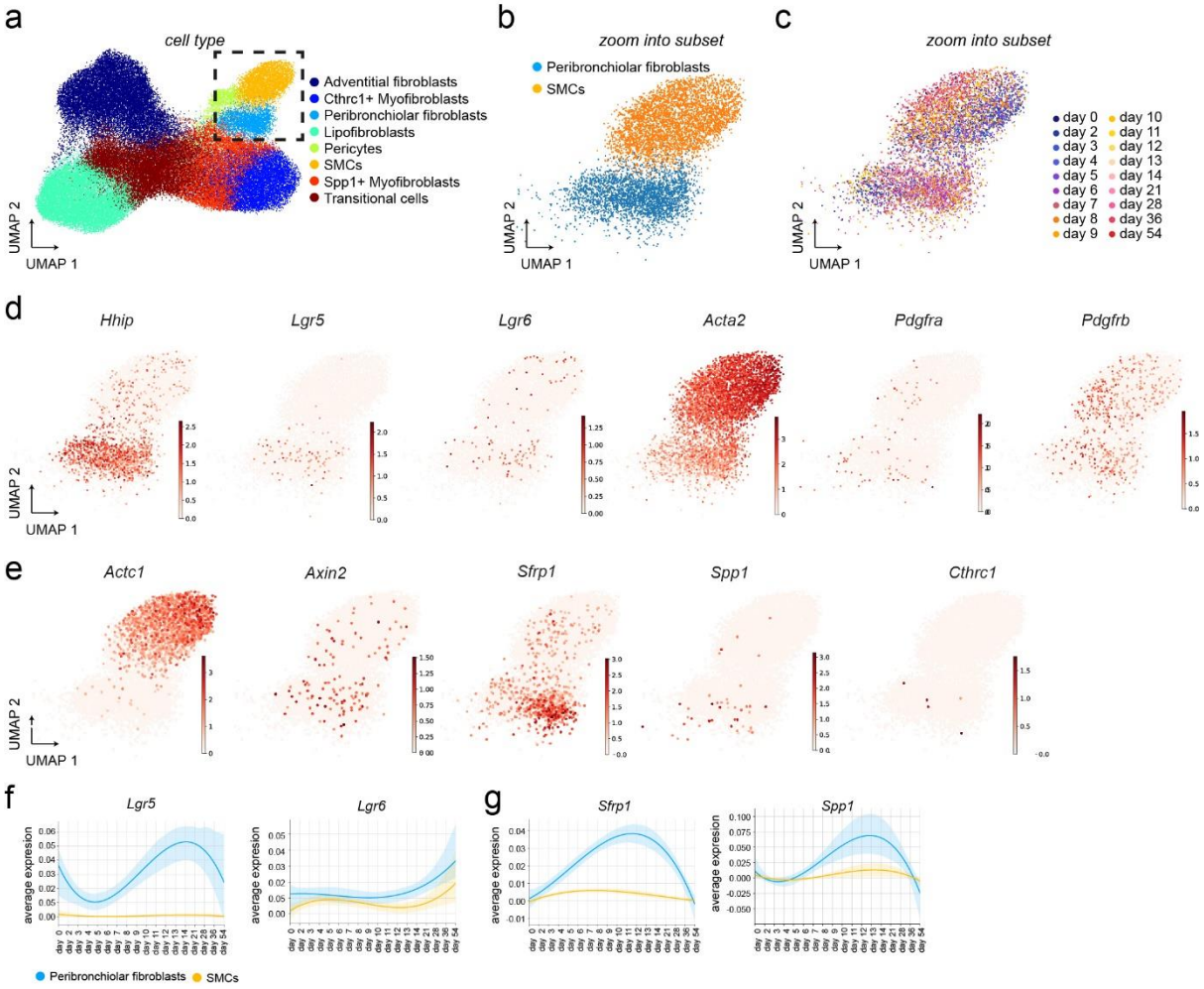
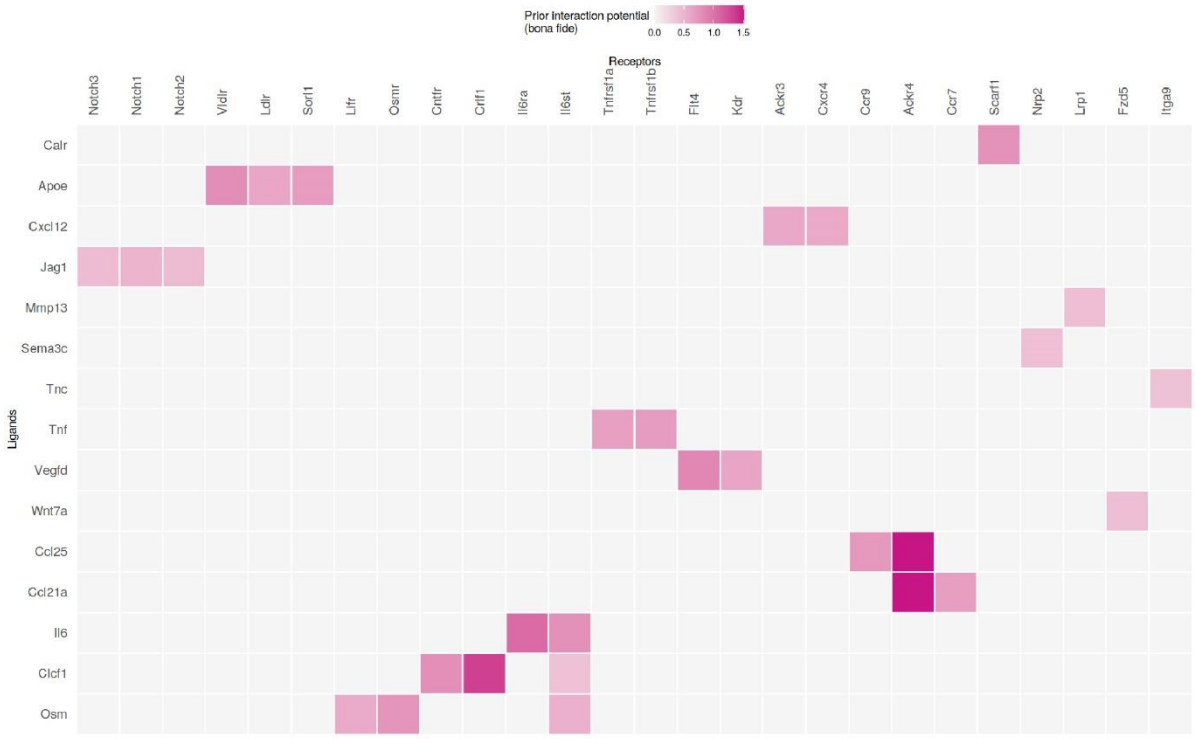


Figure S5. Sub-clustering/zoom in on Hhip+ Peribronchiolar fibroblasts from time course data. a) UMAP embeddings display cells colored by cell type identity. Rectangle indicates zoom into a subset of Peribronchiolar fibroblasts and Smooth Muscle Cells (SMCs). **b-c)** UMAPS embeddings display subset of cells colored by **b)** cell type identity **c)** time point and gene expression of **d)** Hhip, Lgr5, Lgr6, Acta, Pdgfra, Pdgfrb as well as **e)** Actc1, Axin2, Sfrp1, Spp1 and Cthrc1. **f-g) T**he line plots show the smoothed mean expression over time for the indicated genes within the subset of Peribronchiolar fibroblasts And SMCs with a confidence interval across the mouse replicates.

a



b

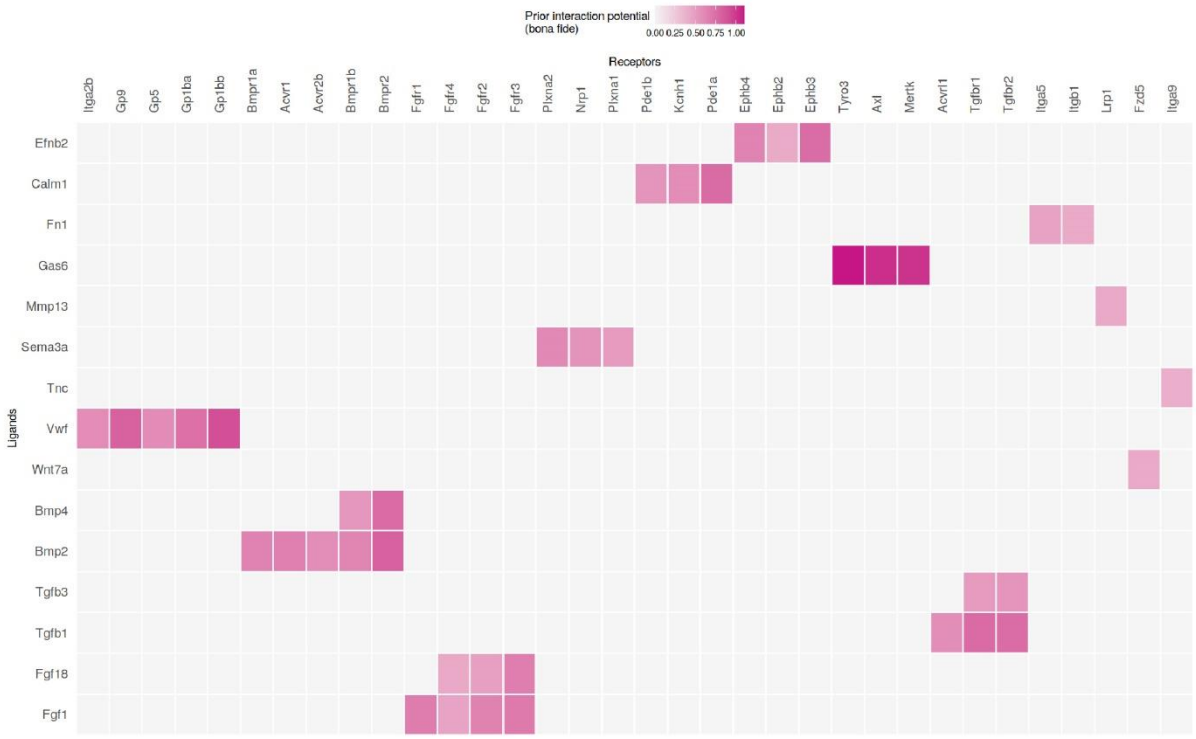


Figure S6. Ligand-receptor pairs. a-b) Heatmaps show the regulatory potential for the given ligand and the potential bona fide receptors in a) the Transition fibroblast fate and b) the Cthrc 1 Myofibroblast fate.

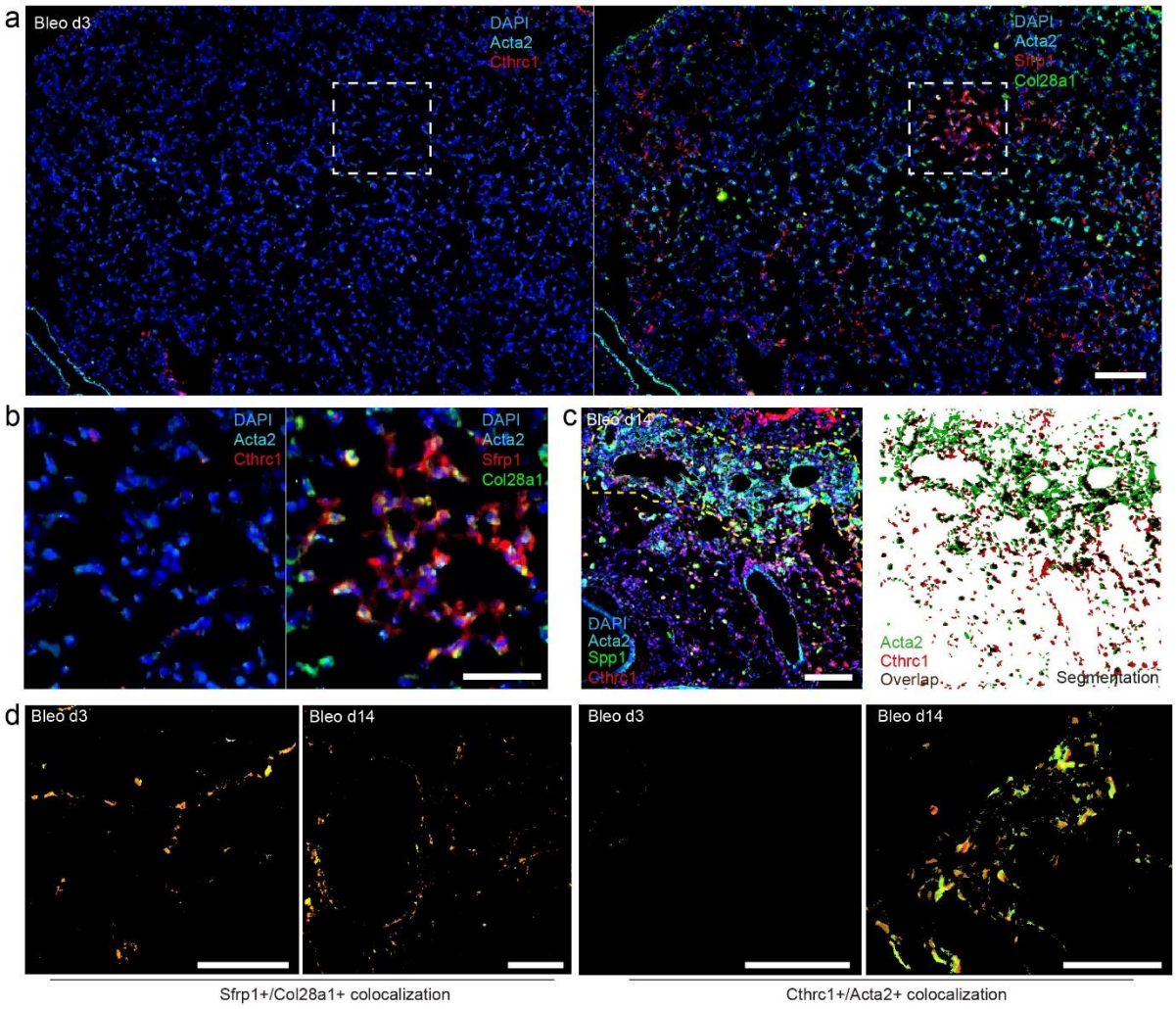


Figure S7. SFRP1+ transitional fibroblasts emerging in parenchymal lung regions and image segmentation. (a) Iterative staining (4i) of parenchymal mouse lung tissue at day 3 after injury indicating emergence of COL28A1 (green)/SFRP1 (red) double positive cells in the left image. The same cells and regions remained negative for CTHRC1 (red) and ACTA2 (cyan). Scale bar = 100 µm. Dashed box indicates magnified view, shown in (**b).** Iterative staining (4i) of parenchymal lung tissue at day 3 after injury indicating COL28A1 (green)/SFRP1 (red) double positive cells in the left image. The same cells, as indicated by white dashed lines, were found to be CTHRC1 (red) and ACTA2 (cyan) negative. Scale bar = 50 µm. **(c)** Representative image demonstrating segmentation which was used for the quantification of appearance of ACTA2, CTHRC1, SPP1, COL28A1 and SFRP1 positive cells at various time points after bleomycin treatment. Yellow dashed line indicates a densely fibrotic region. Cell nuclei stained with DAPI in blue, ACTA2 in cyan, SPP1 in green and CTHRC1 in red. The left panel shows software-based segmentation of ACTA2 (green) and CTHRC1 (red). Overlapping segmented objects are depicted in brown. Scale bar = 100 µm. **(d)** Colocalizing SFRP1/COL28A1 and CTHRC1/ACTA2 fluorescence signals indicated by a colocalization mask at day 3 and day 14 after injury. Scale bars = 100 µm.

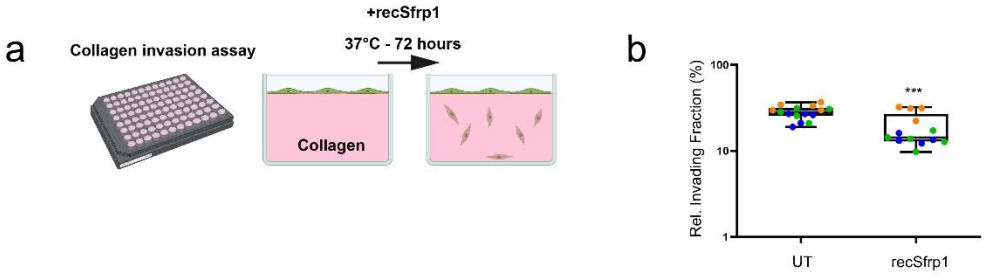


Figure S8. Direct treatment of invading primary human fibroblasts by SFRP1 decreases invasiveness.
(a) Primary human fibroblasts were applied on top of a collagen ECM matrix together with recombinant SFRP1 and left to invade for 72-96 hours. **(b**) Analysis of the cell's invasive capacity was found to be significantly decreased by direct Sfrp1 treatment. Data are expressed as the mean +/- SD. Unpaired two-tailed t-test. ***p < 0.001, n = 3 (fibroblasts from 3 different patients).

Supplementary Methods

Animals

Animal handling, bleomycin/PBS administration, and organ withdrawal were performed in accordance with the Declaration of Helsinki conventions for the use and care of animals and the governmental and international guidelines and ethical oversight by the local government for the administrative region of Upper Bavaria (Germany), registered under ROB-55.2-2532.Vet_02-16-208.

Data availability

ScRNA-seq data was deposited to the Gene Expression Omnibus database. The high temporal mesenchymal enriched Dropseq can be found with the accession code GSE207851, and the Tcf21 labeled mesenchymal 10x data with the accession code GSE207687. Microarray data of primary human fibroblasts (Sfrp1-siRNA knockdown) can be found with the accession code GSE207561.

Human tissue and ethics statement

Primary human lung fibroblasts of non-CLD donors were obtained from the CPC-M bioArchive at the Comprehensive Pneumology Center (CPC Munich, Germany). Participants provided written informed consent to participate in this study, in accordance with approval by the local ethics committee of the LMU, Germany (Project 333-10).

Animals

Pathogen-free female C57BL/6J mice were purchased from Charles River Germany and *Tcf21*iCre(+/-);R26tdTomato(+/-) were maintained in the AG Morty at the Max-Planck-Institut für Herz- und Lungenforschung in Bad Nauheim. Animals were maintained at the appropriate biosafety level at constant temperature and humidity with a 12-hour light cycle. Animals were allowed food and water ad libitum. Animal handling, bleomycin/PBS administration, and organ withdrawal were performed in accordance with the governmental and international guidelines and ethical oversight by the local government for the administrative region of Upper Bavaria (Germany), registered under ROB-55.2-2532.Vet_02-16-208.

Experimental design and bleomycin treatment

Mice were randomly divided into two groups: saline-only (PBS), or bleomycin (Bleo). Lung injury and pulmonary fibrosis were induced by single-dose administration of bleomycin hydrochloride (Sigma Aldrich, Germany), which was dissolved in sterile PBS and given at 2U/kg bodyweight by oropharyngeal instillation. The control group was treated with sterile PBS only. Up to two mice were randomly sacrificed at designated time points (days 0, 2-14, 21, 28, 35, 56) after instillation. With a total of 30 mice, only one mouse was used at days 0, 9, 11, 21, 28, 54. Treated animals were continuously under strict observation with respect to abnormal behavior and signs of body weight loss. *Tcf21*iCre(+/-);R26tdTomato(+/-)-lineage labeled mice, both female and male, were treated with 100mg/kg Tamoxifen 3 days before the start of the bleomycin treatment (3.5U/kg) or NaCl (each 2 mice per lineage) and sacrificed at day 14 (total of 8 mice, 4 lineage-positive, 4 lineage negative, 2 PBS and 2 bleomycin each).

Generation of single cell suspensions from whole mouse lung tissue

Lung single cell suspensions were generated as previously described [1]. Briefly, after euthanasia, the mouse lung was perfused with sterile saline through the heart and all lobes were removed, minced, and transferred for mild enzymatic digestion for 20-30 min at 37°C in an enzymatic mix containing dispase (50 caseinolytic U/ml), collagenase (2 mg/ml), elastase (1 mg/ml), and DNase (30 µg/ml). Cells were collected by straining the digested tissue suspension through a 40-micron mesh and red blood cell lysis (RBC Lysis Buffer, ThermoFisher) was performed for 1 min. After centrifugation at 300 x g for 5 minutes, single cells were taken up in 1 ml of PBS supplemented with 10% fetal calf serum, counted, and critically assessed for single cell separation and overall cell viability.

Negative selection by Magnetic-activated cell sorting (MACS)

For mesenchymal cell enrichment by negative MACS selection, cells were stained with 10 µl per 10 million cells of CD326-AlexaFluor647 antibody (Biolegend, 118212), CD31-APC (Invitrogen, 17-0311-82), CD45-APC (Biolegend, 103112), Lyve1-AF647 (Invitrogen, 50-0443-82) and Ter119-APC (Biolegend, 116218) for 30 min at 4°C in the dark. After washing, the labeled cells were incubated with 100 µl per 10 million cells of MACS dead cell removal beads (Miltenyi Biotec, 130-090-101) and 20 µl of each microbeads specific against AlexaFluor647 (Miltenyi Biotec, 130-091-395) and APC (Miltenyi Biotec, 130-090-855) for 15 min at 4°C. MACS LS columns (Miltenyi Biotec, 130-042-401) were prepared according to the manufacturer's instructions. Cells were applied to the columns and positively labeled cells, bound by the microbeads, were retained in the magnetic column, while the flow-through containing the negatively enriched mesenchymal cells was collected on ice and applied to the Dropseq workflow.

Lineage and negative selection of Tcf21 cells by Fluorescence-activated cell sorting (FACS)

Cells were counted and stained with anti-mouse Epcam CD326-AlexaFluor647 antibody (1:100, Biolegend, 118212), CD31-APC (1:250, Invitrogen, 17-0311-82), CD45-APC (1:250, Biolegend, 103112), Lyve1-AF647 (Invitrogen, 50-0443-82), Ter119-APC (1:100, Biolegend, 116218), and Zombie Violet for viability (Biolegend, 423113). All stainings were performed per 10 Mio cells in 1ml. Cells were stained with Zombie violet according to manufacturer's instructions for 15 min at RT, washed once with PBS and blocked with anti-mouse FC block for 10min on ice. Staining occurred for 20min at 4°C in the dark in the FACS buffer (PBS + 0.1% BSA), and cells were washed once and filtered with a cell strainer (70µM) before subjecting them to FACS sorting. Mesenchymal cells were sorted as ZombieViolet-/APC- and TCF21 lineage was distinguished by the dt-Tomato signal in the PE channel. Data acquisition

was performed in a BD Fortessa flow cytometer (Becton Dickinson, Heidelberg, Germany). 30.000 cells per lineage and mouse were sorted into cooled RPMI media and subject to standard 10x Genomics workflow.

Single cell RNA-sequencing using the Dropseq method

Dropseq experiments were performed according to the original protocols and as previously described[40]. For Dropseq, cells were aliquoted in PBS supplemented with 0.04% of bovine serum albumin at a final concentration of 100 cells/µl. Using the microfluidic device, single cells (100/µl) were co-encapsulated in droplets with barcoded beads (120/µl, purchased from ChemGenes Corporation, Wilmington, MA) at rates of 4000 µl/hr. Droplet emulsions were collected for 10-20 min/each prior to droplet breakage by perfluorooctanol (Sigma-Aldrich). After breakage, beads were harvested, and the hybridized mRNA transcripts reverse transcribed (Maxima RT, ThermoFisher). Unused primers were removed by the addition of exonuclease I (New England Biolabs), following which, beads were washed, counted, and aliquoted for pre-amplification (2000 beads/reaction, equals ca. 100 cells/reaction) with 12 PCR cycles (Smart PCR primer: AAGCAGTGGTATCAACGCAGAGT (100 µM), 2x KAPA HiFi Hotstart Ready-mix (KAPA Biosystems), cycle conditions: 3 min 95°C, 4 cycles of 20s 98°C, 45s 65°C, 3 min 72°C, followed by 8 cycles of 20s 98°C, 20s 67°C, 3 min 72°C, then 5 min at 72°C)[27]. PCR products of each sample were pooled and purified twice by 0.6x clean-up beads (CleanNA), following the manufacturer's instructions. Prior to tagmentation, complementary DNA (cDNA) samples were loaded on a DNA High Sensitivity Chip on the 2100 Bioanalyzer (Agilent) to ensure transcript integrity, purity, and amount. For each sample, 1 ng of pre-amplified cDNA from an estimated 1000 cells was tagmented by Nextera XT (Illumina) with a custom P5-primer (Integrated DNA Technologies). Single-cell libraries were sequenced in a 100 bp paired-end run on the Illumina HiSeq4000 using 0.2 nM denatured sample and 5% PhiX spike-in. For priming of read 1, 0.5 µM Read1CustSeqB was used (primer sequence: GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC).

Single cell RNA-sequencing using the 10x Genomics method

The sorted cells were diluted to 1000 cells/µl and loaded on the 10× Chromium Next GEM Chip G with a targeted cell recovery of 10,000. The following steps were completed according to the manufacturer's protocol (Chromium Next GEM Single Cell 3' Reagent Kits v3.1). Libraries have been pooled according to their minimal required read counts (35,000 or 50,000 reads/cell for 3' gene expression libraries, 20,000 reads/cell for 5' gene expression libraries, and 5000 reads/cell for TCR libraries). Illumina paired-end sequencing was performed with 150 or 200 (3' gene expression) on a NovaSeq 6000.

Processing of the high-temporal resolution bleomycin mesenchymal data set

For the high-resolution mesenchymal lung data set, the Dropseq computational pipeline was used (version 2.0) as previously described[40]. Briefly, STAR (version 2.5.2a) was used for mapping [2]. Reads were aligned to the mm10 reference genome (provided by the Dropseq group, GSE63269). Downstream analysis was performed using the scanpy package (v1.6.0). The top 5000 cells with the highest number of transcripts were read in per sample and subjected to quality control. We assessed the quality of our libraries and excluded barcodes with less than 200 genes detected and retained those with several transcripts between 200 to maximal 6000. A high proportion (> 10%) of transcript counts derived from mitochondria-encoded genes may indicate low cell quality, and we removed these unqualified cells from downstream analysis.

To mitigate effects of background mRNA contamination, we employed the R library SoupX [3]. We manually set the contamination fraction to 0.3 and corrected the count matrices with adjustCounts(). The expression matrices were normalized with scran's size factor based approach [4] and log transformed via scanpy's pp.log1p() function. Variable genes were selected sample-wise (flavor = "cell_ranger"), excluding known cell cycle genes. Those genes being ranked among the top 4,000 in at least 6 samples were used as input for principal component analysis (8,209 genes). To filter the data further, the cells were clustered and clusters expressing no mesenchymal cell marker Col1a1 were excluded from the data set. The aligned bam files were used as input for Velocyto [5] to derive the counts of unspliced and spliced reads in loom format. Next, the sample-wise loom files were combined, normalized and log transformed using scvelos (https://github.com/theislab/scvelo) functions normalize_per_cell() and log1p(). After merging the loom information to the exported .h5ad file using scvelos merge() function the object was scaled and the neighborhood graph constructed with Batch balanced KNN (BBKNN)[6] to account for the different PCR cycles used in the experiment with neighbors_within_batch set to 15 and n_pcs to 40. Several HVGS genes causing an artificial split up of clusters were removed manually. Two dimensional visualization and clustering were carried out with the scanpy functions tl.louvain() at resolution two and tl.umap().

Time course differential expression analysis: to identify genes that show differential expression patterns across time within a given cell type we performed the following analysis (already described [7]). We used the R packages splines and lmtest for our modeling approach. Within each cell type we modeled gene expression as a binomial response where the likelihood of detection of each gene within each mouse sample was the dependent variable. Therefore, the sample size of the model was the number of mouse samples ($n = 28$) and not the number of cells. To assess significance, we performed a likelihood-ratio test between the following two models. For the first model, the independent variables contained an offset for the log-transformed average total UMI count and natural splines fit of the time course variable with two degrees of freedom. The independent variables of the second model just contained the offset for the log-transformed average total UMI count. The dependent variable of both models was the number of cells with UMI count greater than zero out of all cells for a given cell type and mouse sample.

Trajectory inference with CellRank

To model the cell state dynamics within the mesenchymal cells, the CellRank algorithm [8] was used. The time course single cell data was separated into three distinct time frames day2-day5, day6-d21, and day28-day54 and each was processed separately. On the subsetted embeddings, the velocity information recovered, and CellRank initiated by defining the combined kernel with weighing the velocity kernel *0.8 and the connectivity kernel wit *0.2, before computing the sure components and macrostates on the cell type level, until each cell type was represented by at least one macrostate; multiple macrostates per cell type were combined into one. Terminal cell states were set for each time frame separately as shown in the figure and the absorption probabilities were calculated. For each terminal state, the lineage drivers were computed.

Cell-cell communication analysis with NicheNet

NicheNet analysis was used to identify putative upstream regulator and ligand-receptor interactions regulating trajectories towards the transitional and Cthrc1+ myofibroblast state using the R package

nichenetr [9] (version 1.0). To analyze niche signals that drive trajectories towards these two states, we defined them as receiver cells and all cell types contained in a whole mouse lung gene expression dataset from bleomycin and PBS mice at day 14 generated with 10x Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (data not shown) as sender cells.
To define a geneset of interest we used CellRank to compute the top 200 unique lineage driver genes towards the respective cell states. Next, we identified a list of potentially active ligands based in sender cells (% expressed > 10%) and ranked these potential ligands according to their ability (Pearson's correlation coefficient) to predict the gene set of interest. We then selected the top 25 ligands for subsequent analysis and visualization.

Processing of the Tcf21-lineage labeled mesenchymal 10x data

After sequencing, the processing of next-generation sequencing reads of the scRNA-seq data was performed using CellRanger version 3.1.0 (10× Genomics) with a modified genome, by using the mus musculus reference genome (mm10) and manually adding the sequence of the dt-Tomtato gene. We performed unsupervised analysis for multiple subsets of the data with scanpy (v1.6.0). Quality control was performed similarly as described for the Drop-seq data set. We first filtered genes that were expressed in at least 10 cells, scaled cell-wise expression vectors to a total count of 10,000, and logp1-transformed the data and selected highly variable genes (flavor = "cell_ranger"). Barcodes with less than 250 genes detected and a mitochondrial fraction of less than 10% were excluded, and those with a few transcripts between 500 to maximal 50 000 were retained. Again, the expression matrices were normalized with scran's size factor based approach [4] and log transformed via scanpy's pp.log1p() function. Variable genes were selected sample-wise (flavor = "cell_ranger") and known cell cycle genes were excluded. Those genes being ranked among the top 4,000 in at least 3 samples were used as input for principal component analysis (8,162 genes). We then performed a principal component analysis, based on which we computed the k-nearest neighbor graph (n_pcs = 40, n_neighbors = 20). For two-dimensional visualization we used tl.umap(). Clustering was established using the Leiden and Louvain algorithm.

Multiplexed in situ hybridization - SCRINSHOT

SCRINSHOT experiments were performed to an earlier version of the method, now published [10]. Briefly, after euthanasia, two healthy wild-type mouse lungs were washed with PBS and immediately inflated with 4% paraformaldehyde. For frozen sections, tissue was fixed for 1h at room temperature, embedded in OCT media on dry ice and stored at -80°C. Thin lung sections (7 µm) were cut on a cryostat. Multiplexed in situ hybridization was performed as described previously. Briefly, frozen sections and their cells were permeabilized with 0.1 M HCl in DEPC water at room temperature for 3 minutes and dehydrated in an ethanol series. An incubation chamber was mounted, and nonspecific mRNA binding was blocked by adding oligo-dT fragments. Three to five gene specific hybridization probes (padlock probes) were added for 5min at 55°C and for 90 min at 45°C. Ligation of the circular padlock probes was ensured by incubation with SplitR enzyme at room temperature overnight. Rolling circle amplification (RCA) of the padlock circle at 30°C overnight followed by a degradation step for non-specific products with uracil-N-glycosylase (UNG) for 45 minutes at 37°C, provided an amplified number of binding sites for labeled detection probes. One specific labeled probe per padlock probe, with probes for the same gene sharing one color, were hybridized to their binding sites for 60 min at room temperature in the dark. After washing, the tissue was hydrated, the incubation chamber removed, and the slide mounted with a cover slip for microscopy. Per hybridization cycle, three genes

and DAPI could be visualized. After successful imaging, the cover slip and mounting media were removed, the tissue dehydrated again, and the procedure repeated from the UNG step onwards with different labeled probes. Probes used in the experiment were directed against: Postn, Gucy1a3, Aspn, Hhip, Myh11, Acta2, Tcf21, Serpinf1, Clec3b, Col1a2, Rtkn2, Scgb1a1, Sftpc, Clalcr. Respective sequences for all padlock and detection probes can be found in Table S5.

Cell culture, siRNA knock-downs and treatments

Primary human fibroblasts (phFbs) and primary mouse fibroblasts (pmFbs) were cultivated in 24-well plates with a density of 1.0 x10^5 cells/ well and in 6-well plates with a cell density of 2.5x10^5 cells/ well. DMEM-F12 medium supplemented with 20 % FBS and 100 U/ml of penicillin/streptomycin was used. For TGF-β1 treatments, cells were initially synchronized using starvation medium (DMEM-F12, 1% FBS and 100 U/ml penicillin/streptomycin) for 24 hours prior to any instillation. Subsequently, cells were induced with 1 ng/ml of human recombinant TGF-β1 (R&D) for 48 hours. Cells were then harvested for protein and mRNA isolation. phFbs were transfected with siRNA oligonucleotides against human Sfrp1 using Lipofectamine RNAiMAX (Invitrogen) for 48 hours, according to the manufacturer's instructions. The siRNAs s12713, s12414 and s12715 (ThermoFisher) were used to target *Sfrp1*. Scrambled control siRNAs (AM4611 and AM4642; ThermoFisher) were used as a negative control. Furthermore, efficiency of siRNA transfection was tested over a period of 6 days, medium exchange was performed after 48 hours and then analyzed further for additional 24, 48 and 96 hours. Cells were harvested and protein isolations were carried out at each timepoint. TGF-β1 treatment of SFRP1-depleted phFbs was accomplished 48 hours post-transfection with siRNAs using starvation medium (DMEM-F12, 1% FBS and 100 U/ml penicillin/streptomycin) with 1 ng/ml of human recombinant TGF-β1 (R&D) for 24 hours. Subsequently, the cells were trypsinized and used for the invasion assays or harvested for protein and mRNA isolation.

Primary mouse fibroblast isolation

C57BL/6J mice were weighed first to calculate dosage of ketamine/Xylazine solution accordingly (100 µl/10 g weight). The mice were anesthetized using 1:1 ratio of ketamine/Xylazine. The mouse lungs were first flushed with 1X PBS via the right ventricle until almost white. The thorax was cleaned next, and the lungs removed. The dissected lung lobes were placed in ice cold sterile 1X PBS within a 6- well plate. Using a sterile scalpel, the lobes were dissected into small 1-2mm pieces and immediately transferred into a 50ml falcon containing collagenase 1 (diluted 5mg per 50µl of 1X PBS). Collagenase digestion was carried on at 37°C at 400 rpm for 1 hour and next transferred into a 70-micron filter placed on a fresh falcon tube. Using the head of a syringe pistol the digested lung pieces were scratched onto the filter and rinsed thoroughly with sterile 1X PBS. The final suspension was centrifuged for 5 mins at 400 rpm at 4°C. Lastly, the supernatant was discarded, and the pellet was resuspended in fresh cell culture medium (20% FBS and 100U/ml of penicillin/streptomycin supplemented DMEM-F12 media) and cultured under standard cell culture conditions.

Fibroblast Invasion assay

Collagen G matrices in 96-well plates (Corning) were prepared as previously described [11]. Primary human fibroblasts were transfected with siRNA oligonucleotides against human Sfrp1 for 48 hours and treated in 6-well plates with or without 1 ng/ml TGFβ1 in 1% FBS for 24 hours. For rescue experiments

the cells were treated with 1 ng/ml TGFβ1 in combination with 1 µg/ml recombinant SFRP1 (R&D Systems) for 24 hours. Succeeding gelation of the collagen, $2x10^4$ cells/well were seeded on top of the collagen plug in 10% FBS. For RHO inhibition cells were seeded on top of the collagen and treated with 0.5 µg/ml RHO inhibitor (CT04; Cytoskeleton, Inc). Next, primary human (myo)fibroblasts were allowed to infiltrate the collagen matrix in 10% FBS for 96 hours under standard cell culture conditions of 37°C and 5% CO_2. Thereafter, each well was washed twice with PBS and subsequently fixed with 4% paraformaldehyde for 30 minutes at room temperature. Nuclei was stained with DAPI (1:1000 in PBS) overnight. The following day, z-stacks of invaded cell nuclei were imaged using a laser scanning confocal microscope (Zeiss LSM710). Furthermore, quantification of embedded cells and invasion capacity were analyzed with Imaris software (Oxford Instruments) using its spot algorithm.

RHOA activation assay

Activation of RhoA was quantified using the RHOA G-LISA assay kit (Cytoskeleton Inc.) according to the manufacturer's protocol. Briefly, cell lysates from cultured primary human lung fibroblasts were prepared. Concentrations of the cell lysates were analyzed with the Precision Red™ Advanced Protein Assay Reagent (included in the kit). Absorbance was measured by a plate reader at 600 nm and a concentration range between 0.4-2.0 mg/ml was finally used. Subsequently, lysates were added to the coated wells and attachment with the RHO-GTP affinity wells were enhanced with the binding buffer. Following a series of washing steps, anti-RHOA primary antibody and subsequently secondary HRP-labeled secondary antibody were used to quantify the bound RHOA-GTP within the cell lysates. RHOA activation was finally quantified by measuring the absorbance at 490 nm using a microplate spectrophotometer (Tecan reader).

Protein isolation and Western Blotting

Cultured cells were first washed twice with sterile 1X PBS and subsequently scraped from the wells using a cell scratcher in 200µl lysis buffer (RIPA buffer supplemented with 1x Roche complete mini protease inhibitor cocktail and Phospho-Stop phosphatase inhibitor). The cell suspension was then transferred to a vial placed on ice and further mixed on a rotor at 4°C for 1 hour. Following this mixing phase, lysates were subjected to centrifugation at 15,000 rpm for 15 minutes at 4°C to separate the supernatant containing the total protein and pellet with the cell debris. For immunoblotting, samples collected were mixed with Laemmli-SDS-loading buffer and then separated based on molecular weight using a standard 10% SDS PAGE (20 - 30 mA per gel). Proteins were then transferred to methanol activated PVDF (Millipore) membranes (350 mA for 60 minutes). Next, membranes were blocked with 5% milk in 1xTBST (0.1% Tween®20 in TBS) and incubated with primary antibodies at 4°C overnight, followed by HRP-conjugated secondary antibodies for 2 hours at room temperature. On the day of visualization, membranes were initially washed in 1X TBST and then briefly incubated with the chemiluminescent substrates (SuperSignal®, ThermoFisher). The following primary antibodies were used for western blotting: SFRP1 (ab126613; Abcam), CTHRC1 (ab855739; Abcam), α-SMA (A5228; Sigma-Aldrich), and β-Actin−Peroxidase (A3854; Sigma-Aldrich) for normalization.

Microarray analysis

Total RNA was isolated using the PEQGold Total RNA Kit (PeqLab) according to the manufacturer's instructions including gDNA elimination. The Agilent 2100 Bioanalyzer was used to assess RNA quality and RNA with RIN>7 was used for microarray analysis. Total RNA (150 ng) was amplified using the WT

PLUS Reagent Kit (ThermoFisher Scientific Inc., Waltham, USA). Amplified cDNA was hybridized on Human ClariomS arrays (ThermoFisher Scientific). Staining and scanning (GeneChip Scanner 3000 7G) was done according to the manufacturer's instructions. Transcriptome Analysis Console (TAC; version 4.0.0.25; ThermoFisher Scientific) was used for quality control and to obtain annotated normalized SST-RMA gene-level data. Statistical analyses were performed by utilizing the statistical programming environment R (R Development Core Team Ref1). Genewise testing for differential expression was done employing the paired limma *t*-test and Benjamini-Hochberg multiple testing correction (FDR < 10%). To reduce background, gene sets were filtered using DABG p-values<0.05 in at least one sample per pair and in at least two of three pairs per analysis. Array data has been submitted to the GEO database at NCBI (GSE207561).

Fibroblast cell-morphology assessment

To prepare cells for morphology assessment, $5x10^3$ cells/well were plated in 96-well plates using DMEM-F12 media supplemented with 20% FBS and 100 U/ml penicillin-streptomycin. The cells were kept in the incubator for 24h to allow cell attachment and growth under standard cell culture conditions of 37°C and 5% CO2. Subsequently, cells were fixed the next day with 4% PFA for 30 minutes at RT. Next, fixed cells were stained with DAPI (nuclear dye) and Phalloidin (label F-actin) diluted in 1X PBS and incubated overnight at 4°C. Lastly, cells were washed twice with 1X PBS and stored in fresh PBS until imaging. Afterwards, samples were washed three times with PBS for 10 minutes each. Images were acquired with a LSM 710 as z-stacks. Confocal fluorescent z-stacks were volume rendered with Imaris 7.4.0 software (Bitplane). Subsequently, the cell shape parameters were further quantified using Cell Profiler.

Immunofluorescence and iterative stainings (4i-FFPE), microscopy, and image segmentation
For immunofluorescent staining, cells cultured in imaging-plates were fixed in 4% PFA in PBS at 37°C for 30 minutes at room temperature and then permeabilized with 0.5% Triton-X in 4% PFA for 15 minutes. Phalloidin (ThermoFisher) (1:300) and DAPI (1:1000) were mixed in 1% bovine serum albumin (BSA, Sigma) in PBS and added to the samples and incubated at 4°C overnight. Lastly, cells were washed three times with PBS for 10 minutes each. Images were acquired with a LSM 710 as z-stacks and a LD C-Apochromat 40x/1.1 NA water objective lens (Carl Zeiss). For fibroblasts seeded on collagen gels, the cells were fixed with 4% paraformaldehyde (PFA) diluted in 1X PBS at room temperature for 30 mins. Hoechst (Pierce) was diluted in 1% BSA in PBS and incubated at 4°C overnight. Next day the cells were washed by rinsing three times with 1X PBS. Cells were finally imaged in PBS with a LSM 710 as z-stacks. Formalin-fixed paraffin-embedded (FFPE) lung tissue sections from PBS- and Bleomycin-treated mice, as well as from healthy donors and IPF patients, were first placed in a incubator at 60°C for an hour, which was followed by tissue deparaffinization process. Using a Microm HMS 740 Robot-Stainer (ThermoFisher Scientific), the slides containing the tissue sections were automatically incubated with several chemicals as described in table S7. Next, the tissue section containing slides were placed in R-Universal buffer (Aptum Biologics) followed by transfer to an antigen retrieval buffer containing pressure chamber (2100 Retrieval, Aptum Biologics) After 30 mins inside the pressure chamber, the slides were washed once in 1X Tris buffer for 10 min and then incubated in 5% BSA in PBS for 40 mins at room temperature. Subsequently, tissue sections were stained with primary antibodies overnight at 4°C under humid conditions. Next day, the slides were washed twice in 1X PBS for 10 min, and further incubated with fluorescently labeled secondary antibodies for 2 hours at room

temperature under humid conditions. Following two additional washes, slides were then counterstained with DAPI for 1 hour at room temperature, washed again two times with 1X PBS for 10 min and subsequently dried at room temperature. Finally, tissue slides were mounted (Dako mounting medium) and kept in the dark at 4°C until further analysis.
For iterative stainings of FFPE lung tissue sections from PBS- and Bleomycin-treated mice, the iterative indirect immunofluorescence imaging (4i) procedure was adapted to FFPE sections (4i-FFPE) as originally described by Gut et. al. for cultured cells before [12]. Briefly, the FFPE sections were prepared as described above for regular immunofluorescence staining until the blocking step. Then FFPE sections were incubated in 5% BSA in PBS containing 150 mM Maleimide for 45 minutes. FFPE sections were then washed six times in 1X PBS. Next, tissue sections were stained with primary antibodies and subsequently with secondary antibodies and DAPI, similar as for regular immunofluorescence stainings. Following staining steps, slides with FFPE sections were kept in humid conditions and mounted with an imaging buffer (700 mM N-Acetyl-Cyteine in ⅘ ddH2O ⅕ HEPES. pH 7.4) with an unfixed coverslip. After imaging the coverslips and imaging buffer were gently removed and washed 6 times with ddH_2O. Antibodies were then eluted by incubating the samples in an elution buffer (0.5 M L-glycine, 3 M Urea, 3M Guanidium Chloride, 70 mM TCEP-HCl in ddH_2O. pH 2.5) thrice. After each elution step samples were washed 6 times with ddH2O. Next round of staining was performed identically starting with the blocking step. Image segmentation for the quantification of cell-type distributions over time was performed by using 1049 µm * 1049 µm (1.1 mm^2) ROIs from whole mouse lung sections, which were stained with the appropriate antibodies for Cthrc1, Spp1, Acta2, Sfrp1, and Col28a1, and subsequently scanned with a digital slide scanner (Zeiss Axioscan 7). ROIs were cropped and imported into Imaris (Bitplane) software. During segmentation using Imaris' surface algorithm, low intensity objects as well as objects smaller than <10 µm were filtered out. Highly large objects representing mainly structures of blood vessels and airways (e.g., smooth-muscle cells) were manually removed as objects after final segmentation. The remaining objects were counted using Imaris' statistical tools.

mRNA Isolation, cDNA synthesis and qRT-PCR

RNA extraction from cultured cells was performed using the PeqGold RNA kit (Peqlab) according to the manufacturer's instructions. The concentration of the isolated RNA was assessed spectrophotometrically at a wavelength of 260 nm (NanoDrop 1000). cDNA was synthesized with the GeneAMP PCR kit (Applied Biosystems (Foster City, CA, USA)) utilizing random hexamers using 1 µg of isolated RNA for one 301 reaction. Denaturation was performed in an Eppendorf Mastercycler with the following settings: 302 303 lid=45°C, 70°C for 10 minutes and 4°C for 5 minutes. Reverse transcription was performed in an Eppendorf Mastercycler with the following settings: lid=105°C, 20°C for 10 minutes, 42°C for 60 minutes and 99°C for 5 minutes. qRT-PCR reactions were performed in triplicates with SYBR Green I Master in a LightCycler® 480II (Roche, Switzerland) with standard conditions: 95°C for 5 min followed by 45 cycles of 95°C for 5 s (denaturation), 59°C for 5 s (annealing) and 72°C for 20 s (elongation). Target genes were normalized to HPRT expression. Primer sequences used for qPCR are detailed in Table S7.

Code availability

No custom code was developed for the analysis. Packages and parameters used are given in the methods section.

References

1. Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, Mayr CH, Tsidiridis G, Lange M, Mattner LF, Yee M, Ogar P, Sengupta A, Kukhtevich I, Schneider R, Zhao Z, Voss C, Stoeger T, Neumann JHL, Hilgendorff A, Behr J, O'Reilly M, Lehmann M, Burgstaller G, Königshoff M, Chapman HA, Theis FJ, Schiller HB. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* 2020; 11: 3559.

2. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15–21.

3. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* 2020; 9. Available from: http://dx.doi.org/10.1093/gigascience/giaa151.

4. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* 2016; 5: 2122.

5. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. RNA velocity of single cells. *Nature* 2018; 560: 494–498.

6. Polański, K., Young, M. D., Miao, Z., Meyer, K. B., Teichmann, S. A., & Park, J. E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics (Oxford, England)*, *36*(3), 964–965.

7. Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, Mayr CH, Tsidiridis G, Lange M, Mattner LF, Yee M, Ogar P, Sengupta A, Kukhtevich I, Schneider R, Zhao Z, Voss C, Stoeger T, Neumann JHL, Hilgendorff A, Behr J, O'Reilly M, Lehmann M, Burgstaller G, Königshoff M, Chapman HA, Theis FJ, Schiller HB. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* 2020; 11: 3559.

8. Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, Lickert H, Ansari M, Schniering J, Schiller HB, Pe'er D, Theis FJ. CellRank for directed single-cell fate mapping. *Nat. Methods* 2022; 19: 159–170.

9. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 2020; 17: 159–162.

10. Sountoulidis A, Liontos A, Nguyen HP, Firsova AB, Fysikopoulos A, Qian X, Seeger W, Sundström E, Nilsson M, Samakovlis C. SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLoS Biol.* 2020; 18: e3000675.

11. Oehrle B, Burgstaller G, Irmler M, Dehmel S, Grün J, Hwang T, Krauss-Etschmann S, Beckers J, Meiners S, Eickelberg O. Validated prediction of pro-invasive growth factors using a transcriptome-wide invasion signature derived from a complex 3D invasion assay. *Sci. Rep.* 2015; 5: 12673.

12. Gut, G., Herrmann, M. D., & Pelkmans, L. (2018). Multiplexed protein maps link subcellular organization to cellular states. *Science (New York, N.Y.)*, *361*(6401), eaar7042. https://doi.org/10.1126/science.aar7042