Title
RUNX2 promotes fibrosis via an alveolar-to-pathological fibroblast transition

Authors
Fang, Yinshan
Chung, Sanny SW
Xu, Le
et al.

Publication Date
2025-02-05

DOI
10.1038/s41586-024-08542-2

Peer reviewed

RUNX2 promotes fibrosis via an alveolar-to-pathological fibroblast transition

https://doi.org/10.1038/s41586-024-08542-2

Received: 27 March 2023

Accepted: 18 December 2024

Published online: 05 February 2025

Check for updates

Yinshan Fang[1], Sanny S. W. Chung[1], Le Xu[2], Chenyi Xue[3], Xue Liu[4], Dianhua Jiang[4], Rongbo Li[2], Yohei Korogi[2], Ke Yuan[5], Anjali Saqi[6], Hanina Hibshoosh[6], Yuefeng Huang[7], Chyuan-Sheng Lin[8], Tatsuya Tsukui[9], Dean Sheppard[9], Xin Sun[2]✉ & Jianwen Que[1]✉

A hallmark of pulmonary fibrosis is the aberrant activation of lung fibroblasts into pathological fibroblasts that produce excessive extracellular matrix[1–3]. Thus, the identification of key regulators that promote the generation of pathological fibroblasts can inform the development of effective countermeasures against disease progression. Here we use two mouse models of pulmonary fibrosis to show that LEPR$^+$ fibroblasts that arise during alveologenesis include SCUBE2$^+$ alveolar fibroblasts as a major constituent. These alveolar fibroblasts in turn contribute substantially to CTHRC1$^+$POSTN$^+$ pathological fibroblasts. Genetic ablation of POSTN$^+$ pathological fibroblasts attenuates fibrosis. Comprehensive analyses of scRNA-seq and scATAC-seq data reveal that RUNX2 is a key regulator of the expression of fibrotic genes. Consistently, conditional deletion of *Runx2* with *Lepr^{creERT2}* or *Scube2^{creERT2}* reduces the generation of pathological fibroblasts, extracellular matrix deposition and pulmonary fibrosis. Therefore, LEPR$^+$ cells that include SCUBE2$^+$ alveolar fibroblasts are a key source of pathological fibroblasts, and targeting *Runx2* provides a potential treatment option for pulmonary fibrosis.

Pulmonary fibrosis is characterized by the substantial presence of aberrant activated fibroblasts and excessive deposition of extracellular matrix (ECM) proteins (for example, collagens) that in turn leads to a progressively dysfunctional lung[1–4]. Multiple sources of fibrotic fibroblasts have been associated with pulmonary fibrosis. These include GLI1$^+$ stromal cells, TBX4$^+$ resident fibroblasts, AXIN2$^+$ myofibrogenic progenitor cells and PDGFRA$^+$ADRP$^+$ lipofibroblasts[5–8]. More recently, single-cell RNA sequencing (scRNA-seq) analyses have shown that fibrotic changes are associated with the presence of CTHRC1$^+$ pathological fibroblasts in mouse models and in human idiopathic pulmonary fibrosis (IPF)[9]. Results from RNA velocity analyses suggest that these CTHRC1$^+$ cells are derived from alveolar fibroblasts[9], which are essential for maintaining normal alveolar architecture[10,11]. However, cell-fate mapping evidence is lacking. Moreover, although multiple signalling pathways[12–14], including TGFβ and PDGF, have been connected to pulmonary fibrosis, the key downstream regulators remain to be identified, which is in part due to the uncertain origins of these pathological fibroblasts. Thus, investigation of the major sources of pathological fibroblasts is important for uncovering their underlying molecular mechanisms and for developing effective therapies.

Recently, *Lepr* has been shown to mark a subpopulation of stromal cells in the bone marrow[15,16]. After myelofibrosis is induced through the ectopic expression of the glycoprotein hormone thrombopoietin, stromal cells labelled with *Lepr^{cre}* become a major contributor to myofibroblasts[16]. Here we use unbiased scRNA-seq analyses to show that *Lepr* is expressed in mouse lung mesenchymal cells. We generate a *Lepr^{creERT2}* mouse line and used it along with the existing *Lepr^{cre}* mouse line to trace LEPR$^+$ lung mesenchymal cells that arise during neonatal alveologenesis. Notably, the majority of labelled cells are SCUBE2$^+$ alveolar fibroblasts and they give rise to pathological fibroblasts (CTHRC1$^+$POSTN$^+$) after bleomycin challenge, as revealed by cell-fate mapping and scRNA-seq analyses. Genetic ablation of POSTN$^+$ cells attenuates pulmonary fibrotic changes. Results from computational analyses combined with single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) shows that the transcription factor RUNX2 is a key regulator that promotes the generation of pathological fibroblasts. Conditional deletion of *Runx2* with *Lepr^{creERT2}* or *Scube2^{creERT2}* blocks the generation of pathological fibroblasts and ECM deposition. Together, our findings support the concept that alveolar fibroblasts, a major constituent of LEPR$^+$ fibroblasts, serve as a key source of pathological fibroblasts, the generation of which depends on RUNX2.

LEPR$^+$ cells contribute to pathological fibroblasts

A re-analysis of scRNA-seq data of developing mouse lungs[17] revealed that *Lepr* transcripts were present in multiple mesenchymal cells, including alveolar fibroblasts, secondary crest myofibroblasts, adventitial fibroblasts and pericytes and endothelial cells (Fig. 1a,b

[1]Columbia Center for Human Development and Division of Digestive and Liver Disease, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, USA. [2]Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, CA, USA. [3]Division of Cardiology, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY, USA. [4]Department of Medicine and Women's Guild Lung Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [5]Division of Pulmonary Medicine, Department of Pediatrics, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA. [6]Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, USA. [7]Department of Microbiology and Immunology, Columbia University, New York, NY, USA. [8]Department of Pathology and Cell Biology and Transgenic Mouse Shared Resource, Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA. [9]Division of Pulmonary, Critical Care, Allergy and Sleep, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ✉e-mail: xinsun@ucsd.edu; jq2240@cumc.columbia.edu

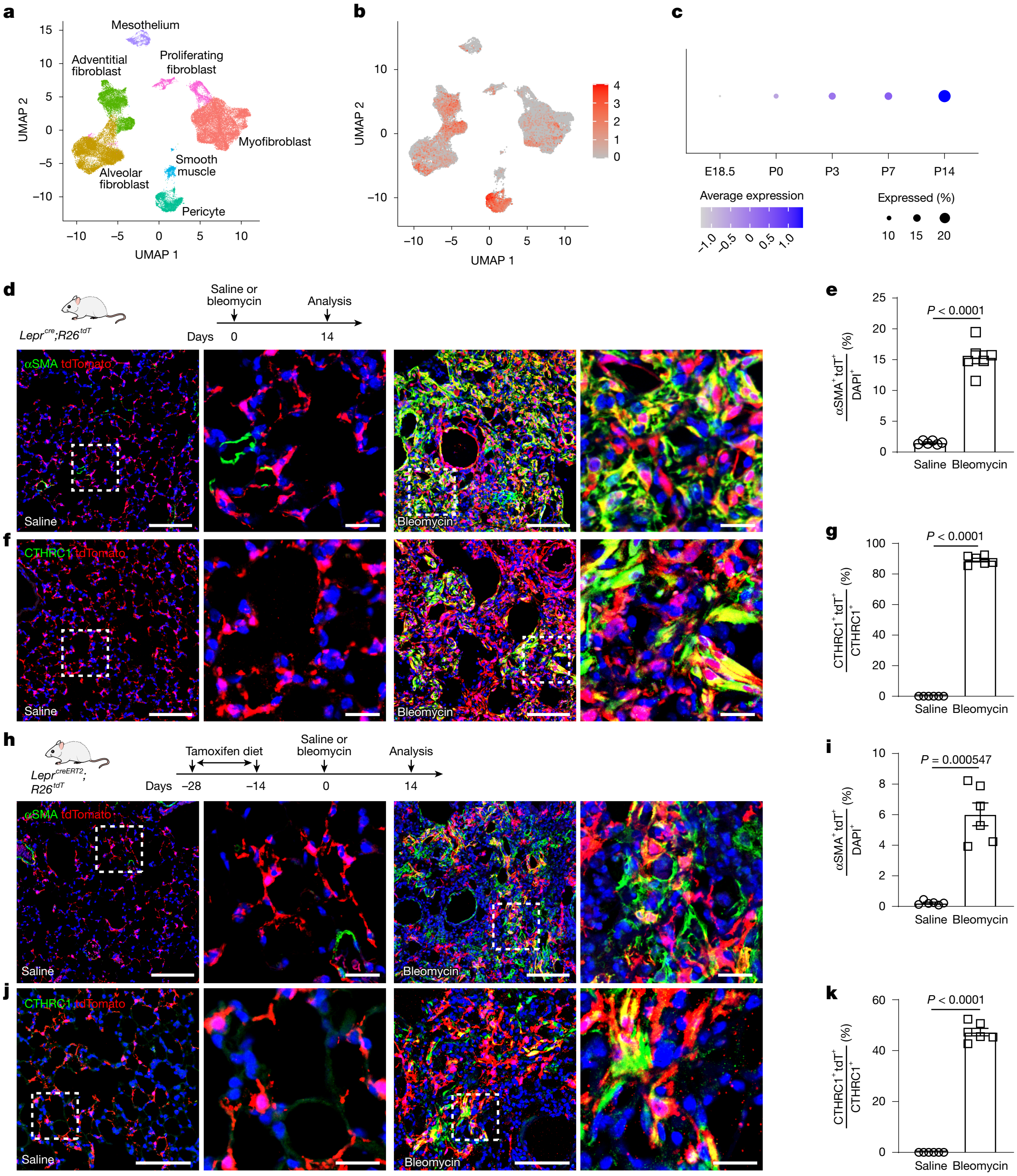


Fig. 1 | *Lepr*-expressing lung mesenchymal cells generate pathological fibroblasts during pulmonary fibrosis. a, Uniform manifold approximation and projection (UMAP) plot showing multiple mesenchymal cell populations in the mouse lungs collected at E18.5, P0, P3, P7 and P14. **b**, Expression of *Lepr* in lung mesenchymal cells. **c**, Increased expression of *Lepr* in lung mesenchymal cells during alveologenesis. **d**, Top, schematic depicting the treatment of *Lepr*^{cre};*R26*^{tdT} mice with bleomycin. Bottom, representative images of αSMA^{+}tdT^{+} cells in lung tissue. **e**, Quantification analysis showing that αSMA^{+}tdT^{+} cells are increased following bleomycin treatment (saline, $n = 6$; bleomycin, $n = 6$). **f**, Representative images showing immunostaining of CTHRC1 and tdTomato in lung tissue. **g**, Quantification analysis showing the enrichment of CTHRC1^{+}tdT^{+} pathological fibroblasts after bleomycin challenge (saline, $n = 6$; bleomycin, $n = 6$). **h**, Top, schematic showing the use of tamoxifen-containing chow and bleomycin challenge of *Lepr*^{creERT2};*R26*^{tdT} mice. Bottom, representative images of αSMA^{+}tdT^{+} cells in lung tissue. **i**, Quantification analysis showing that αSMA^{+}tdT^{+} cells are increased following bleomycin treatment (saline, $n = 6$; bleomycin, $n = 6$). **j**, Representative images showing immunostaining of CTHRC1 and tdTomato in lung tissue. **k**, Quantification analysis showing that CTHRC1^{+}tdT^{+} pathological fibroblasts are increased after bleomycin challenge (saline, $n = 6$; bleomycin, $n = 6$). Data are mean ± s.e.m. Data are representative of at least three independent experiments. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction (**e**,**g**,**i**,**k**). For **d**, **f**, **h** and **j**, images on the right are magnified views of the dashed square in the left-hand images. Scale bars, 100 μm (20 μm in magnified views).

and Extended Data Fig. 1a–e). The level of *Lepr* expression gradually increased during postnatal alveologenesis and was abundant at postnatal day 14 (P14) but not embryonic day 18.5 (E18.5) (Fig. 1c). In line with this finding, few tdTomato-positive (tdT^{+}) cells were observed in the lungs of *Lepr*^{cre};*Rosa26*^{tdtomato} (*R26*^{tdT}) mice at P0 (Extended Data Fig. 1f). By contrast, we observed an increased presence of tdT^{+} cells in the lung parenchyma at P10, and tdT^{+} mesenchymal cells were observed throughout the alveoli at P30 and up to P60 (Extended Data Fig. 1f). Consistent with the scRNA-seq data, few tdT^{+}αSMA^{+} cells (<0.1% of tdT^{+}αSMA^{+} cells) were present in the peribronchiolar and perivascular areas, which suggests that a small number of airway and vascular smooth cells expressed *Lepr* (Extended Data Fig. 1g). We generated a *Lepr*^{creERT2} knock-in mouse strain to confirm the derivatives of LEPR$^+$ cells (Extended Data Fig. 1h). Consistently, tdT$^+$ cells were observed in the lungs of P14 and P20 *Lepr*^{creERT2};*R26*^{tdT} mice that received a single dose of tamoxifen at P7 and P14, respectively (Extended Data Fig. 1i,j). When tamoxifen-containing chow was used to label LEPR$^+$ cell derivatives in the adult lung, we observed an extensive presence of tdT$^+$ cells in the lung parenchyma (Extended Data Fig. 1k). Notably, 75% of tdT$^+$ cells were in close contact with alveolar epithelial type 2 (AT2) cells in the alveoli, a result indicative of alveolar fibroblasts (Extended Data Fig. 1k). Consistently, a re-analysis of scRNA-seq data from adult mouse lungs[9,18,19] showed that *Lepr* and the alveolar fibroblast marker *Scube2* were colocalized in alveolar fibroblasts (Extended Data Fig. 1l–p). Moreover, in situ hybridization with a *Scube2* probe showed that 63.54 ± 1.36% (mean ± s.e.m.) tdT$^+$ cells expressed *Scube2* (Extended Data Fig. 1q,r), whereas 53.09 ± 1.07% SCUBE2$^+$ cells expressed tdTomato (Extended Data Fig. 1s). These findings indicate that the *Lepr*^{creERT2} mouse strain efficiently labels alveolar fibroblasts.

At homeostasis, we did not detect apparent differences in the migration, production of ECM or response to TGFβ1 stimulation between *Lepr*^{cre}-labelled fibroblasts and non-labelled lung fibroblasts (Extended Data Fig. 2a–d). We asked whether the derivatives of LEPR$^+$ cells contribute to pathological fibroblasts during fibrosis. We used two mouse models to address this issue: bleomycin-induced pulmonary fibrosis and silica-induced pulmonary fibrosis. Adult *Lepr*^{cre};*R26*^{tdT} mice were treated with bleomycin, and lungs were examined 14 days after injury (Fig. 1d). Bleomycin induced heterogeneous injuries and fibrosis in the lungs[20,21], and tdT$^+$ cells were significantly expanded in the areas that showed severe damage (Fig. 1d–g and Extended Data Fig. 2e–h). tdT$^+$ cells were positive for the myofibroblast marker αSMA and the pathological fibroblast marker CTHRC1 (Fig. 1d,f). tdT$^+$ cells also expressed collagen I (Extended Data Fig. 2e), one of the major components of ECM deposited by pathological fibroblasts[9,22]. Enrichment of tdT$^+$αSMA$^+$ fibroblasts following bleomycin treatment was further confirmed by fluorescence-activated cell sorting (FACS) analysis (2.83 ± 0.39% (saline) compared with 11.74 ± 0.95% (bleomycin), $P = 0.001051$) (Extended Data Fig. 2i–k). We also challenged *Lepr*^{creERT2};*R26*^{tdT} mice with bleomycin after feeding the mice with tamoxifen-containing chow (Fig. 1h). An extensive number of tdT$^+$ cells expressing αSMA and CTHRC1 were detected in fibrotic foci following bleomycin challenge. This result provides confirmation that LEPR$^+$ cell derivatives expanded when fibrosis occurred (Fig. 1h–k). We then used the silica-induced fibrosis model to further test the contribution of LEPR$^+$ cell derivatives to pulmonary fibrosis in both *Lepr*^{cre};*R26*^{tdT} and *Lepr*^{creERT2};*R26*^{tdT} mice (Extended Data Fig. 2l–q). Silica treatment led to extensive fibrosis surrounding the terminal airways (Extended Data Fig. 2l,o), and tdT$^+$αSMA$^+$ cells were abundant in the fibrotic foci surrounding the terminal bronchioles in the lungs of both *Lepr*^{cre};*R26*^{tdT} and *Lepr*^{creERT2};*R26*^{tdT} mice (Extended Data Fig. 2m,n,p,q). Together, these findings demonstrate that LEPR$^+$ mesenchymal cells that arise during alveologenesis become a key source of pathological fibroblasts that contribute to pulmonary fibrosis.

LEPR$^+$ fibroblasts generate pathological fibroblasts

We next aimed to define the fibroblast subpopulations derived from LEPR$^+$ cells. First, we sorted tdT$^+$ cells from the lungs of *Lepr*^{cre};*R26*^{tdT} mice treated with bleomycin or saline and performed scRNA-seq analysis (Extended Data Fig. 3a). *tdTomato* transcripts were detected in all cells (Extended Data Fig. 3b). Although the majority of the *Lepr*^{cre}-labelled cells were mesenchymal cells, minor populations of epithelium, endothelium and immune cells were also identified (Extended Data Fig. 3c–e). In line with this result, *Lepr* is known to be expressed in immune cells in the bone marrow[15]. To better characterize LEPR$^+$ lung mesenchymal cells and their derivatives in adult lungs, *Lepr*^{creERT2};*R26*^{tdT} mice were fed with tamoxifen-containing chow before bleomycin challenge (Fig. 2a). tdT$^+$ lung mesenchymal cells (EpCAM$^-$CD45$^-$CD31$^-$) were FACS-sorted from the lungs following saline or bleomycin treatment and then analysed by scRNA-seq (Extended Data Fig. 4a–e). In the lungs of saline-treated mice, alveolar fibroblasts and adventitial fibroblasts stood out as two major subpopulations together with minor populations, including peribronchial fibroblasts and pericytes (Extended Data Fig. 4c,d). As expected, bleomycin treatment induced fibrotic changes in the lungs, and the transcript levels of fibrotic genes, including *Acta2*, *Col1a1*, *Col3a1*, *Tnc* and *Fn1*, were significantly increased (Extended Data Fig. 4f).

Re-clustering of *Lepr*^{cre}-labelled fibroblast populations and *Lepr*^{creERT2}-labelled fibroblast populations revealed four subpopulations: alveolar fibroblasts, adventitial fibroblasts, pathological fibroblasts and proliferating fibroblasts (Fig. 2b–e and Extended Data Figs. 3f–i and 4g). Alveolar fibroblasts were identified using the alveolar fibroblast signature, which included the expression of *Scube2, Npnt* and *Inmt*, as previously reported[9,23] (Fig. 2c,f,g and Extended Data Figs. 3h,j–l and 4h,i). Similarly, adventitial fibroblasts were identified using the reported adventitial fibroblast signature[9,23], including the expression of *Pi16* and *Dcn* (Extended Data Figs. 3h,m,n and 4j–l). Immunostaining also confirmed that tdT$^+$ cells expressed PI16 in the adventitial cuff area (Extended Data Figs. 3o and 4m), a region where adventitial fibroblasts are by definition located[9,24,25]. At homeostasis, *Lepr*^{cre}-labelled fibroblasts and *Lepr*^{creERT2}-labelled fibroblasts were predominantly alveolar and adventitial fibroblasts (Fig. 2e and Extended Data Fig. 3i). After bleomycin challenge, both alveolar fibroblast and adventitial fibroblast populations were reduced (Fig. 2e and Extended Data Fig. 3i), which was accompanied by the strong emergence of pathological fibroblasts and proliferating fibroblasts. These changes were associated with the increased expression of cell cycle regulators (for example, CCNA2 and CDK1) (Fig. 2c,e, Extended Data Fig. 3h,i and Supplementary Tables 1 and 2).

Pathological fibroblasts were recently defined by the expression of fibrotic signature genes, including *Cthrc1* and *Postn*[9,23]. Notably, the bleomycin-induced fibroblast population derived from *Lepr*^{cre}-labelled cells and *Lepr*^{creERT2}-labelled cells was also characterized by these genes (Fig. 2c,h, Extended Data Fig. 3h,p,q and Supplementary Tables 1 and 2). Immunostaining confirmed that *Lepr*^{cre}-labelled cells and *Lepr*^{creERT2}-labelled cells in fibrotic foci expressed CTHRC1 (Fig. 1f,j) and POSTN (Extended Data Figs. 3r and 4n). Results from trajectory analysis suggested that *Lepr*^{cre}-labelled alveolar fibroblasts and *Lepr*^{creERT2}-labelled alveolar fibroblasts give rise to pathological fibroblasts (Fig. 2i,j and Extended Data Figs. 3s–u and 5a,b), which is consistent with recent studies[9,26]. Differential gene expression analysis confirmed the enrichment of *Scube2*, *Inmt* and *Npnt* in *Lepr*^{creERT2}-labelled alveolar fibroblasts (Fig. 2k). By contrast, fibrosis-associated genes, including *Col1a1* and *Tnc* along with *Cthrc1* and *Postn*, were highly expressed in pathological fibroblasts (Fig. 2k). Findings from pathway analysis indicated that ECM-associated pathways, including ECM–receptor interaction, ECM organization and collagen biosynthesis, were activated in pathological fibroblasts (Fig. 2l and Extended Data Fig. 5c,d). In line with these findings, *Tgfb1* transcripts and the TGFβ signalling pathway signature were enriched in *Lepr*^{creERT2}-labelled pathological fibroblasts (Fig. 2m and Extended Data Fig. 5e), which indicated the activation of the fibrotic program following bleomycin challenge. Notably, the level of *Lepr* transcripts was significantly decreased after bleomycin treatment (Extended Data Figs. 3v and 5f), which indicated that *Lepr* is not activated during fibrosis.

Lepr^{cre} and *Lepr*^{creERT2} labelled a minor pericyte population (Extended Data Figs. 3c and 4c), and the contribution of pericytes to pathological fibroblasts (myofibroblasts) during lung fibrosis is controversial[27,28]. Therefore, we generated a *Higd1b*^{creERT2} knock-in mouse strain to specifically fate map pericytes in the lung and heart, regions

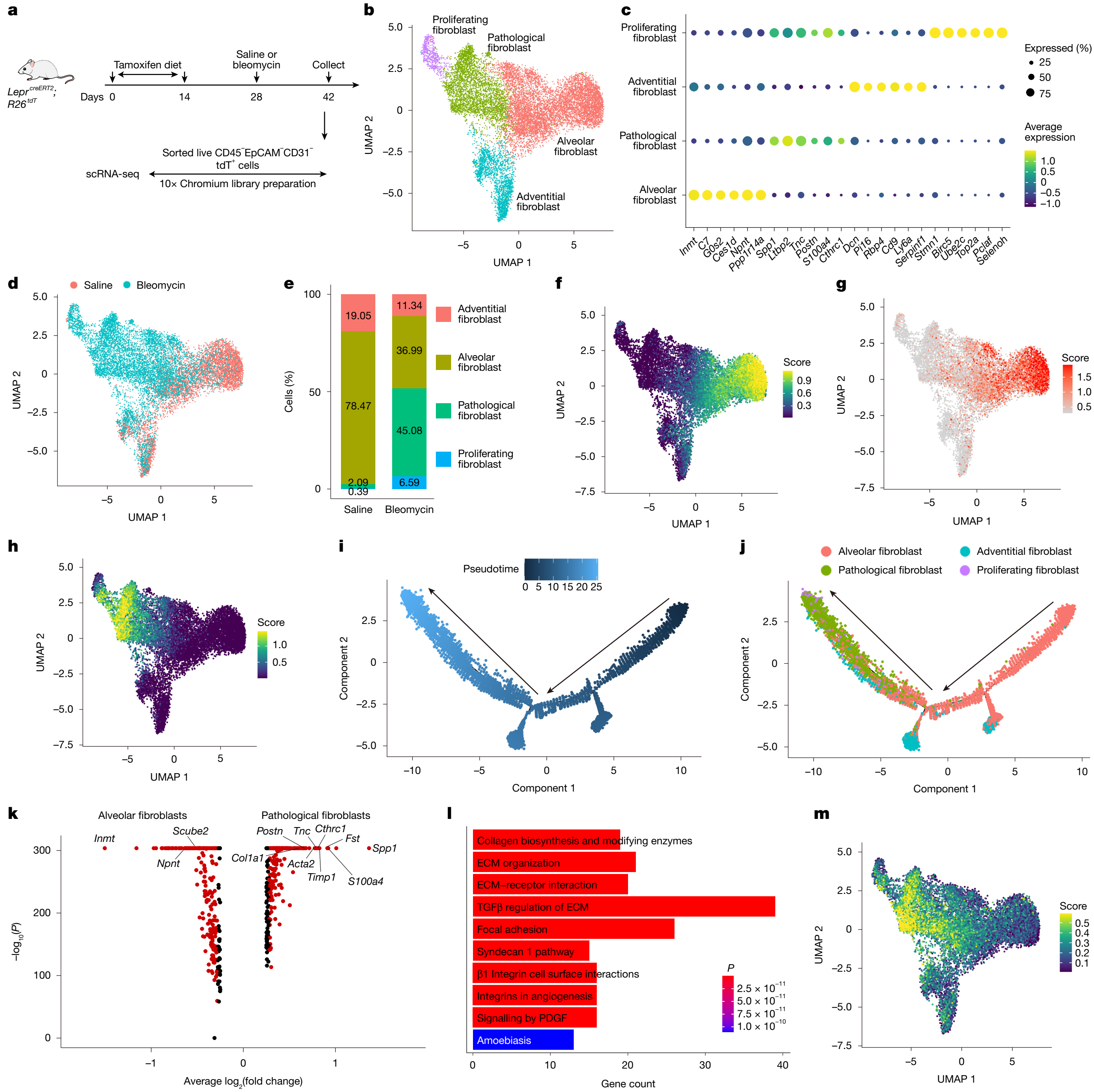


Fig. 2 | scRNA-seq analysis reveals *Lepr*creERT2-labelled pathological fibroblasts induced by bleomycin treatment. **a**, Schematic of the scRNA-seq experimental design. **b**, UMAP plot showing four different *Lepr*creERT2-labelled lung fibroblast populations. **c**, Dot plot showing the representative markers for each fibroblast population. **d**, UMAP of cells from saline-treated lungs and from bleomycin-treated lungs. **e**, Frequency of each fibroblast population in saline-treated lungs and bleomycin-treated lungs. **f**, UMAP plot showing the score of the alveolar fibroblast signature. **g**, UMAP plot showing the expression of *Scube2* in alveolar fibroblasts. **h**, UMAP plot showing the score of the pathological fibroblast signature. **i**,**j**, Trajectory analysis with Monocle2 showing cells ordered by pseudotime (**i**) and cell type (**j**). **k**, Volcano plot showing the differentially expressed genes between alveolar fibroblasts and pathological fibroblasts. **l**, The top ten signalling pathways associated with genes enriched in pathological fibroblasts. **m**, UMAP plot showing the score of the TGFβ signalling pathway signature.

where *Higd1b* is exclusively expressed[29] (Extended Data Fig. 6a–e). We challenged adult *Higd1b*creERT2;*R26*tdT mice with bleomycin after three doses of tamoxifen injection. A few tdT+ cells co-expressing αSMA were observed in fibrotic foci (Extended Data Fig. 6f). However, tdT+ cells rarely expressed CTHRC1 (Extended Data Fig. 6g,h), collagen I or Ki67 (Extended Data Fig. 6i,j). This result suggests that *Higd1b*creERT2-labelled pericytes contribute to αSMA+ cells that surround endothelial cells, but they rarely generate pathological fibroblasts. However, we cannot completely rule out the possibility that a minor subpopulation of pericytes that do not express *Higd1b* contribute to fibrosis (Extended Data Fig. 6a,b). Taken together, our scRNA-seq data confirmed that *Lepr*cre-labelled fibroblasts and

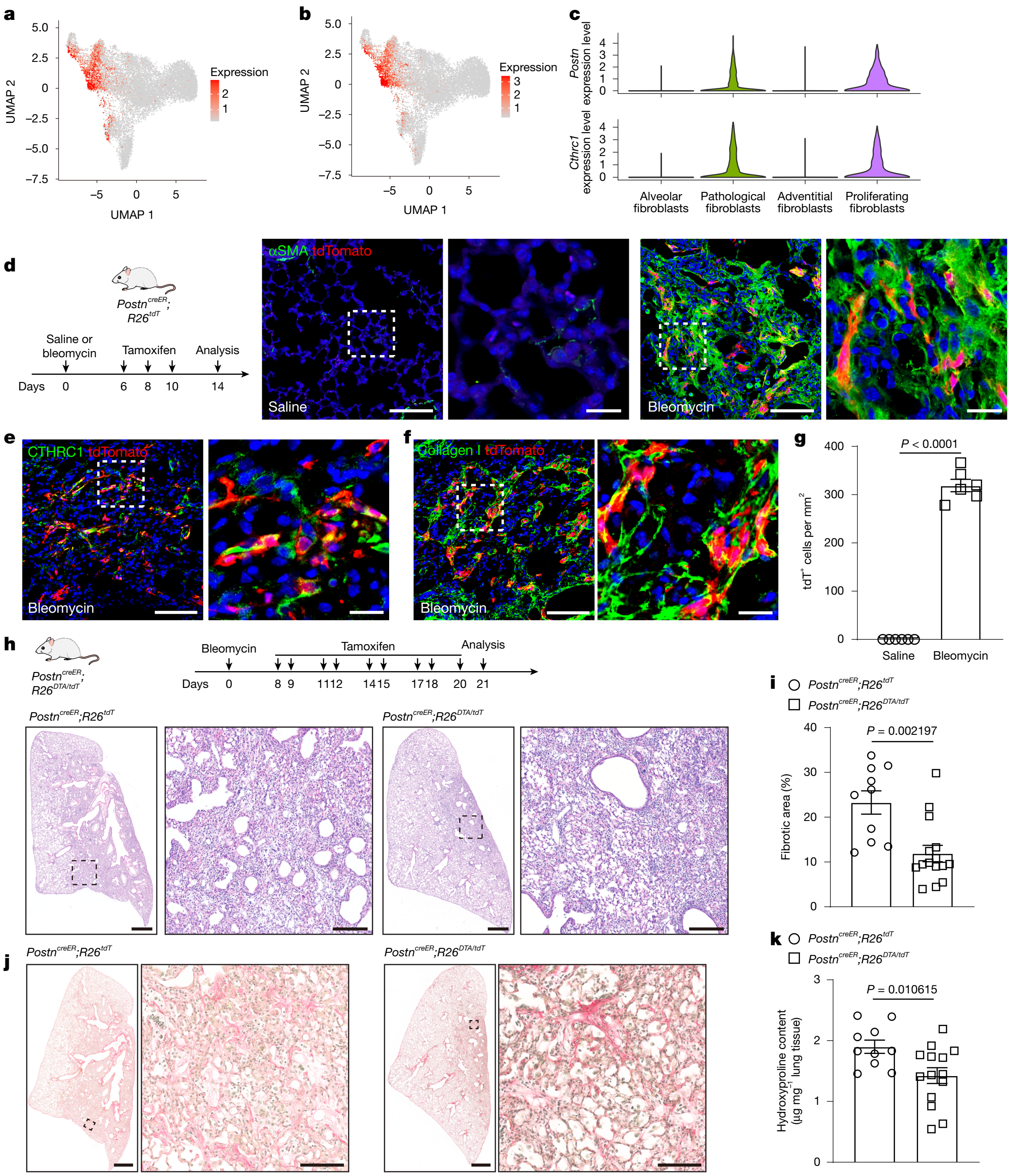


Fig. 3 | Ablation of POSTN⁺ pathological fibroblasts attenuates pulmonary fibrosis. a,**b**, UMAP plot showing the expression of *Postn* (**a**) and *Cthrc1* (**b**) in pathological fibroblasts from the lungs of $Lepr^{creERT2}$;$R26^{tdT}$ mice challenged with saline or bleomycin. **c**, Enriched expression of *Postn* and *Cthrc1* in pathological and proliferating fibroblasts. **d**, Left, schematic of tamoxifen injection and bleomycin challenge of $Postn^{creER}$;$R26^{tdT}$ mice. Right, representative immunostaining images of αSMA and tdTomato in lung tissue. **e**,**f**, Representative images showing CTHRC1⁺tdT⁺ cells (**e**) and collagen I⁺tdT⁺ cells (**f**). **g**, Quantification analysis showing that tdT⁺ cells are increased after bleomycin challenge (saline, *n* = 6; bleomycin, *n* = 6). **h**, Top, schematic of bleomycin challenge and tamoxifen injection of $Postn^{creER}$;$R26^{DTA}$ mice. Bottom, representative images of lung samples stained with haematoxylin and eosin (H&E). **i**, Quantification analysis showing that fibrotic areas are decreased in the lungs of $Postn^{creER}$;$R26^{DTA}$ mice ($Postn^{creER}$;$R26^{tdT}$, *n* = 10; $Postn^{creER}$;$R26^{DTA/tdT}$, *n* = 15). **j**, Representative images of Picro-Sirius Red staining. **k**, Reduced hydroxyproline content in $Postn^{creER}$;$R26^{DTA}$ mice following bleomycin challenge ($Postn^{creER}$;$R26^{tdT}$, *n* = 10, $Postn^{creER}$;$R26^{DTA/tdT}$, *n* = 14). Data are mean ± s.e.m. Data are representative of at least three independent experiments. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction (**g**,**i**,**k**). For **d**–**f**, **h** and **j**, images on the right are magnified views of the dashed square in the left-hand images. Scale bars, 20 µm (magnifications in **d**–**f**), 100 µm (main images in **d**–**f** and magnification in **j**), 200 µm (magnification in **h**), 1 mm (main images in **h** and **j**).

$Lepr^{creERT2}$-labelled fibroblasts include alveolar fibroblasts as a major constituent. After injury, these labelled cells are converted into pathological fibroblasts that express *Cthrc1* and *Postn*. In consideration of recent findings[9,26], LEPR⁺ fibroblasts probably represent a key contributor to pathological fibroblasts.

Ablation of *Postn*⁺ cells attenuates pulmonary fibrosis

Our results from scRNA-seq analyses demonstrated that pathological fibroblasts are characterized by the expression of *Postn* and *Cthrc1*, which were rarely expressed in normal lung fibroblasts (Fig. 3a–c).

Consistently, cell-fate mapping with the knock-in *Postn-MerCreMer* ($Postn^{creER}$) mouse line[30] confirmed that few mesenchymal cells were labelled in the lungs of adult mice treated with saline (Fig. 3d). By contrast, numerous tdT$^+$ cells expressing the myofibroblast marker αSMA, the pathological fibroblast marker CTHRC1 and the ECM component collagen I were present in the fibrotic area following bleomycin challenge (tdT$^+$ cells, 0.00 ± 0.00 per mm^2 (saline) compared with 319.00 ± 12.87 per mm^2 (bleomycin), $P < 0.0001$) (Fig. 3d–g). A re-analysis of published scRNA-seq data[31] also showed that *Postn* is expressed at high levels in pathological fibroblasts after silica exposure (Extended Data Fig. 7a–d). We further validated this finding with the silica injury model, in which substantial accumulation of lineage-labelled fibroblasts (tdT$^+$αSMA$^+$) was observed in the fibrotic areas (tdT$^+$ cells, 0.00 ± 0.00 per mm^2 (saline) compared with 335.80 ± 16.65 per mm^2 (silica), $P < 0.0001$) (Extended Data Fig. 7e,f). Given the abundant presence of POSTN$^+$ pathological fibroblasts during fibrosis, we predicted that ablation of these cells will attenuate fibrotic changes. $Postn^{creER};R26^{DTA}$ mice were treated with bleomycin followed by tamoxifen injection (Fig. 3h). Pulmonary fibrosis was significantly attenuated after tamoxifen injection, as determined by the quantification of fibrotic areas and hydroxyproline content (fibrotic area, 23.28 ± 2.59% (control) compared with 11.90 ± 1.83% (diphtheria toxin (DTA)), $P = 0.002197$; hydroxyproline content, 1.90 ± 0.11 μg mg^{-1} lung (control) compared with 1.43 ± 0.13 μg mg^{-1} lung (DTA), $P = 0.010615$) (Fig. 3h–k), without apparent effects on the pulmonary vasculature (Extended Data Fig. 7g). These findings demonstrate that *Postn* is switched on during the conversion of LEPR$^+$ lung fibroblasts to pathological fibroblasts and that ablation of POSTN$^+$ cells reduces pulmonary fibrosis.

RUNX2 in alveolar–pathological fibroblast conversion

The Enrichr program has been instrumental for enrichment analyses and the identification of key transcription factors that contribute to tissue regeneration and disease[32–34]. We therefore used this program to predict transcription factors that regulate the conversion of alveolar fibroblasts to pathological fibroblasts. Among multiple transcription factors identified, RUNX2 was predicted to regulate the most fibrosis-associated genes (Fig. 4a). Notably, our scRNA-seq analyses of $Lepr^{cre}$-labelled cells and $Lepr^{creERT2}$-labelled cells showed that *Runx2* transcripts were substantially increased and highly enriched in the pathological fibroblast population (Fig. 4b and Extended Data Fig. 8a–c). In parallel, we performed scATAC-seq to profile chromatin accessibility. Mesenchymal cells (EpCAM$^-$CD45$^-$CD31$^-$) were sorted from the lungs of mice treated with bleomycin or saline and subjected to scATAC-seq (Extended Data Fig. 8d). Similar to our scRNA-seq analyses of $Lepr^{cre}$-labelled and $Lepr^{creER}$-labelled populations, multiple fibroblast subpopulations were identified (Extended Data Fig. 8e–g). Again, the pathological fibroblast population was only present in bleomycin-treated lungs (Extended Data Fig. 8e,f). ChromVAR[35] analysis revealed that the RUNX2-binding motif was preferentially enriched in pathological fibroblasts (Fig. 4c). Consistently, chromatin accessibility at the *Runx2* locus was highly enriched in pathological fibroblasts (Fig. 4d). Moreover, the transcript level of *Runx2* was increased in pathological fibroblasts in the lungs of bleomycin-treated mice, as shown by a re-analysis of published scRNA-seq datasets[9,19,36,37] (Extended Data Fig. 8h–l). The findings were further confirmed by quantitative PCR with reverse transcription (RT–qPCR) analysis and the extensive presence of RUNX2$^+$tdT$^+$ cells in the lungs of $Lepr^{cre};R26^{tdT}$ mice following bleomycin challenge (Fig. 4e,f and Extended Data Fig. 8m). Taken together, these data suggest that RUNX2 functions as a key transcription factor that regulates the generation of pathological fibroblasts. Moreover, our re-analysis of a scRNA-seq dataset[38] revealed that *Runx2* is co-expressed with *Cthrc1* in an extensive number of active hepatic stellate cells during liver fibrosis induced by carbon tetrachloride (CCl_4) (Extended Data Fig. 8n–q).

Runx2 insufficiency attenuates pulmonary fibrosis

We next tested whether *Runx2* promotes the generation of pathological fibroblasts during pulmonary fibrosis. $Lepr^{cre};Runx2^{f/f}$ and $Lepr^{creERT2};Runx2^{f/f}$ mice were challenged with bleomycin. Deletion of *Runx2* with $Lepr^{cre}$, which also labelled bone marrow stromal cells, did not alter the lung morphology or the proportion of monocytes and neutrophils in the bone marrow at homeostasis (Extended Data Fig. 9a–e and Supplementary Fig. 1), although they both have been associated with the pathogenesis of pulmonary fibrosis[39,40]. Specific ablation of *Runx2* in $Lepr^{cre}$-derived cells attenuated bleomycin-induced lung fibrosis, as the fibrotic areas were reduced from 24.97 ± 2.24% to 10.96 ± 1.58% ($P = 0.000145$; Extended Data Fig. 9f,g). In line with this finding, lungs from $Lepr^{cre};Runx2^{f/f}$ mutant mice exhibited reduced hydroxyproline content (1.67 ± 0.09 μg mg^{-1} lung (control) compared with 1.30 ± 0.08 μg mg^{-1} lung (mutant), $P = 0.008515$) and αSMA$^+$ myofibroblasts were also decreased in mutant mice (17.25 ± 0.95% (control) compared with 7.68 ± 0.57% (mutant), $P < 0.0001$) (Extended Data Fig. 9h–j). Loss of *Runx2* in the lungs of $Lepr^{creERT2};Runx2^{f/f}$ mice similarly reduced fibrotic areas and hydroxyproline content (fibrotic area, 23.97 ± 2.01% (control) compared with 15.34 ± 1.25% (mutant), $P = 0.003504$; hydroxyproline content, 1.68 ± 0.15 μg mg^{-1} lung (control) compared with 1.26 ± 0.11 μg mg^{-1} lung (mutant), $P = 0.039141$) (Fig. 4g–j and Extended Data Fig. 9k). Moreover, αSMA$^+$ myofibroblasts and CTHRC1$^+$ pathological fibroblasts were decreased in lungs of $Lepr^{creERT2};Runx2^{f/f}$ mice (αSMA$^+$ myofibroblasts, 16.03 ± 0.85% (control) compared with 10.10 ± 0.55% (mutant), $P = 0.000301$; CTHRC1$^+$ pathological fibroblasts, 142.70 ± 8.78 per field (control) compared with 85.67 ± 6.16 per field (mutant), $P = 0.000492$) (Fig. 4k,l and Extended Data Fig. 9l,m). Given that SCUBE2$^+$ alveolar fibroblasts are the major constituent of LEPR$^+$ cell derivatives, we asked whether deletion of *Runx2* using $Scube2^{creERT2}$ attenuates bleomycin-induced fibrosis (Fig. 4m). The fibrotic area and hydroxyproline content were reduced after *Runx2* deletion (fibrotic area, 21.45 ± 1.84% (control) compared with 11.92 ± 1.55% (mutant), $P = 0.001504$; hydroxyproline content, 1.79 ± 0.13 μg mg^{-1} lung (control) compared with 1.07 ± 0.11 μg mg^{-1} lung (mutant), $P = 0.000654$) (Fig. 4m–p and Extended Data Fig. 9n). Moreover, αSMA$^+$ myofibroblasts and CTHRC1$^+$ pathological fibroblasts were also substantially reduced (αSMA$^+$ myofibroblasts, 16.88 ± 0.33% (control) compared with 8.32 ± 0.65% (mutant), $P < 0.0001$; CTHRC1$^+$ pathological fibroblasts, 139.70 ± 7.07 per field (control) compared with 61.33 ± 7.17 per field (mutant), $P < 0.0001$) (Fig. 4q,r and Extended Data Fig. 9o,p). Together, these findings support the conclusion that RUNX2 is important for the conversion of alveolar fibroblasts into pathological fibroblasts during pulmonary fibrosis.

We next re-analysed published scRNA-seq datasets[41] to determine whether *RUNX2* expression is also increased in the lungs of human patients with IPF. Four fibroblast populations were identified: alveolar fibroblasts, adventitial fibroblasts, pathological fibroblasts and proliferating fibroblasts (Fig. 5a and Extended Data Fig. 10a,b). *LEPR* was expressed in alveolar fibroblasts and adventitial fibroblasts, but less so in pathological fibroblasts, whereas *SCUBE2* was barely expressed in human alveolar fibroblasts (Extended Data Fig. 10c,d). *RUNX2* was predominantly expressed by POSTN$^+$CTHRC1$^+$ pathological fibroblasts, which also expressed high levels of ACTA2 and COL1A1 together with other ECM proteins (Fig. 5b–d and Extended Data Fig. 10e–g). Notably, pathological fibroblasts from human and mouse lungs shared key marker genes, including *CTHRC1*, *POSTN*, *TNC*, *ACTA2*, *FN1*, *SPARC*, *MMP14* and various collagens, although some genes were differentially expressed, which is possibly due to the cross-species differences (Extended Data Fig. 10h and Supplementary Table 3). Further analysis of scRNA-seq data confirmed the increased expression of *RUNX2* transcripts in IPF pathological fibroblasts (Fig. 5e,f), which is in line with the bulk RNA-seq data published by two different groups[42,43] (Fig. 5g and Extended Data Fig. 10i). Numerous RUNX2$^+$αSMA$^+$ fibroblasts

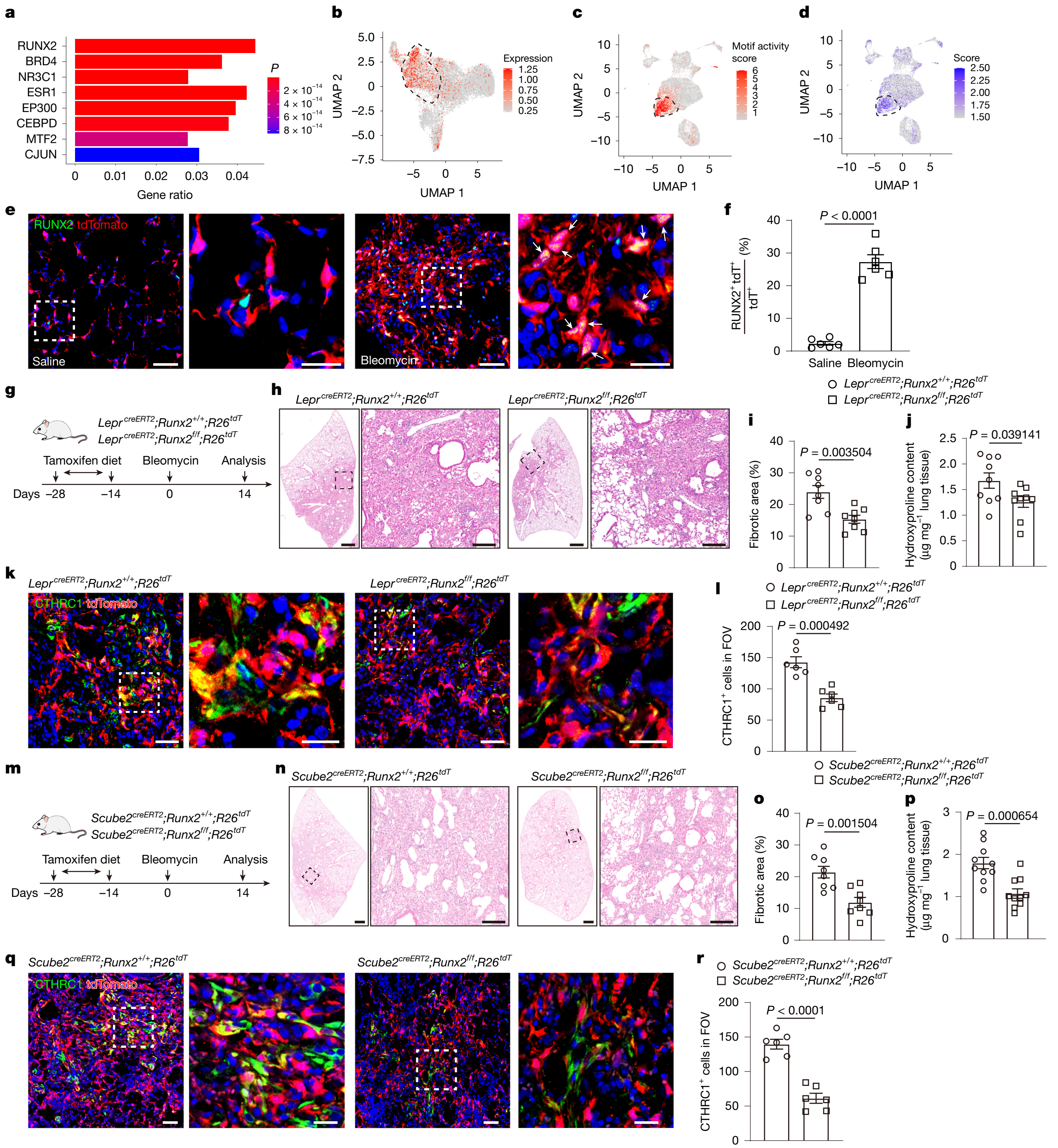


Fig. 4 | *Runx2* insufficiency blocks the generation of pathological fibroblasts and attenuates lung fibrosis. a, The top eight transcription factors predicted to regulate fibrotic genes enriched in pathological fibroblasts. **b**, scRNA-seq UMAP plot showing *Runx2* expression in lung mesenchymal cells of *Lepr*^{creERT2};*R26*^{tdT} mice challenged with saline or bleomycin. **c**, scATAC-seq UMAP plot showing the RUNX2 motif MA0511.2. **d**, scATAC-seq UMAP plot showing *Runx2* chromatin accessibility. **e**, RUNX2^{+}tdT^{+} cells (arrows) in lungs treated with saline or bleomycin. **f**, Quantification of RUNX2^{+}tdT^{+} cells (saline, $n = 6$; bleomycin, $n = 6$). **g**, Schematic of the experimental design. **h**, Representative H&E-stained images. **i**, Quantification of fibrotic areas in the lungs of control ($n = 8$) and *Lepr*^{creERT2};*Runx2*$^{f/f}$;*R26*^{tdT} mice ($n = 8$). **j**, Hydroxyproline content in the lungs of control ($n = 9$) and *Lepr*^{creERT2};*Runx2*$^{f/f}$;*R26*^{tdT} mice ($n = 9$) following bleomycin challenge. **k**, Representative images showing CTHRC1 and tdTomato in the lungs. **l**, Quantification of CTHRC1^{+} cells in the lungs of control ($n = 6$) and *Lepr*^{creERT2};*Runx2*$^{f/f}$;*R26*^{tdT} mice ($n = 6$). **m**, Schematic of the experimental design. **n**, Representative H&E-stained images. **o**, Quantification of fibrotic areas in the lungs of control ($n = 8$) and *Scube2*^{creERT2};*Runx2*$^{f/f}$;*R26*^{tdT} mice ($n = 8$). **p**, Hydroxyproline content in bleomycin-treated lungs of control ($n = 10$) and *Scube2*^{creERT2};*Runx2*$^{f/f}$;*R26*^{tdT} ($n = 10$) mice. **q**, Representative images showing CTHRC1 and tdTomato in the lungs. **r**, Quantification of CTHRC1^{+} cells in the lungs of control ($n = 6$) and *Scube2*^{creERT2};*Runx2*$^{f/f}$;*R26*^{tdT} ($n = 6$). Data are representative of at least three independent experiments. Data are mean ± s.e.m. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction (**f**,**i**,**j**,**l**,**o**,**p**,**r**). For **e**, **h**, **k**, **n** and **q**, images on the right are magnified views of the dashed square in the left-hand images. Scale bars, 20 μm (magnifications in **e**, **k** and **q**), 50 μm (main images in **e**, **k** and **q**), 200 μm (magnifications in **h** and **n**), 1 mm (main images in **h** and **n**). FOV, field of view.

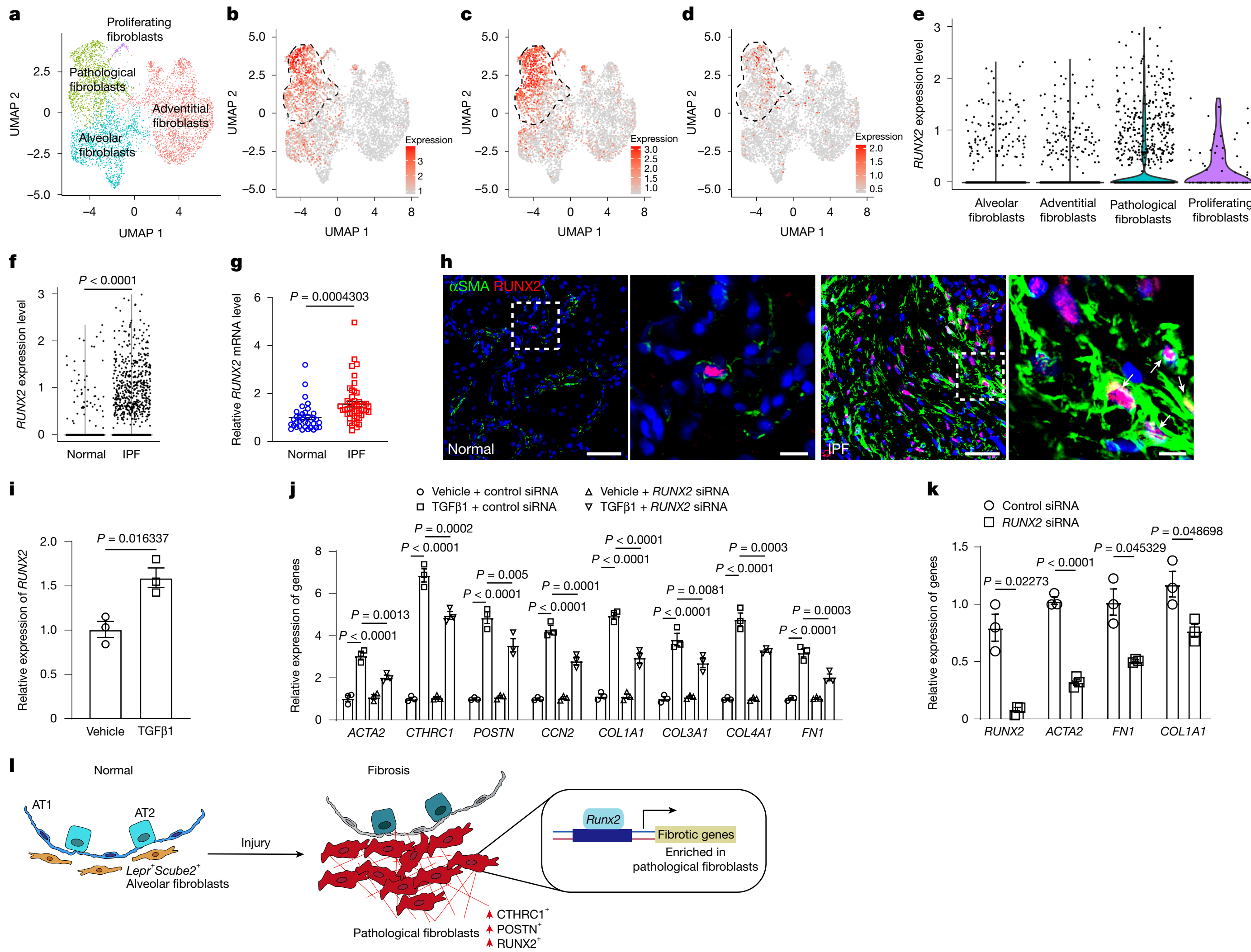


Fig. 5 | RUNX2 mediates ECM production in human IPF fibroblasts. a, UMAP plot showing four different lung fibroblast populations in human normal and IPF lungs. **b**,**c**, UMAP plot showing expression of *POSTN* (**b**) and *CTHRC1* (**c**) in the pathological fibroblast population. **d**, UMAP plot showing expression of *RUNX2*. Pathological fibroblast population is indicated by dashed outline. **e**, Enriched expression of *RUNX2* in pathological fibroblasts. **f**, Violin plot of scRNA-seq data showing that *RUNX2* transcripts are significantly increased in IPF lungs. **g**, The increased transcript levels of *RUNX2* in IPF lungs revealed by re-analysis of a bulk RNA-seq dataset[42] (Gene Expression Omnibus (GEO) accession number GSE124685) (normal, $n = 35$; IPF, $n = 49$). **h**, Representative immunostaining of αSMA and RUNX2 in normal human lungs ($n = 5$) and IPF lungs ($n = 8$). The arrows indicate αSMA$^+$RUNX2$^+$ cells. Images on the right are magnified views of the dashed square in the left-hand images. Scale bars, 50 μm (10 μm in magnified views). **i**, RT–qPCR analysis of *RUNX2* in freshly isolated primary normal human lung fibroblasts treated with TGFβ1 or vehicle (vehicle, $n = 3$; TGFβ1, $n = 3$). **j**, siRNA-mediated knockdown of *RUNX2* affects the expression of fibrosis-associated genes in freshly isolated primary normal human lung fibroblasts treated with TGFβ1 or vehicle. The transcripts of *ACTA2*, *CTHRC1*, *POSTN*, *CCN2*, *COL1A1*, *COL3A1*, *COL4A1* and *FN1* were determined by RT–qPCR ($n = 3$ for each group). **k**, RT–qPCR analysis of human IPF lung fibroblasts transfected with *RUNX2* siRNA or control siRNA (control siRNA, $n = 3$; *RUNX2* siRNA, $n = 3$). **l**, Schematic showing that LEPR$^+$SCUBE2$^+$ alveolar fibroblasts contribute to CTHRC1$^+$POSTN$^+$ pathological fibroblasts, and that increased *Runx2* activation regulates the expression of fibrotic genes during disease progression. Data are representative of at least three independent experiments. Data are mean ± s.e.m. Statistical analysis was performed using two-sided Wilcoxon rank-sum test (**f**), unpaired two-tailed *t*-test with Welch's correction (**g**,**i**,**k**) or two-way analysis of variance with multiple comparisons test with Sidak's correction (**j**). AT1, alveolar type 1 cell.

were also observed in IPF lungs compared with normal lungs, in which RUNX2$^+$ cells were rarely present (36.38 ± 6.57 per mm^2 (IPF) compared with 2.20 ± 1.07 per mm^2 (normal), $P = 0.00115$) (Fig. 5h). Consistently, the expression level of *RUNX2* was significantly increased in primary human normal lung fibroblasts after TGFβ1 treatment (Fig. 5i), and knockdown of *RUNX2* with short interfering RNA (siRNA) suppressed TGFβ1-induced pathological fibroblast differentiation and ECM production (Fig. 5j). We isolated fibroblasts from IPF lung samples and tested whether knockdown of *RUNX2* affected the expression of fibrotic genes. *RUNX2* knockdown led to a substantial reduction in the transcript levels of *ACTA2*, *FN1* and *COL1A1* (Fig. 5k), which indicated that RUNX2 also has a functional role in human pulmonary fibrosis.

Discussion

The obscure cell of origin for pulmonary fibrosis contributes to the uncertainty of the factors that drive initial fibrotic changes. Our comprehensive scRNA-seq analyses uncovered *Lepr* as a marker that labels most of the lung fibroblasts that persist in the adult lung. Cell-fate mapping confirmed that after injury, LEPR$^+$ cell derivatives, the majority of which are alveolar fibroblasts, differentiated into CTHRC1$^+$POSTN$^+$

pathological fibroblasts. Genetic ablation of pathological fibroblasts significantly attenuated pulmonary fibrosis. Further analyses revealed that RUNX2 is a key driver for pulmonary fibrosis, and inactivation of *Runx2* blocked the transition of alveolar fibroblasts to pathological fibroblasts (Fig. 5l).

Our study demonstrated that both $Lepr^{cre}$ and $Lepr^{creERT2}$ alleles labelled lung fibroblasts starting at the neonatal stage. These labelled fibroblasts became a key source of pathological fibroblasts during pulmonary fibrosis in two different injury models. In line with a previous study of myelofibrosis[16], $LEPR^{+}$ stromal cells seem to be a key contributor to fibrosis in both lung and bone marrow. Our scRNA-seq analyses revealed that *Cthrc1* and *Postn* are expressed in the pathological fibroblasts derived from $Lepr^{cre}$-labelled cells and $Lepr^{creERT2}$ labelled cells, which suggests that these $LEPR^{+}$ fibroblasts are significant contributors of previously identified pathological fibroblasts[9]. Consistently, we observed the expansion of $Lepr^{cre}$-labelled cells and $Lepr^{creERT2}$-labelled cells in bleomycin-treated lungs and silica-treated lungs, and genetic ablation of $Postn^{creER}$-labelled pathological fibroblasts attenuated fibrosis. Previous studies showed that fibrosis is attenuated in the lungs of mutant animals lacking *Postn*[44,45], which suggests that POSTN has a functional role during fibrotic changes. Of note, a recent study[26] showed that $Scube2^{creERT2}$-labelled cells differentiate into $POSTN^{+}CTHRC1^{+}$ pathological fibroblasts, similar to the $Lepr^{creERT2}$-labelled cells described here. *Lepr* is broadly expressed in several mesenchymal cell types, and our $Lepr^{cre}$ and $Lepr^{creER}$ alleles label a broad collection of cells, with alveolar fibroblasts constituting a predominant proportion. In the future, it will be interesting to explore whether $LEPR^{-}$ mesenchymal cells also meaningfully contribute to pulmonary fibrosis.

Our scATAC-seq and computational analyses identified RUNX2 as a pivotal transcription regulator of fibrotic drivers. In support of this finding, deletion of *Runx2* with $Lepr^{creERT2}$ or $Scube2^{creERT2}$ reduced the generation of pathological fibroblasts. RUNX2 mediates TGFβ signalling function during bone development[46,47], and overexpression of *Runx2* leads to fibrotic changes in vascular smooth muscle cells with increased expression of *Col1a1* and *Col1a2* (ref. 48). We found that RUNX2 is also important for the expression of fibrotic genes in fibrotic fibroblasts isolated from human IPF lung samples. Notably, a previous study[49] showed that RUNX2 is increased in AT2 cells but reduced in fibroblasts of IPF samples, as measured by microarray and immunofluorescence staining. That study also demonstrated that siRNA-mediated knockdown of *RUNX2* promoted the expression of *COL1A1* and *ACTA2* in TGFβ1-stimulated fibroblasts[49]. By contrast, our scRNA-seq findings indicated that *Runx2* is activated during the conversion of alveolar fibroblasts into pathological fibroblasts in both mouse models and IPF samples. Genetic deletion of *Runx2* blocked the conversion of alveolar fibroblasts to pathological fibroblasts and reduced lung fibrosis. Previously, TBX4 and PU.1 have been shown to regulate the differentiation of fibroblasts and contribute to pulmonary fibrosis[6,50]. Our scRNA-seq analyses showed that the level of *Tbx4* transcripts was similar between alveolar fibroblasts and pathological fibroblasts. Moreover, pathological fibroblasts ($CTHRC1^{+}POSTN^{+}$) exhibited minimal expression of PU.1 (data not shown), which suggests that PU.1 acts in different cell populations during pulmonary fibrosis. Further elucidation of the relationship and function of each player during fibrosis should offer new opportunities to intervene in the aggressive progression towards end-stage disease.

In summary, we demonstrated that $LEPR^{+}$ cells that arise at the early postnatal stage include alveolar fibroblasts as major derivatives. Cell-fate mapping combined with scRNA-seq analyses confirmed that $LEPR^{+}$ fibroblasts become a major contributor to pathological fibroblasts during fibrosis. Deletion of *Runx2* blocked the differentiation of alveolar fibroblasts into pathological fibroblasts. Our results provide a potential therapeutic target to treat pulmonary fibrosis.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08542-2.

1. Noble, P. W., Barkauskas, C. E. & Jiang, D. Pulmonary fibrosis: patterns and perpetrators. *J. Clin. Invest.* **122**, 2756–2762 (2012).
2. Lagares, D. et al. ADAM10-mediated ephrin-B2 shedding promotes myofibroblast activation and organ fibrosis. *Nat. Med.* **23**, 1405–1415 (2017).
3. Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
4. Mayr, C. H. et al. Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers. *EMBO Mol. Med.* **13**, e12871 (2021).
5. Kramann, R. et al. Perivascular $Gli1^{+}$ progenitors are key contributors to injury-induced organ fibrosis. *Cell Stem Cell* **16**, 51–66 (2015).
6. Xie, T. et al. Transcription factor TBX4 regulates myofibroblast accumulation and lung fibrosis. *J. Clin. Invest.* **126**, 3063–3079 (2016).
7. El Agha, E. et al. Two-way conversion between lipogenic and myogenic fibroblastic phenotypes marks the progression and resolution of lung fibrosis. *Cell Stem Cell* **20**, 261–273 (2017).
8. Zepp, J. A. et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* **170**, 1134–1148 (2017).
9. Tsukui, T. et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* **11**, 1920 (2020).
10. Lee, J. H. et al. Anatomically and functionally distinct lung mesenchymal populations marked by Lgr5 and Lgr6. *Cell* **170**, 1149–1163 (2017).
11. Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118–1123 (2018).
12. Abdollahi, A. et al. Inhibition of platelet-derived growth factor signaling attenuates pulmonary fibrosis. *J. Exp. Med.* **201**, 925–935 (2005).
13. Frangogiannis, N. Transforming growth factor-β in tissue fibrosis. *J. Exp. Med.* **217**, e20190103 (2020).
14. Boyd, D. F. et al. Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* **587**, 466–471 (2020).
15. Zhou, B. O., Yue, R., Murphy, M. M., Peyer, J. G. & Morrison, S. J. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. *Cell Stem Cell* **15**, 154–168 (2014).
16. Decker, M. et al. Leptin-receptor-expressing bone marrow stromal cells are myofibroblasts in primary myelofibrosis. *Nat. Cell Biol.* **19**, 677–688 (2017).
17. Negretti, N. M. et al. A single-cell atlas of mouse lung development. *Development* **148**, dev199512 (2021).
18. Curras-Alonso, S. et al. An interactive murine single-cell atlas of the lung responses to radiation injury. *Nat. Commun.* **14**, 2445 (2023).
19. Saito, S. et al. Angiopoietin-like 4 is a critical regulator of fibroblasts during pulmonary fibrosis development. *Am. J. Respir. Cell Mol. Biol.* **69**, 328–339 (2023).
20. Knudsen, L. et al. Alveolar micromechanics in bleomycin-induced lung injury. *Am. J. Respir. Cell Mol. Biol.* **59**, 757–769 (2018).
21. Niethamer, T. K. et al. Defining the role of pulmonary endothelial cell heterogeneity in the response to acute lung injury. *eLife* **9**, e53072 (2020).
22. Upagupta, C., Shimbori, C., Alsilmi, R. & Kolb, M. Matrix abnormalities in pulmonary fibrosis. *Eur. Respir. Rev.* **27**, 180033 (2018).
23. Melms, J. C. et al. A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).
24. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
25. Dahlgren, M. W. et al. Adventitial stromal cells define group 2 innate lymphoid cell tissue niches. *Immunity* **50**, 707–722 (2019).
26. Tsukui, T., Wolters, P. J. & Sheppard, D. Alveolar fibroblast lineage orchestrates lung inflammation and fibrosis. *Nature* **631**, 627–634 (2024).
27. Rock, J. R. et al. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc. Natl Acad. Sci. USA* **108**, E1475–E1483 (2011).
28. Hung, C. et al. Role of lung pericytes and resident fibroblasts in the pathogenesis of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **188**, 820–830 (2013).
29. Baek, S. H. et al. Single cell transcriptomic analysis reveals organ specific pericyte markers and identities. *Front. Cardiovasc. Med.* **9**, 876591 (2022).
30. Kanisicak, O. et al. Genetic lineage tracing defines myofibroblast origin and function in the injured heart. *Nat. Commun.* **7**, 12260 (2016).
31. Ogawa, T., Shichino, S., Ueha, S., Bando, K. & Matsushima, K. Profibrotic properties of $C1q^{+}$ interstitial macrophages in silica-induced pulmonary fibrosis in mice. *Biochem. Biophys. Res. Commun.* **599**, 113–119 (2022).
32. Nguyen, H. C. B., Adlanmerini, M., Hauck, A. K. & Lazar, M. A. Dichotomous engagement of HDAC3 activity governs inflammatory responses. *Nature* **584**, 286–290 (2020).
33. Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
34. Vierstra, J. et al. Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).
35. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

36. Geng, Y. et al. PEAR1 regulates expansion of activated fibroblasts and deposition of extracellular matrix in pulmonary fibrosis. *Nat. Commun.* **13**, 7114 (2022).
37. Parimon, T. et al. Syndecan-1 promotes lung fibrosis by regulating epithelial reprogramming through extracellular vesicles. *JCI Insight* **5**, e129359 (2019).
38. Yang, W. et al. Single-cell transcriptomic analysis reveals a hepatic stellate cell-activation roadmap and myofibroblast origin during liver fibrosis in mice. *Hepatology* **74**, 2774–2790 (2021).
39. Misharin, A. V. et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J. Exp. Med.* **214**, 2387–2404 (2017).
40. Leslie, J. et al. FPR-1 is an important regulator of neutrophil recruitment and a tissue-specific driver of pulmonary fibrosis. *JCI Insight* **5**, e125937 (2020).
41. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, eaba1983 (2020).
42. McDonough, J. E. et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* **4**, e131597 (2019).
43. Sivakumar, P. et al. RNA sequencing of transplant-stage idiopathic pulmonary fibrosis lung reveals unique pathway regulation. *ERJ Open Res.* **5**, 00117–02019 (2019).
44. Naik, P. K. et al. Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **303**, L1046–L1056 (2012).
45. Uchida, M. et al. Periostin, a matricellular protein, plays a role in the induction of chemokines in pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **46**, 677–686 (2012).
46. Seo, H. S. & Serra, R. Tgfbr2 is required for development of the skull vault. *Dev. Biol.* **334**, 481–490 (2009).
47. Wu, M., Chen, G. & Li, Y. P. TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* **4**, 16009 (2016).
48. Raaz, U. et al. Transcription factor Runx2 promotes aortic fibrosis and stiffness in type 2 diabetes mellitus. *Circ. Res.* **117**, 513–524 (2015).
49. Mummler, C. et al. Cell-specific expression of runt-related transcription factor 2 contributes to pulmonary fibrosis. *FASEB J.* **32**, 703–716 (2018).
50. Wohlfahrt, T. et al. PU.1 controls fibroblast polarization and tissue fibrosis. *Nature* **566**, 344–349 (2019).

Methods

Human lung tissue collection

Lung tissue samples from healthy donors and from patients with end-stage IPF undergoing transplantation were obtained from the Department of Pathology at Columbia University Medical Center under a protocol approved by the Columbia University Institutional Review Board (AAAS4094) and from Cedars Sinai Medical Center (Pro00032727). Informed consent was obtained from each donor or authorized representatives. All experiments using human lung tissue were performed in accordance with the approved protocol mentioned above.

Mice

Lepr^{cre} (ref. 51), *Col1a1*^{EGFP} (ref. 52) and *Runx2*$^{flox/flox}$ (ref. 53) mouse strains have been previously described. The *Scube2*^{creERT2} mouse strain was shared by D. Sheppard. *Lepr*^{creERT2} and *Higd1b*^{creERT2} mouse strains were generated in the Que Laboratory (see below for details). *Postn-MerCreMer* (029645), *Rosa26-tdtomato* (007914), *Rosa26-DTA* (009669) and *ACTB-Flpe* (005703) mice were purchased from The Jackson Laboratory. All mice used in the experiments were between the ages of 8 and 12 weeks and maintained on a C57BL/6 background (both male and female). Mice were housed with a 12-h light–dark cycle at 18–23 °C and 40–60% humidity in the animal facility at Columbia University Medical Center. To induce Cre recombinase activity in the *Higd1b*^{creERT2} mouse strain, mice were intraperitoneally injected with 200 mg kg^{-1} body weight tamoxifen (Sigma, T5648) dissolved in sunflower oil. *Lepr*^{creERT2} and *Scube2*^{creERT2} mice were fed with tamoxifen-containing chow (0.5 g kg^{-1}, Inotiv, TD.130857) for 2 weeks to induce Cre recombinase activity followed by at least a 2-week washout time before exposure to bleomycin or silica. Sample sizes were empirically determined based on experience or previous published relevant studies. No blinding method was applied. Sex-matched and age-matched mice were randomly assigned to experiments. All mouse experiments and care were conducted in accordance with the procedures approved by the Institutional Animal Care and Use Committee at Columbia University (AABM6565).

Generation of *Lepr*$^{IRES\text{-}creERT2\text{-}P2A\text{-}EGFP}$ and *Higd1b*$^{creERT2\text{-}P2A\text{-}EGFP}$ mouse strains

To generate the *Lepr*$^{IRES\text{-}creERT2\text{-}P2A\text{-}EGFP}$ mouse strain, a targeting construct containing an internal ribosome entry site (IRES), a cDNA encoding tamoxifen-inducible Cre recombinase (CreERT2), a P2A self-cleaving peptide sequence and EGFP followed by a FRT-flanked neomycin-resistance cassette was generated and electroporated into KV1 (129-C57BL/6 hybrid) embryonic stem cells. The targeting vector was inserted immediately before the 5′ end of the stop codon in the last exon of the *Lepr* gene by homologous recombination. Following G418 selection, the targeted embryonic stem cell clones were validated by PCR analysis and injected into C57BL/6N blastocysts to generate chimeras. The chimeras were crossed with *ACTB-Flpe* mice to identify germline transmission of the targeted allele and to remove the neomycin cassette. A similar strategy was used to generate the *Higd1b*$^{creERT2\text{-}P2A\text{-}EGFP}$ knock-in mouse strain, but the targeting vector was inserted to replace the start codon in exon 2 of the *Higd1b* locus. The targeting construct for *Higd1b*$^{creERT2\text{-}P2A\text{-}EGFP}$ contains a cDNA encoding tamoxifen-inducible Cre recombinase (CreERT2), a P2A self-cleaving peptide sequence and EGFP followed by a FRT-flanked neomycin-resistance cassette.

Lung injury mouse models

C57BL/6 mice (8 weeks old) mice were used for lung injury models. Intratracheal administration of bleomycin or silica was performed as previously described[54,55]. In brief, mice were anaesthetized and 1.75 unit kg^{-1} bleomycin (Fresenius Kabi, USP) or 200 mg kg^{-1} silica suspension in PBS was delivered through intratracheal injection with a 30-gauge needle. Lung tissue samples were collected at the indicated time points.

Tissue preparation and histology

Mice were euthanized with isoflurane and lungs were inflated and fixed with 4% paraformaldehyde overnight. Lung tissue samples were dehydrated and processed as previously described[23,54]. Sections (7 μm) were cut and collected for further histology staining and immunostaining. H&E staining was performed as previously described[23,54]. Picro-Sirius Red staining was performed according to the instructions of a commercial kit (VitroVivo Biotech, VB-3017). A Leica Aperio AT2 microscope slide scanner was used to obtain whole section images.

Hydroxyproline assay

The hydroxyproline content in lung tissue was measured using a commercial hydroxyproline assay kit (Cell Biolabs, STA-675) as previously described[56]. In brief, the lung tissue samples were homogenized in distilled water, and the samples were mixed with 12 N hydrochloric acid and incubated for 24 h at 95 °C to hydrolyse the homogenized tissue. Following clarification, the hydrolysed samples were filtered through a 0.45 μm syringe filter into tubes and dried in an oven to remove the residual hydrochloric acid. After incubation with chloramine T mixture for 30 min at room temperature, the samples were incubated with Ehrlich's reagent for 45 min at 60 °C. Following incubation at 4 °C for 5 min, the samples were centrifuged at 6,000*g* for 15 min at room temperature. The supernatants were transferred to microplate wells and the absorbance was measured on a microplate reader using 540–560 nm as the primary wavelength.

Immunofluorescence staining

Immunostaining was performed as previously described[23,54,57]. In brief, paraffin sections were dewaxed and rehydrated through gradient ethanol. Antigen retrieval was performed with high-pressure heating in a commercial antigen unmasking solution (Vector Laboratory, H-3300) for 2 min. The sections were washed in PBS, permeabilized and blocked with blocking buffer (0.2% Triton X-100 and 5% normal donkey serum in PBS) for 1 h at room temperature. The sections were incubated with the following primary antibodies: anti-αSMA (Santa Cruz, sc-32251, 1:200); anti-tdTomato (Biorbyt, orb182397, 1:1,000); anti-RFP (Rockland, 600-401-379, 1:500); anti-collagen I (SouthernBiotech, 1310-01, 1:200); anti-PI16 (R&D systems, AF4929, 5 μg ml^{-1}); anti-CTHRC1 (MaineHealth Institute for Research, Vli55, 1:250); anti-POSTN (Abcam, ab215199, 1:200); anti-ERG (Abcam, ab92513, 1:200); anti-endomucin (Santa Cruz, sc-65495, 1:200); anti-RUNX2 (Cell Signaling Technology, 12556S, 1:200); anti-Ki67 (Cell Signaling Technology, 9129S, 1:200); anti-NG2 (Millipore Sigma, AB5320, 1:200); and anti-ProSPC (Abcam, ab211326, 1:500). Antibodies were diluted in blocking buffer and incubated at 4 °C overnight. Following extensive washing with PBS three times, the sections were incubated with fluorophore-conjugated secondary antibodies for 2 h at room temperature. DAPI was used to counterstain the nuclei. After washing with PBS, the sections were mounted using Fluoromount-G (SouthernBiotech, 0100-20). A Zeiss LSM T-PMT confocal laser-scanning microscope was used for obtaining the images.

RNA in situ hybridization

RNA in situ hybridization was performed using a RNAscope multiplex fluorescent detection kit v.2 (Advanced Cell Diagnostics, 323100) as previously described[58]. In brief, pre-baked paraffin sections were dewaxed and rehydrated, then treated with hydrogen peroxide solution for 10 min at room temperature. After target retrieval for 15 min in 95–105 °C solution, the sections were incubated with protease for 30 min at 40 °C. *Mm-Scube2* probes (Advanced Cell Diagnostics, 488141) were hybridized for 2 h at 40 °C, followed by signal amplification steps. Fluorophore (Akoya Biosciences, FP1487001KT) was incubated for 30 min at 40 °C. The sections were incubated with DAPI for counterstaining and then mounted using Fluoromount-G solution. A Zeiss LSM T-PMT confocal laser-scanning microscope was used for obtaining the images.

Single-cell isolation and flow cytometry analysis

To obtain single-cell suspensions, lung tissue samples were dissected, washed with PBS and minced with a razor blade followed by digestion in a digestion buffer (2 mg ml^{-1} collagenase IV, 2 mg ml^{-1} dispase II and 10 U ml^{-1} DNase I) for 30 min at 37 °C. DMEM containing 10% FBS was added to stop digestion, and the cells were filtered through 100 µm and 40 µm strainers. After centrifugation at 300*g* for 5 min, the cell pellet was resuspended in red blood cell lysis buffer (Sigma), incubated for 2 min at 37 °C, followed by washing with HBSS containing 10% FBS and centrifuged at 300*g* for 5 min. Flow cytometry analysis was performed as previously described[23,54]. In brief, cells were incubated with PE-Cy7-CD45 (BioLegend, 103114, 1:100), APC-EpCAM (BioLegend, 118214, 1:100) and BV711-CD31 (BioLegend, 102449, 1:100) antibodies in FACS buffer (5% FBS with 0.5 mM EDTA in PBS) for 1 h at 4 °C, and then incubated with Live/Dead stain dye for 10 min at room temperature to exclude dead cells. After washing with FACS buffer, the cells were fixed and permeabilized using Fixation/Permeabilization buffer and then incubated with Alexa Fluor 488-αSMA antibody (eBioscience, 53-9760-82, 1:100) at 4 °C for 1 h. BD LSRII and FlowJo (v.10) software were used for obtaining data and analyses, respectively. Live CD45$^-$EpCAM$^-$CD31$^-$ mesenchymal cells were gated for further analysis. For scRNA-seq, single-cell suspensions were applied to sort live tdT$^+$ cells from *Lepr^{cre};R26^{tdT}* mice or live tdT$^+$CD45$^-$EpCAM$^-$CD31$^-$ mesenchymal cells from *Lepr^{creERT2};R26^{tdT}* mice using a BD Influx cell sorter. For the analysis of bone marrow monocytes and neutrophils, the single-cell suspensions were obtained from femur bone marrow as previously described[59,60]. Bone marrow cells were incubated with eFluor 450-CD45 (eBioscience, 48-0451, 1:100), Alexa Fluor 700-Ly6G (BioLegend, 127622, 1:100), Alexa Fluor 488-Ly6C (BioLegend, 128022, 1:100) and PE-Cy7-CD11b (eBioscience, 25-0112, 1:100) antibodies in FACS buffer for 1 h at 4 °C. The flow cytometry data were collected on a BD LSRII and analysed using FlowJo (v.10) software. CD45$^+$Ly6G$^-$CD11b$^+$Ly6C$^+$ monocytes and CD45$^+$Ly6G$^+$CD11b$^+$Ly6C$^+$ neutrophils were gated for analysis.

scRNA-seq analysis

Sorted tdT$^+$ cells were loaded onto a Chromium Controller instrument (10x Genomics) at the Single Cell Analysis Core of Genome Center at Columbia University. 10x Single Cell 3′ V2 and V3 chemistry kits were used to produce single-cell barcoded droplets and to prepare libraries. An Illumina NovaSeq 6000 instrument was used to sequence the resulting libraries and to obtain the fastq files. Reads were aligned to a custom reference containing the mouse genome GRCm38/mm10 with the tdTomato-WPRE-polyA sequence, and the unique molecular identifier (UMI) counts were obtained using Cell Ranger (v.3.1.0 and v.7.1.0) software. We re-analysed publicly available scRNA-seq data of developing mouse lung[17] (GEO accession numbers GSE160876 and GSE165063), normal adult mouse lungs[9,18,19] (GEO accession numbers GSE132771, GSE201698 and GSE211713), bleomycin-challenged mouse lungs[9,19,36,37] (GEO accession numbers GSE131800, GSE132771, GSE183545 and GSE201698), silica-exposed mouse lungs[31] (GEO accession number GSE184854), CCl_4-treated mouse liver[38] (GEO accession number GSE171904) and human patients with IPF[41] (GEO accession number GSE136831) and publicly available bulk RNA-seq data of human patients with IPF[42,43] (GEO accession numbers GSE124685 and GSE134692). The R package Seurat (v.4.4.0)[61,62] was used for further analysis by importing the raw count matrices. Low-quality cells were excluded as determined by the number of gene transcripts, UMI counts and the percentage of mitochondrial transcripts. UMI count normalization and variable feature identification were performed using the Seurat NormalizeData and FindVariableFeatures functions, respectively. Two sample objects were integrated by identifying integration anchors using Seurat FindIntegrationAnchors and IntegrateData. Seurat merge and RunFastMNN functions were used to integrate multiple Seurat objects with batch correction. ScaleData, principal component analysis and nonlinear dimensional reduction UMAP were applied to perform dimensional reduction. Cell clusters were identified using Seurat FindNeighbors and FindClusters functions. The Seurat FindAllMarkers function was used with the default parameters to identify the differentially expressed genes for each cluster. Cell clusters were manually annotated according to the marker genes in each cluster. Fibroblast clusters were selected for re-clustering, and four specific clusters (alveolar fibroblasts, adventitial fibroblasts, pathological fibroblasts and proliferating fibroblasts) were identified on the basis of the reported marker genes. To obtain the module scores for feature expression, the average expression levels of gene signatures on a single-cell level were calculated using the AddModuleScore function. The differentially expressed genes between 'alveolar fibroblast' and 'pathological fibroblast' were identified using the Seurat FindMarkers function. R packages Monocle2 (ref. 63) and Monocle3 (ref. 64) were used to perform pseudotime analysis, specifying 'alveolar fibroblast' as roots of the pseudotime. The Enrichr program[33,65,66] was used to predict transcription factors and signalling pathways by importing the enriched genes in 'pathological fibroblast'.

scATAC-seq analysis

For scATAC-seq, 8-week-old C57BL/6 wild-type mice were intratracheally administered with 1 unit kg^{-1} bleomycin or saline after anaesthesia with ketamine (100 mg kg^{-1}) and xylazine (10 mg kg^{-1}). At day 14, the lungs were perfused with 12 ml cold DPBS (Life Technology) and then inflated with dissociation buffer (PRMI1640 (Thermo Scientific) with 10% FBS, 1 mM HEPES (Life Technology), 1 mM $MgCl_2$ (Life Technology), 1 mM $CaCl_2$ (Sigma-Aldrich), 0.525 mg ml^{-1} collagenase D (Roche), 5 unit ml^{-1} dispase (Stemcell Technologies) and 0.05 mg ml^{-1} DNase I (Roche)). Minced lung was placed in dissociation buffer for 30 min at 37 °C, and cell solution was filtered through 70 µm and 40 µm strainers. Ammonium–chloride–potassium lysis was used to remove blood cells, and cell pellets were resuspended in FACS buffer. Antibodies used for flow cytometry included BV510-CD45 (BioLegend, 103138, 1:100), PE-Epcam (BioLegend, 118206, 1:100), PE-CD31 (BioLegend, 102408, 1:100) and APC-PDGFRα (BioLegend, 135908, 1:100). DAPI was used to label dead cells. A total of 18,000 cells from each group was subjected to scATAC-seq following the standard protocol by the Center for Epigenomics at the University of California, San Diego. Sample preparation and library construction were performed as previously described[67]. For data analysis, read alignment (mouse genome mm10) and cell barcode demultiplexing were conducted using 10x Genomics Cell Ranger ATAC (v.2.1.0) with default settings. Quality control, transcriptional start site enrichment, data normalization, dimensional reduction and UMAP-based cell clustering were performed using Signac[68]. Analysis of differentially accessible peaks was performed using FindAllMarkers. ChromVAR[35] was used to analyse differential transcription factor binding motif activities between groups of cells. The expression features of marker genes and the projection of motif enrichment scores were profiled and visualized using the R package ggplot2 (v.3.5.1).

Cell migration assay

Around 5×10^4 sorted tdT$^+$EGFP$^+$ and tdT$^-$EGFP$^+$ lung fibroblasts without mycoplasma contamination cultured in DMEM supplemented with 10% FBS were added into the top chamber of a BioCoat Control cell insert. To trigger cell migration, 10 ng ml^{-1} PDGF-BB was added to the bottom chamber. After incubation for 24 h at 37 °C with 5% CO_2, the cells were fixed with 4% paraformaldehyde for 30 min at room temperature. The cells on the upper side of the insert membrane were removed by scrubbing with a cotton swab. DAPI was used for counterstaining. Images were obtained using an EVOS M5000 system.

RNA extraction and RT–qPCR

Total RNA extraction was performed according to the instructions of a commercial RNA extraction kit (Qiagen, 74134). A SuperScript III First-Strand SuperMix kit (ThermoFisher, 18080400) was used to synthesize the first-strand cDNA from RNA. To quantify cDNA, iTaq Universal SYBR Green Supermix (Bio-Rad, 1725122) was used according to the manufacturer's instructions for a QuantStudio 5 Real-Time PCR system. At least three technical and biological replicates were performed. The primer sequences used in this study are listed in Supplementary Table 4.

siRNA-mediated knockdown of *RUNX2* in human normal lung fibroblasts treated with TGFβ1 and in human IPF fibroblasts

Human normal and IPF lung fibroblast isolation and culture were performed as previously described[69]. The isolated human lung fibroblasts were tested to be negative for mycoplasma contamination. Human normal lung fibroblasts were transfected with *RUNX2* siRNA (Santa Cruz, sc-37145) or a non-targeting control siRNA (Santa Cruz, sc-37007) using Lipofectamine RNAiMAX Reagent (ThermoFisher, 13778) according to the manufacturer's instructions while being treated with 5 ng ml^{-1} TGFβ1 for 48 h. In brief, Lipofectamine RNAiMAX Reagent and siRNA were diluted in Opti-MEM medium. Diluted siRNA was added to Lipofectamine RNAiMAX Reagent at a 1:1 ratio before applying to the cells. The cells were incubated at 37 °C with 5% CO_2. Human IPF lung fibroblasts were also transfected with *RUNX2* siRNA or a negative control siRNA. Cells were incubated in the mixture for 48 h and collected for further analysis. The oligonucleotide sequences of siRNA are listed in Supplementary Table 5.

Quantification and statistical analysis

Whole-slide digital images of lung lobes were used to quantify the fibrotic area of pulmonary fibrosis. For quantification of tdT^+ cells, $EGFP^+$ cells, $αSMA^+$ cells, $CTHRC1^+$ cells and $RUNX2^+$ cells, at least 10 random fields (×20 magnification or 1 mm^2) were captured, and ImageJ (v.1.51) software was used to count the positive cells or to measure the area of $αSMA^+$ cells as previously described[70]. Each group included at least three replicates for all experiments. All data are presented as the mean ± s.e.m. using GraphPad Prism 8. Unpaired two-tailed *t*-test with Welch's correction and two-sided Wilcoxon rank-sum test were used to determine statistical significance. For multiple comparisons, two-way analysis of variance was used with Sidak's correction. $P < 0.05$ or less was considered significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The scRNA-seq and scATAC-seq data generated in this study have been deposited into the GEO (accession numbers GSE229523, GSE276546 and GSE278419). The publicly available scRNA-seq data for mouse lung development (accession numbers GSE160876 and GSE165063), normal adult mouse lungs (accession numbers GSE132771, GSE201698 and GSE211713), bleomycin-challenged mouse lungs (accession numbers GSE131800, GSE132771, GSE183545 and GSE201698), silica-exposed mouse lungs (accession GSE184854), CCl_4-treated mouse liver (accession number GSE171904), human patients with IPF (accession number GSE136831) and bulk RNA-seq data for human patients with IPF (accession numbers GSE124685 and GSE134692) were used for analyses. Source data are provided with this paper.

Code availability

The codes used in this study for scRNA-seq and scATAC-seq analyses are available from the corresponding authors upon request. No custom code was generated.

51. DeFalco, J. et al. Virus-assisted mapping of neural inputs to a feeding center in the hypothalamus. *Science* **291**, 2608–2613 (2001).
52. Yata, Y. et al. DNase I-hypersensitive sites enhance α1(I) collagen gene expression in hepatic stellate cells. *Hepatology* **37**, 267–276 (2003).
53. Takarada, T. et al. An analysis of skeletal development in osteoblast-specific and chondrocyte-specific runt-related transcription factor-2 (Runx2) knockout mice. *J. Bone Miner. Res.* **28**, 2064–2069 (2013).
54. Fang, Y. et al. Epithelial Wntless regulates postnatal alveologenesis. *Development* **149**, dev199505 (2022).
55. Fang, Y. et al. Follistatin like-1 aggravates silica-induced mouse lung injury. *Sci. Rep.* **7**, 399 (2017).
56. Xu, C. et al. Hepatic neddylation deficiency triggers fatal liver injury via inducing NF-κB-inducing kinase in mice. *Nat. Commun.* **13**, 7782 (2022).
57. Jiang, M. et al. Transitional basal cells at the squamous–columnar junction generate Barrett's oesophagus. *Nature* **550**, 529–533 (2017).
58. Jiang, M. et al. VEGF receptor 2 (KDR) protects airways from mucus metaplasia through a Sox9-dependent pathway. *Dev. Cell* **56**, 1646–1660 (2021).
59. Janssen, H. et al. Monocytes re-enter the bone marrow during fasting and alter the host response to infection. *Immunity* **56**, 783–796 (2023).
60. Shao, Z. et al. DNA-PKcs has KU-dependent function in rRNA processing and haematopoiesis. *Nature* **579**, 291–296 (2020).
61. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
62. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
63. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
64. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
65. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
66. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
67. Wang, A. et al. Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of SARS-CoV2 host genes. *eLife* **9**, e62522 (2020).
68. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
69. Liu, X. et al. HER2 drives lung fibrosis by activating a metastatic cancer signature in invasive lung fibroblasts. *J. Exp. Med.* **219**, e20220126 (2022).
70. Konkimalla, A. et al. Transitional cell states sculpt tissue topology during lung regeneration. *Cell Stem Cell* **30**, 1486–1502 (2023).

Acknowledgements We thank the colleagues in the Que Laboratory for their critical input to the study. This work is partly supported by NIH grants R01HL152293, R01HL159675 and Department of Defense grant W81XWH2110196 (to J.Q.), NIH grants R01HL172990 and P01HL108793 (to D.J.), American Heart Association award 24CDA1268568 and the Pulmonary Fibrosis Foundation Scholars Program 1272558 (to X.L.). Flow cytometry was performed at the Columbia Center for Translational Immunology (CCTI) Flow Cytometry Core at Columbia University Medical Center, supported in part by the Office of the Director, National Institutes of Health under the awards S10RR027050 and S10OD020056. The CCHD microscopy core is supported by S10 OD032447 from the NIH. This research was also funded in part through the NIH/NCI Cancer Center Support Grant P30CA013696 and used the Genetically Modified Mouse Models/GMMMSR Core and the Genomics and High Throughput Screening Shared Resource.

Author contributions Y.F. and J.Q. designed experiments, analysed data and wrote the manuscript. Y.F., C.-S.L. and J.Q. generated the $Lepr^{creERT2}$ and $Higd1b^{creERT2}$ knock-in mouse lines. Y.F. performed immunostaining, imaging, flow cytometry and mouse genetics. S.S.W.C. performed immunostaining, imaging and mouse genetics. Y.F. and C.X. performed scRNA-seq analyses. L.X., R.L. and Y.K. performed scATAC-seq analyses. K.Y., Y.H., X.L., D.J., T.T. and D.S. provided materials. A.S. and H.H. oversaw and performed tissue collection. X.S. and J.Q. supervised this work. All authors reviewed and approved the final manuscript.

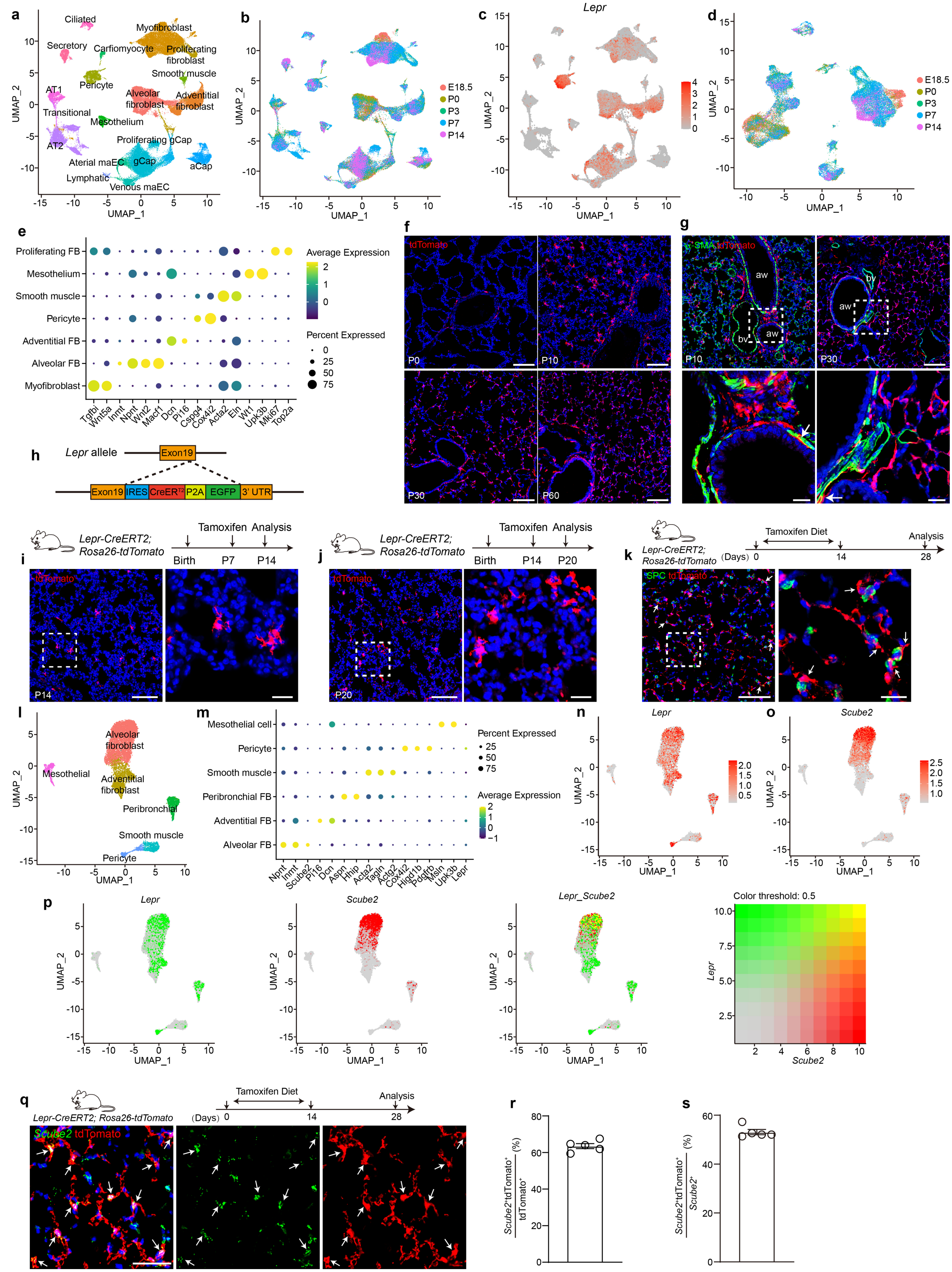
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08542-2.

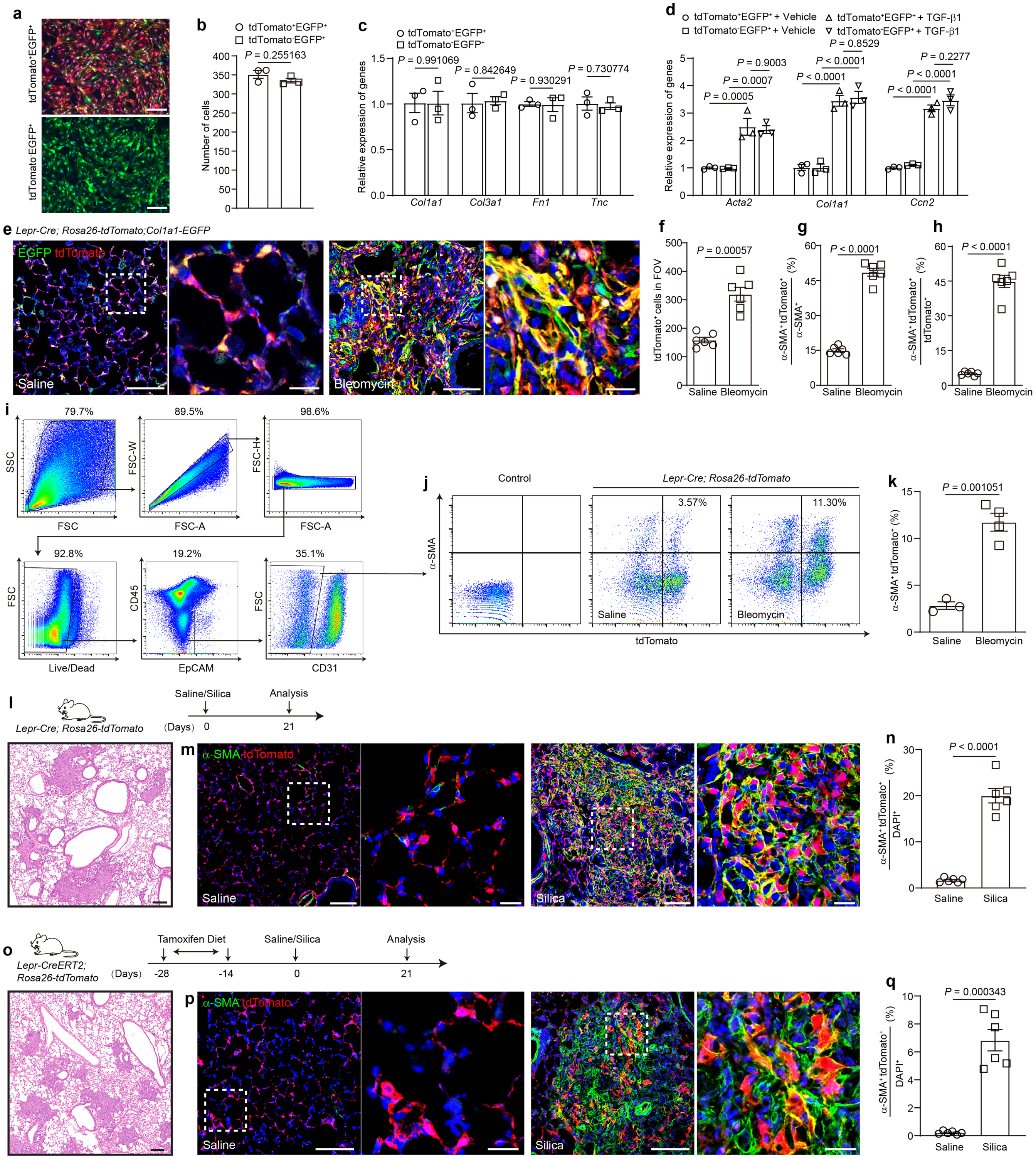
Correspondence and requests for materials should be addressed to Xin Sun or Jianwen Que.

Peer review information *Nature* thanks Christopher Buckley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



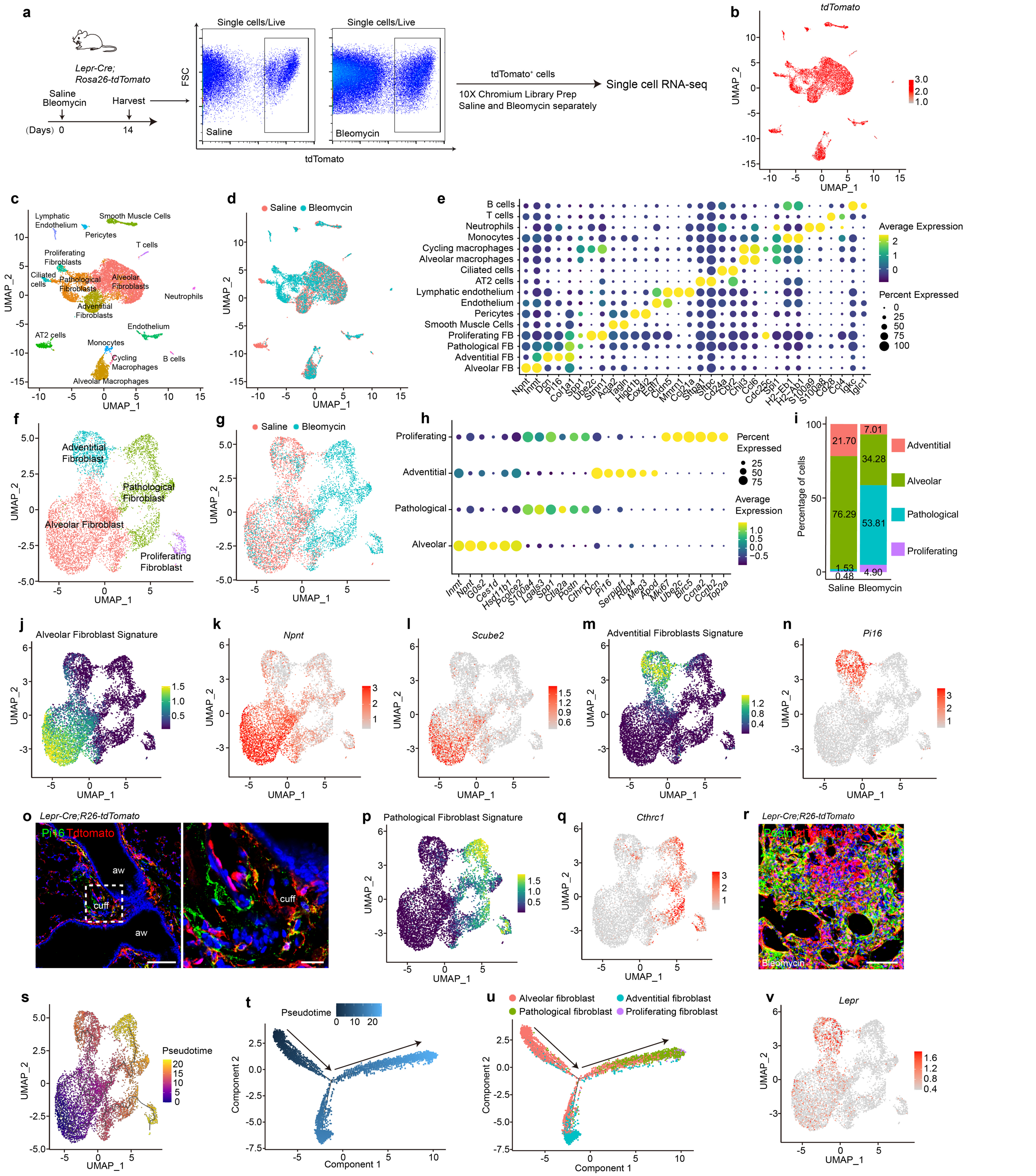
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | *Lepr*⁺ mesenchymal cells in the developing and adult mouse lung. **a**-**b**, UMAP plot showing the different cell populations (**a**) in the lungs collected at E18.5, P0, P3, P7 and P14 (**b**). Re-analysis of the datasets GSE160876 and GSE165063. **c**, UMAP plot showing expression of *Lepr* in the developing lungs. **d**, UMAP plot showing the integration of lung mesenchymal cells at E18.5, P0, P3, P7 and P14. **e**, Dot plot showing the representative markers for each mesenchymal cell population. **f**, tdTomato (tdT) expression in the lungs of *Lepr-Cre;R26tdT* mice at different postnatal stages. P: Postnatal. **g**, tdT and α-SMA expression in the lungs of *Lepr-Cre;R26tdT* mice at different postnatal stages. The arrows indicate rare tdT⁺ smooth muscle cells. aw: airway, bv: blood vessel. **h**, Schematic depicting the generation of *Lepr-CreERT2* mouse strain. **i**-**j**, Schematic diagram for tamoxifen injection and the representative immunostaining image of tdT in the lungs of *Lepr-CreERT2;R26tdT* mice at P14 (**i**) and P20 (**j**). **k**, Schematic depicting the use of tamoxifen-containing chow to feed *Lepr-CreERT2;R26tdT* mice and representative immunostaining images of surfactant protein C (SPC) and tdT in lung tissues. **l**, UMAP plot showing different mesenchymal populations from the adult control lungs that integrate the datasets GSE132771, GSE201698 and GSE211713. **m**, Dot plot showing the representative markers for each mesenchymal population. **n**-**o**, UMAP plot showing expression level of *Lepr* (**n**) and *Scube2* (**o**). **p**, Blended feature plots showing co-expression of *Lepr* and *Scube2*. **q**, Representative images of tdT immunostaining and *Scube2* RNA in situ hybridization. **r**-**s**, Quantification analysis showing the percentage of *Scube2*⁺ tdT⁺ cells in total tdT⁺ cells (**r**, n = 5) and in *Scube2*⁺ alveolar fibroblasts (**s**, n = 5). Data are representative of at least three independent experiments. Data are mean ± SEM. Scale bars: 100 µm (20 µm in magnified views).



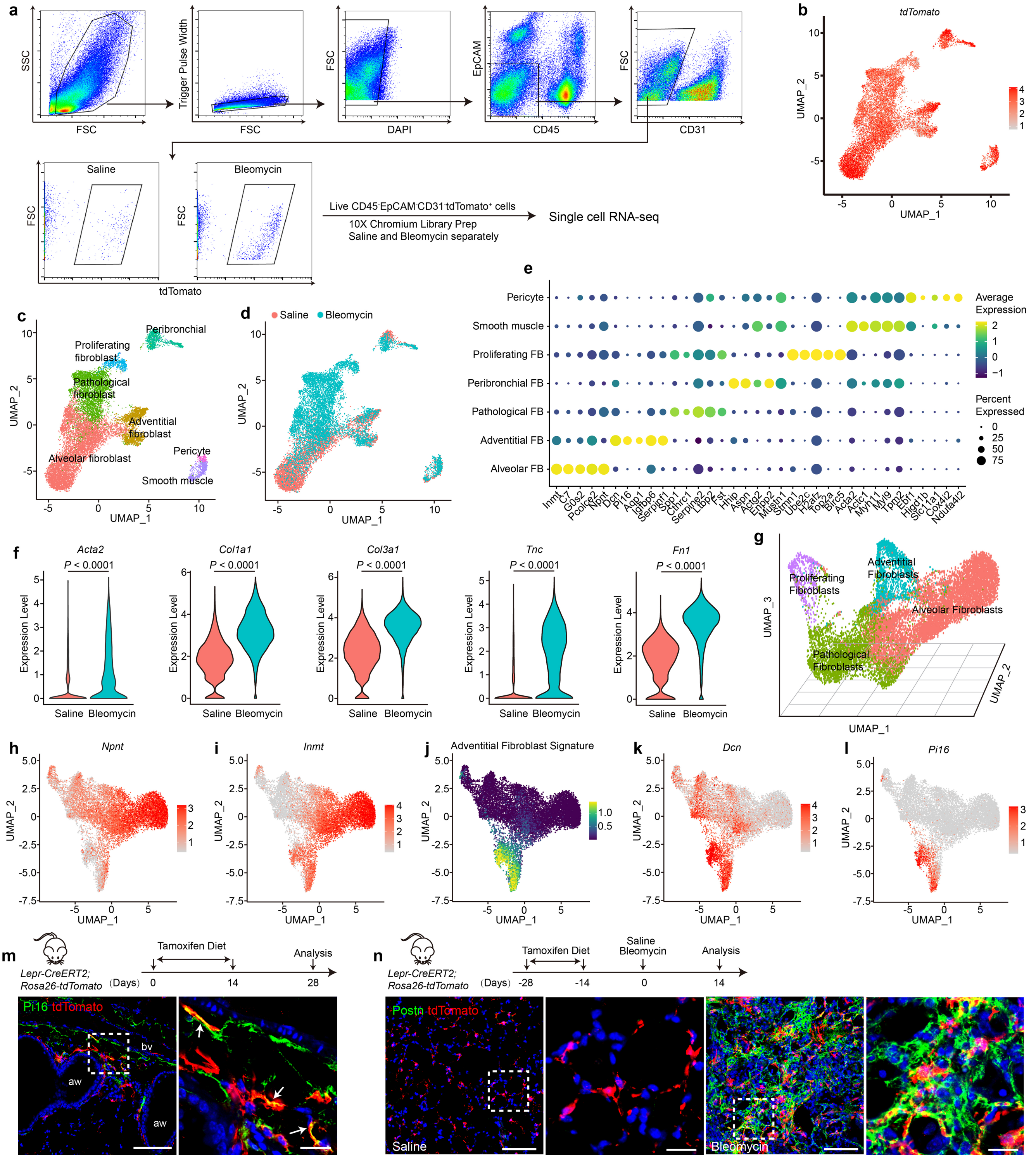
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | *Lepr-Cre* labeled cells contribute to pathological fibroblasts. a, Representative images of migrated tdT$^+$EGFP$^+$ and tdT$^-$EGFP$^+$ lung mesenchymal cells triggered by PDGF-BB. **b**, Quantification of the migrated cells (DAPI$^+$) in **a** (tdT$^+$EGFP$^+$: n = 3, tdT$^-$EGFP$^+$: n = 3). **c**, Relative expression of matrix genes in tdT$^+$EGFP$^+$ and tdT$^-$EGFP$^+$ lung mesenchyme (RT-qPCR) (tdT$^+$EGFP$^+$: n = 3, tdT$^-$EGFP$^+$: n = 3). **d**, Expression of *Acta2*, *Col1a1* and *Ccn2* in tdT$^+$EGFP$^+$ and tdT$^-$EGFP$^+$ lung mesenchymal cells treated with/without TGF-β1 (RT-qPCR) (n = 3 for each group). **e**, tdT and EGFP expression in the lungs of *Lepr-Cre;R26tdT;Col1a1-EGFP* mice challenged with saline or bleomycin. **f**, Quantification of tdT$^+$ cells (Saline: n = 6, Bleomycin: n = 6). **g**-**h**, Quantification of the proportion of α-SMA$^+$tdT$^+$ cells among α-SMA$^+$ cells (**g**) and among tdT$^+$ cells (**h**) (Saline: n = 6, Bleomycin: n = 6). **i**, Gating strategy to identify α-SMA$^+$ tdT$^+$ cells in live mesenchymal cells. **j**, FACS analysis of tdT$^+$ and α-SMA$^+$ cells isolated from the lung of *Lepr-Cre;R26tdT* mice treated with saline or bleomycin. **k**, Flow cytometric quantification for the percentage of α-SMA$^+$ tdT$^+$ cells in the lungs of *Lepr-Cre;R26tdT* mice treated with saline or bleomycin (Saline: n = 3, Bleomycin: n = 4). **l**, Experimental schematic and representative H&E images. **m**, α-SMA$^+$ tdT$^+$ cells in the lungs of *Lepr-Cre;R26tdT* mice exposed to saline or silica. **n**, Quantification of α-SMA$^+$ tdT$^+$ cells (Saline: n = 6, Silica: n = 6). **o**, Experimental schematic and representative H&E images. **p**, α-SMA and tdT expression in lung tissues. **q**, Quantification of α-SMA$^+$ tdT$^+$ cells in the lungs of *Lepr-CreERT2;R26tdT* mice following silica exposure (Saline: n = 6, Silica: n = 6). Data are representative of at least three independent experiments. Data are mean ± SEM. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction (**b**, **c, f**, **g**, **h**, **k, n**, **q**) and two-way analysis of variance with multiple comparisons test with Sidak's correction (**d**). Scale bars: **l**, **o**: 200 μm, **a**, **e**, **m**, **p**: 100 μm (**e**, **m**, **p**: 20 μm in magnified views).

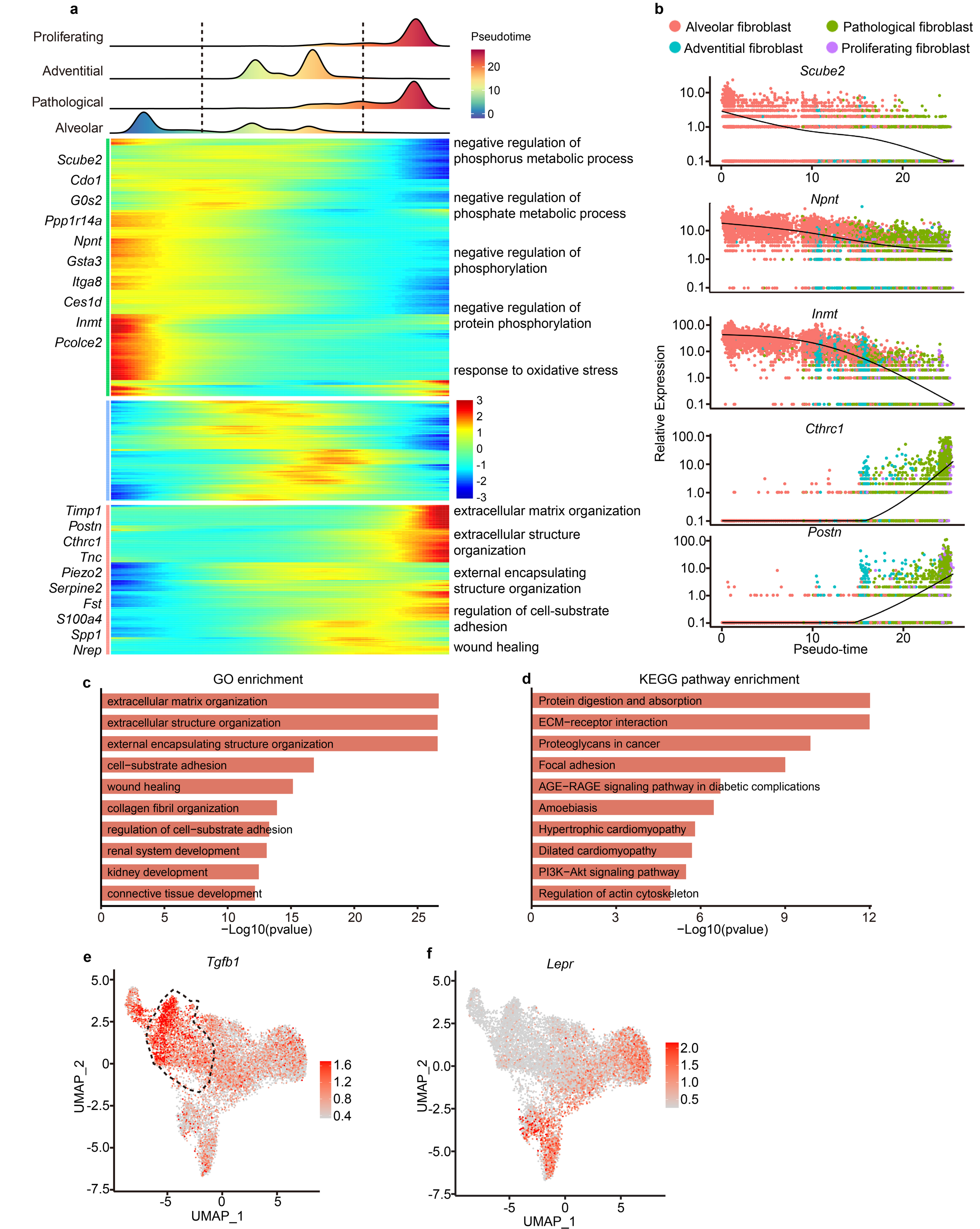


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | scRNA-seq reveals the contribution of *Lepr-Cre* labeled alveolar fibroblasts to pathological fibroblasts. a, Schematic diagram for scRNA-seq experimental design. **b**, UMAP plot showing the transcript levels of *tdTomato*. **c**, UMAP plot showing different cell populations. **d**, Red dots showing the cells from saline-treated lungs and blue dots showing the cells from bleomycin-treated lungs. **e**, Dot plot showing the representative markers for each population. **f**, UMAP plot showing four different fibroblast populations. **g**, Red and blue dots showing the cells from saline- and bleomycin-treated lungs, respectively. **h**, Dot plot showing the representative markers for each fibroblast population. **i**, Frequency of each fibroblast population in saline and bleomycin-treated lungs. **j**, UMAP plot showing the score of alveolar fibroblast signature. **k**-**l**, UMAP plot showing the expression level of the alveolar fibroblast markers *Npnt* (**k**) and *Scube2* (**l**). **m**, UMAP plot showing the adventitial fibroblast signature score. **n**, UMAP plot showing the expression level of the adventitial fibroblast marker *Pi16*. **o**, Representative immunostaining image of Pi16 and tdT in the adventitial cuffs. aw: airway. **p**, UMAP plot showing the pathological fibroblast signature score. **q**, UMAP plot showing the expression level of the pathological fibroblast marker *Cthrc1*. **r**, Representative immunostaining image of Postn and tdT in the lungs of *Lepr-Cre;R26tdT* mice treated with bleomycin. **s**, UMAP plot showing pseudotime analysis with Monocle3. **t**-**u**, Trajectory analysis with Monocle2 showing cells ordered by pseudotime (**t**) and cell type (**u**). **v**, UMAP plot showing the expression of *Lepr* in the mesenchymal cells of the lungs isolated from *Lepr-Cre;R26tdT* mice challenged with saline or bleomycin. Data are representative of at least three independent experiments. Scale bars: **o** and **r**: 100 µm (**o**: 20 µm in magnified views).

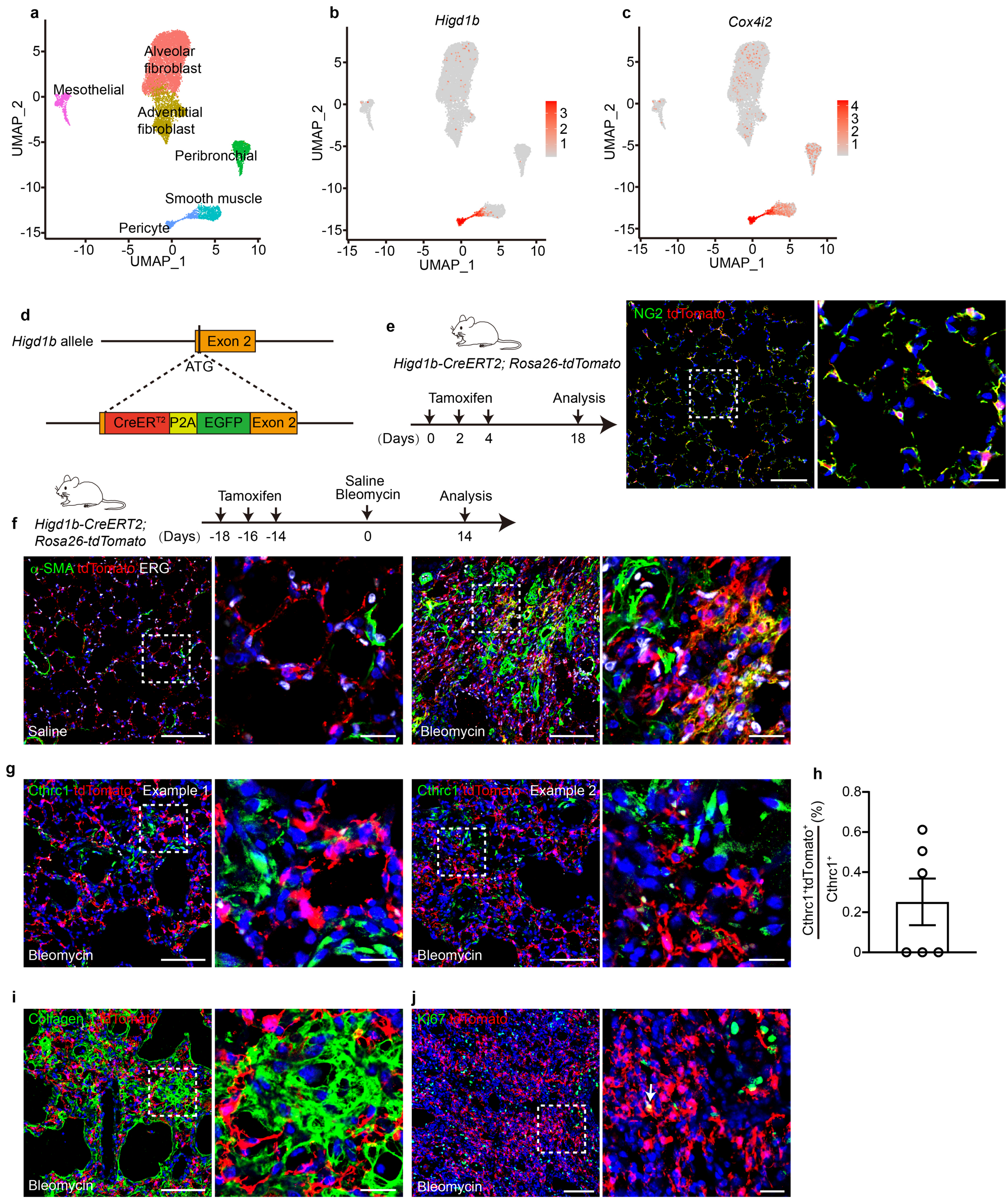


Extended Data Fig. 4 | scRNA-seq analysis reveals the heterogeneous lung fibroblasts labeled by *Lepr-CreERT2*. **a**, Gating strategy to sort tdT⁺ lung mesenchymal cells (CD45⁻EpCAM⁻CD31⁻) from the lungs of *Lepr-CreERT2;R26tdT* mice treated with bleomycin or saline for scRNA-seq. **b**, *tdTomato* transcript levels are shown by UMAP plot. **c**, UMAP plot showing multiple *Lepr-CreERT2* labeled cell populations. **d**, Red and blue dots showing the cells from saline- and bleomycin-treated lungs, respectively. **e**, Dot plot showing the representative markers for each population. **f**, Violin plot showing the significantly increased fibrosis-associated genes upon bleomycin challenge. **g**, 3D UMAP plot showing four different *Lepr-CreERT2* labeled fibroblast populations. **h**-**i**, UMAP plot showing the expression level of the alveolar fibroblast markers *Npnt* (**h**) and *Inmt* (**i**). **j**, UMAP plot showing the adventitial fibroblast signature score. **k**-**l**, UMAP plot showing the expression level of the adventitial fibroblast markers *Dcn* (**k**) and *Pi16* (**l**). **m**, Schematic depicting for the use of tamoxifen (Tmx)-containing chow to feed *Lepr-CreERT2;R26tdT* mice and representative immunostaining image of Pi16 and tdT in the adventitial cuffs. aw: airway. bv: blood vessel. **n**, Schematic depicting for the use of Tmx-containing chow and bleomycin challenge of *Lepr-CreERT2;R26tdT* mice, and representative images of Postn⁺ tdT⁺ pathological fibroblasts in lung tissues. Data are representative of at least three independent experiments. Statistical analysis was performed using two-sided Wilcoxon Rank Sum test (**f**). Scale bars: **m** and **n**: 100 μm (20 μm in magnified views).



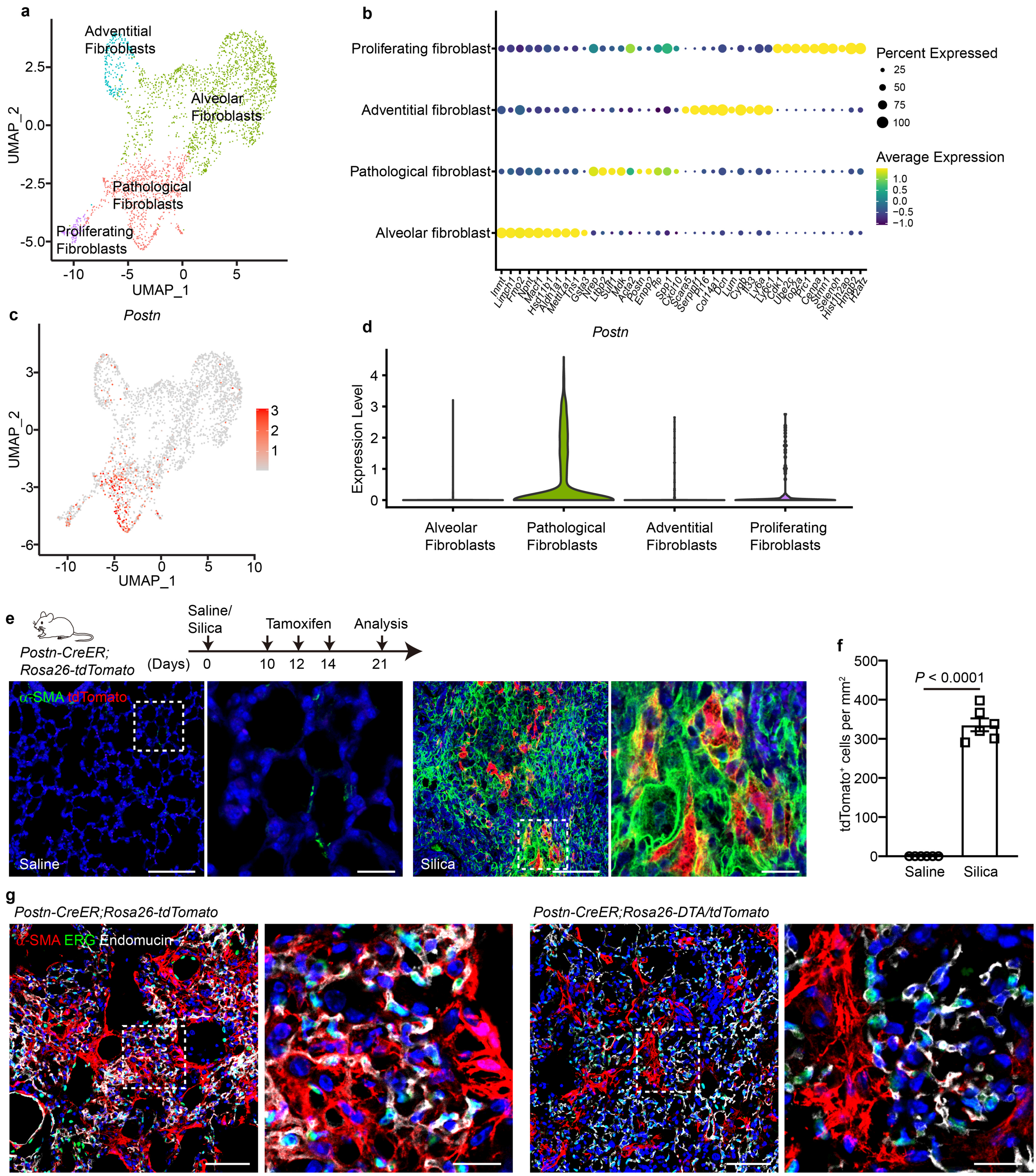
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | *Lepr-CreERT2* labeled alveolar fibroblasts generate pathological fibroblasts. a, The cell distribution of *Lepr-CreERT2* labeled fibroblast clusters along with the color-coded pseudotime (upper panel). Heatmap showing dynamic gene expression in fibroblast clusters (lower panel). Representative genes (left) and enriched pathways (right) in alveolar fibroblasts and pathological fibroblasts are shown, respectively. **b**, The expression levels of the alveolar fibroblast makers *Scube2*, *Npnt*, *Inmt* and the pathological fibroblast markers *Cthrc1*, *Postn* along pseudotime analysis. **c**-**d**, The top ten GO terms (**c**) and KEGG pathways (**d**) associated with genes enriched in pathological fibroblasts. **e**, UMAP plot showing the enrichment of *Tgfβ1* transcript in pathological fibroblasts. **f**, UMAP plot showing the expression of *Lepr* transcript in lung mesenchymal cells of *Lepr-CreERT2;R26tdT* mice challenged with saline or bleomycin.

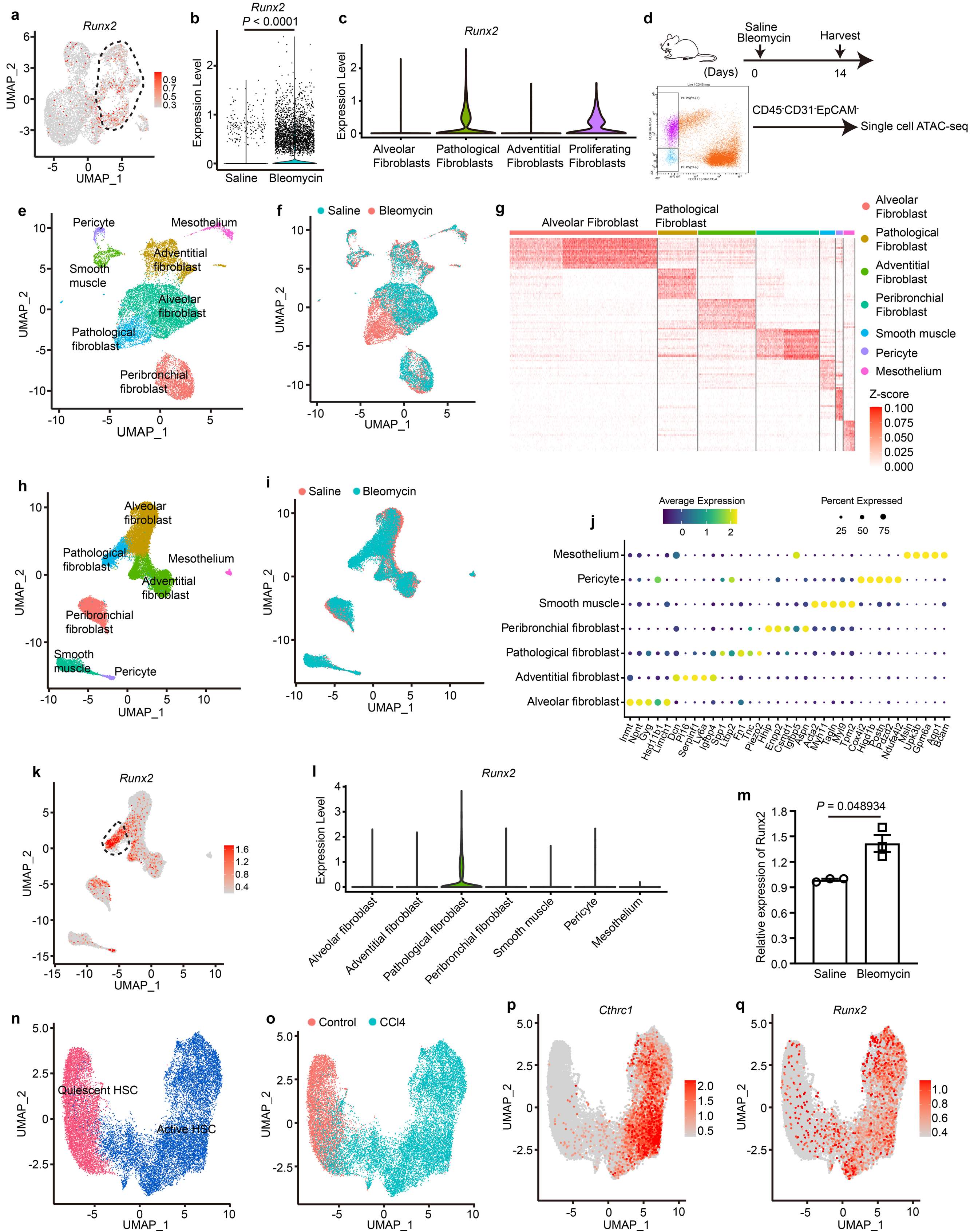


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | A novel *Higd1b-CreERT2* mouse strain specifically labels pericytes which rarely contribute to pathological fibroblasts during lung fibrosis. a, UMAP plot showing different mesenchymal populations from the adult control lungs that integrate the datasets GSE132771, GSE201698 and GSE211713. **b-c**, UMAP plot showing the expression levels of *Higd1b* (**b**) and *Cox4i2* (**c**) in pericytes. Of note, 88% pericytes expressed *Higd1b*. **d**, Schematic depicting the generation of *Higd1b-CreERT2* mouse strain. **e**, Schematic diagram for Tmx injection of adult *Higd1b-CreERT2;R26-tdT* mice and representative images showing immunostaining of pericyte marker NG2 and tdT. **f**, Schematic diagram for Tmx injection and bleomycin challenge of adult *Higd1b-CreERT2;R26-tdT* mice and representative images showing immunostaining of α-SMA, ERG and tdT. **g**, Representative images of two examples showing immunostaining of Cthrc1 and tdT in the lungs of *Higd1b-CreERT2;R26-tdT* mice challenged with bleomycin. **h**, Quantification analysis showing the rare contribution of pericytes to Cthrc1$^+$ pathological fibroblasts in the lungs of *Higd1b-CreERT2;R26-tdT* mice (n = 6). Data are representative of at least three independent experiments. Data are mean ± SEM. **i-j**, Representative images showing immunostaining of tdT, Collagen I (**i**) and Ki67 (**j**) in the lungs of *Higd1b-CreERT2;R26-tdT* mice challenged with bleomycin. Scale bars: 100 µm (20 µm in magnified views).

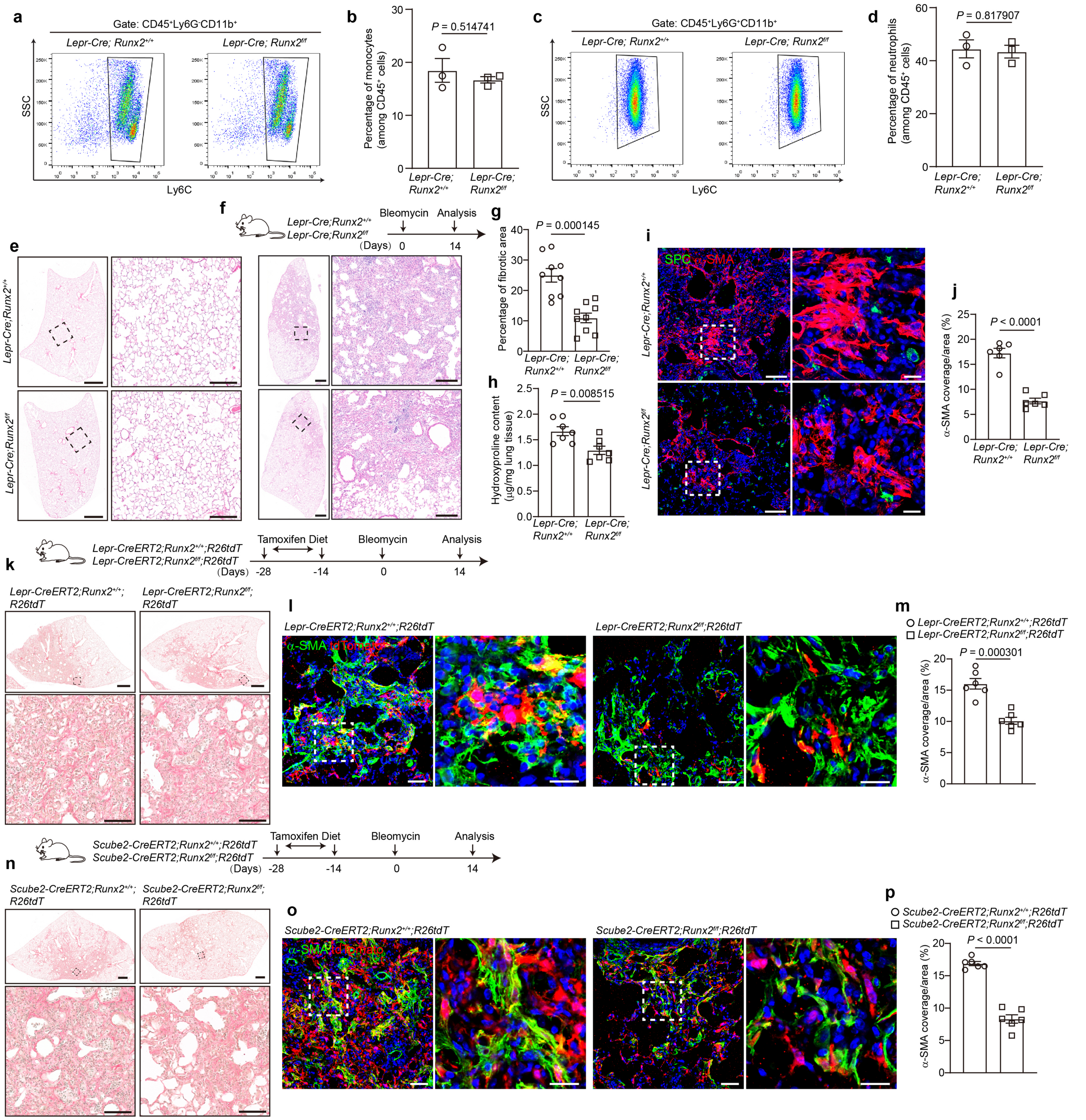


Extended Data Fig. 7 | *Postn*$^{+}$ pathological fibroblasts are identified in silica-challenged lungs. **a**, UMAP plot showing four different lung fibroblast populations in silica-treated lungs. The results were generated by re-analysis of the dataset GSE184854. **b**, Dot plot showing the representative markers for each fibroblast population. **c**-**d**, *Postn* expression shown by UMAP plot (**c**) and Violin Plot (**d**). **e**, Schematic diagram for Tmx injection and silica exposure of *Postn-CreER;R26tdT* mice and representative images showing immunostaining of α-SMA and tdT in lung tissues. **f**, Quantification analysis showing increased tdT^{+} cells upon silica exposure (Saline: n = 6, Silica: n = 6). **g**, Representative images showing immunostaining of α-SMA, ERG and Endomucin in lung tissues. Data are representative of at least three independent experiments. Data are mean ± SEM. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction. Scale bars: 100 μm (20 μm in magnified views).

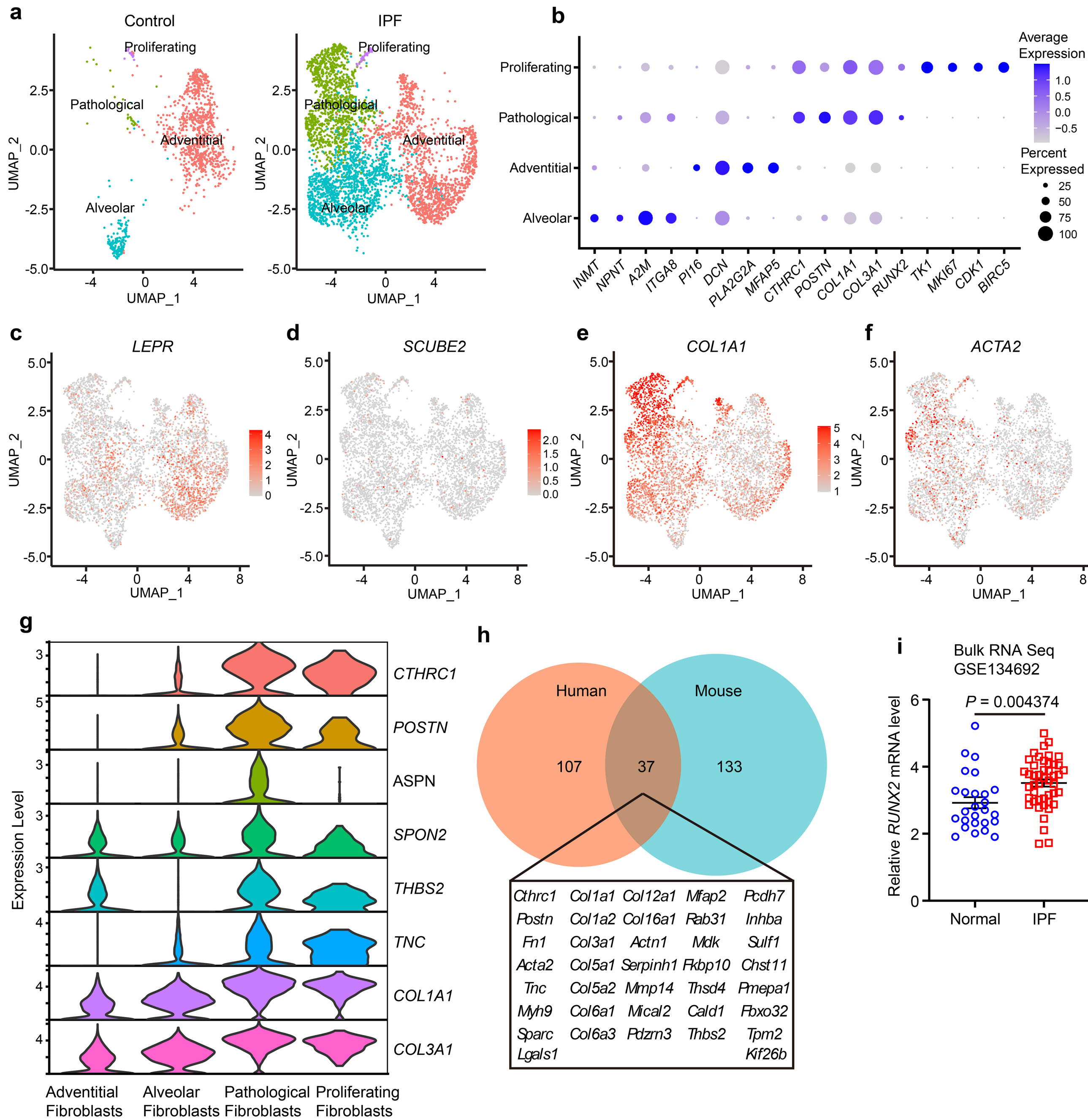


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | The TF-binding motif enrichment identified by scATAC-seq. a, UMAP plot showing the expression of *Runx2* transcripts in lung mesenchymal cells from *Lepr-Cre;R26tdT* mice challenged with saline or bleomycin. **b**, Violin plot showing the increased expression of *Runx2* in the lungs of *Lepr-CreERT2;R26tdT* mice challenged with bleomycin. **c**, Violin plot showing the enrichment of *Runx2* in pathological fibroblasts and proliferating fibroblasts of bleomycin-treated *Lepr-CreERT2;R26tdT* mouse lungs. **d**, Schematic depicting the experimental design for scATAC-seq. **e**, UMAP plot of scATAC-seq showing various lung mesenchymal populations. **f**, Blue and red dots showing the cells from saline- and bleomycin-treated lungs, respectively. **g**, Heatmap showing the representative markers for each lung mesenchymal population. **h**-**i**, UMAP plot showing various lung mesenchymal populations (**h**) from bleomycin-treated (blue dots) and control (red dots) lungs (**i**) that integrate the published scRNA-seq datasets GSE131800, GSE132771, GSE183545 and GSE201698. **j**, Dot plot showing the representative markers for each fibroblast population. **k**-**l**, UMAP plot (**k**) and Violin Plot (**l**) showing the enrichment of *Runx2* transcripts in pathological fibroblasts. **m**, RT-qPCR analysis of *Runx2* transcripts in FACs-sorted tdT$^+$ cells isolated from the lungs of *Lepr-Cre;R26tdT* mice treated with bleomycin or saline (Saline: n = 3, Bleomycin: n = 3). **n**-**o**, UMAP plot showing the quiescent and active hepatic stellate cells (**n**) of control (red dots) and CCl_4-treated (blue dots) mice (**o**). The dataset GSE171904 was used for re-analysis. HSC: Hepatic stellate cell. **p**-**q**. UMAP plot showing the expression of *Cthrc1* (**p**) and *Runx2* (**q**) in hepatic stellate cells. Data are representative of at least three independent experiments. Data are mean ± SEM. Statistical analysis was performed using two-sided Wilcoxon Rank Sum test (**b**) and unpaired two-tailed *t*-test with Welch's correction (**m**).



Extended Data Fig. 9 | Conditional deletion of *Runx2* attenuates bleomycin-induced lung fibrosis. a,**b**, FACS analysis to quantify bone marrow monocytes among CD45$^+$ cells (*Lepr-Cre;Runx2*$^{+/+}$: n = 3, *Lepr-Cre;Runx2*$^{f/f}$: n = 3). **c**,**d**, FACS analysis to quantify bone marrow neutrophils among CD45$^+$ cells (*Lepr-Cre; Runx2*$^{+/+}$: n = 3, *Lepr-Cre;Runx2*$^{f/f}$: n = 3). **e**, Representative H&E staining images of untreated lungs from *Lepr-Cre;Runx2*$^{+/+}$ and *Lepr-Cre;Runx2*$^{f/f}$ mice. **f**, Experimental schematic and representative H&E staining images of the lungs from *Lepr-Cre;Runx2*$^{+/+}$ and *Lepr-Cre;Runx2*$^{f/f}$ mice after bleomycin challenge. **g**, Quantification of fibrotic areas in the lungs of *Lepr-Cre;Runx2*$^{+/+}$ (n = 9) and *Lepr-Cre;Runx2*$^{f/f}$ mice (n = 9). **h**, Hydroxyproline content in the lungs of *Lepr-Cre;Runx2*$^{+/+}$ (n = 7) and *Lepr-Cre;Runx2*$^{f/f}$ mice (n = 7) treated with bleomycin. **i**, α-SMA and SPC expression in the lungs of *Lepr-Cre;Runx2*$^{+/+}$ and *Lepr-Cre;Runx2*$^{f/f}$ mice after bleomycin challenge. **j**, Quantification of α-SMA$^+$ cells in the lungs of *Lepr-Cre;Runx2*$^{+/+}$ (n = 6) and *Lepr-Cre;Runx2*$^{f/f}$ mice (n = 6). **k**, Experimental schematic and representative Picro-Sirius Red staining images. **l**, α-SMA and tdT expression in the lungs of *Lepr-CreERT2;Runx2*$^{+/+}$*;R26tdT* and *Lepr-CreERT2;Runx2*$^{f/f}$*;R26tdT* mice. **m**, Quantification of α-SMA$^+$ cells in the lungs of *Lepr-CreERT2;Runx2*$^{+/+}$*;R26tdT* (n = 6) and *Lepr-CreERT2;Runx2*$^{f/f}$*;R26tdT* mice (n = 6). **n**, Experimental schematic and representative Picro-Sirius Red staining images. **o**, α-SMA and tdT expression in the lungs of *Scube2-CreERT2; Runx2*$^{+/+}$*;R26tdT* and *Scube2-CreERT2;Runx2*$^{f/f}$*;R26tdT* mice. **p**, Quantification of α-SMA$^+$ cells in the lungs of *Scube2-CreERT2;Runx2*$^{+/+}$*;R26tdT* (n = 6) and *Scube2-CreERT2;Runx2*$^{f/f}$*;R26tdT* mice (n = 6). Data are representative of at least three independent experiments. Data are mean ± SEM. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction. Scale bars: **e**, **f**, **k** and **n**: 1 mm (**e** and **f**: 200 μm in magnified views, **k** and **n**: 100 μm in magnified views), **i**, **l** and **o**: 100 μm (20 μm in magnified views).



Extended Data Fig. 10 | Increased expression of *RUNX2* in IPF lungs. **a**, Separated UMAP plots showing fibroblasts from normal human lungs and IPF lungs. These results were generated through the re-analysis of the dataset GSE136831. **b**, Dot plot of scRNA-seq showing the representative markers for each fibroblast population. **c-f**, UMAP plot showing the expression levels of *LEPR* (**c**), *SCUBE2* (**d**), *COL1A1* (**e**) and *ACTA2* (**f**). **g**, Violin plot of scRNA-seq showing the expression levels of ECM associated genes. **h**, Venn diagram showing the common and differential genes expressed in pathological fibroblasts of human IPF and bleomycin-treated mouse lungs. **i**, Increased transcript levels of *RUNX2* in IPF lungs revealed by re-analyzing the bulk RNA seq dataset GSE134692 (Normal: n = 26, IPF: n = 46). Data are mean ± SEM. Statistical analysis was performed using unpaired two-tailed *t*-test with Welch's correction.

nature portfolio

Corresponding author(s): Jianwen Que, Xin Sun

Last updated by author(s): Nov 26, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Zeiss LSM 710 with ZEN 2012 SP5 FP3 software (black, Version 14.0.27.201) was used to obtain fluorescence images.
BD LSRII with FACSDIVA software (version 9.0) was used to collect flow cytometry data.
QuantStudio Design & Analysis software (version 1.5.2) was used for qPCR data collection.
EVOS M5000 Cell Imaging System software (version 1.4.1031.622) was used to obtain fluorescence images.

Data analysis

Zeiss ZEN 3.6 (blue edition, version 3.6.095.09000), R version 4.0.2, Aperio ImageScope software (v12.4.3.5008), GraphPad Prism version 8.0.2, FlowJo version 10.0.7 and ImageJ version 1.51w were used to analyze data. The custom reference and alignment for single cell RNA-seq were made by Cell Ranger version 3.1.0 and 7.1.0. Cell Ranger ATAC version 2.1.0 was used for reads alignment and cell barcode demultiplex of single cell ATAC-seq.
The following R packages were loaded for data analysis.
BSgenome.Mmusculus.UCSC.mm10_1.4.3 BSgenome_1.72.0 rtracklayer_1.64.0 BiocIO_1.14.0
Biostrings_2.72.0 XVector_0.44.0 TFBSTools_1.42.0
JASPAR2020_0.99.10 chromVAR_1.26.0 Signac_1.9.0
monocle3_1.3.7 SingleCellExperiment_1.26.0 SummarizedExperiment_1.34.0
GenomicRanges_1.56.0 GenomeInfoDb_1.40.1 IRanges_2.38.0
S4Vectors_0.42.0 MatrixGenerics_1.16.0 matrixStats_1.3.0
monocle_2.32.0 DDRTree_0.1.5 irlba_2.3.5.1
VGAM_1.1-11 Biobase_2.64.0 BiocGenerics_0.50.0 plotly_4.10.4 htmlwidgets_1.6.4
lubridate_1.9.3 forcats_1.0.0 stringr_1.5.1 purrr_1.0.2 readr_2.1.5
tidyr_1.3.1 tibble_3.2.1 tidyverse_2.0.0 enrichR_3.2 hypeR_2.2.0
DoMultiBarHeatmap_0.1.0 magrittr_2.0.3 rlang_1.1.4 dittoSeq_1.17.0 scales_1.3.0

RColorBrewer_1.1-3	ggsci_3.1.0	ggpubr_0.6.0	ggforce_0.4.2	clustree_0.5.1
ggraph_2.2.1	VennDiagram_1.7.3	futile.logger_1.4.3	Hmisc_5.1-3	DoubletFinder_2.0.4
readxl_1.4.3	dplyr_1.1.4	viridis_0.6.5	viridisLite_0.4.2	limma_3.60.2
cowplot_1.1.3	ggplot2_3.5.1	patchwork_1.2.0	SeuratWrappers_0.2.0	scCustomize_1.1.3
Matrix_1.7-0	SeuratObject_4.1.4	Seurat_4.4.0	ComplexHeatmap_2.10.0	

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The scRNA-seq and scATAC-seq data generated in this study were deposited in Gene Expression Omnibus (GEO) (accession # GSE229523, GSE276546 and GSE278419). The publicly available scRNA-seq data for mouse lung development (accession # GSE160876 and GSE165063), normal adult mouse lungs (accession # GSE132771, GSE201698 and GSE211713), bleomycin-challenged mouse lungs (accession # GSE131800, GSE132771, GSE183545 and GSE201698), silica-exposed mouse lungs (accession # GSE184854), CCl4-treated mouse liver (accession #GSE171904), human IPF patients (accession # GSE136831) and bulk RNA-seq data for human IPF patients (accession # GSE124685 and GSE134692) were used for analysis. Source data are provided with this paper.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	This research examined the expression of RUNX2 in normal and IPF lungs from patients with decoded identities . We included 5 normal samples (2 males + 3 females) and 8 IPF patients (4 males + 4 females).
Population characteristics	IPF patients were from 50-80 years of age, and normal samples were from 50-70 years of age.
Recruitment	De-identified individuals were enrolled into IRB-approved studies at Cedars Sinai Medical Center and Columbia University Medical Center
Ethics oversight	Approved by IRB committee at Columbia University (IRB AAAS4094) and Cedars Sinai Medical Center (IRB Pro00032727)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed. Sample sizes were determined empirically based on experience in our group or previous published relevant studies with similar layout.
Data exclusions	No data was excluded from analyses.
Replication	Each group included at least three replicates for all experiments, except for scRNA-seq analysis. Sample size is indicated in figure and figure legend.
Randomization	Gender- and age-matched mice were randomly allocated for experiments.
Blinding	No human subjects in clinical trials were involved in this study. No blinding was applied as the same investigators performed the experiments and analyzed results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☐	☒	Animals and other organisms
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used

For immunostaining imaging, primary antibodies:
anti-α-SMA (Santa Cruz, sc-32251, 1:200) https://www.scbt.com/p/alpha-actin-antibody-1a4?srsltid=AfmBOopBcRxD8aS_bKMjfvyNBaI7GbYnEkz-iBRPWw-zfGXCQamzfkJo
anti-tdtomato (biorbyt, orb182397, 1:1000) https://www.biorbyt.com/tdtomato-antibody-orb182397.html
anti-RFP (Rockland, 600-401-379, 1:500) https://www.rockland.com/categories/primary-antibodies/rfp-antibody-pre-adsorbed-600-401-379/?srsltid=AfmBOopibPVAcpLpVPaoOi463TVbaZIpUIAMn8dye6QEygIUDFFzOkyO
anti-Collagen I (SouthernBiotech, 1310-01, 1:200) https://www.southernbiotech.com/goat-anti-type-i-collagen-unlb-1310-01
anti-Pi16 (R&D, AF4929, 5 µg/ml) https://www.rndsystems.com/products/mouse-peptidase-inhibitor-16-pi16-antibody_af4929
anti-Cthrc1 (MaineHealth Institute for Research, Vli55, 1:250) https://mhir.org/?page_id=2008
anti-Postn (Abcam, ab215199, 1:200) https://www.abcam.com/en-us/products/primary-antibodies/periostin-antibody-epr20806-ab215199?srsltid=AfmBOoqZOLQ9SsZ48Jq-LYhhlBrLiUtETaLW_0L--uvsh-UU50phtucJ
anti-ERG (Abcam, ab92513, 1:200) https://www.abcam.com/en-us/products/primary-antibodies/erg-antibody-epr3864-ab92513?srsltid=AfmBOop6tD7RdIDcn24PeeZpBasTeVxGu4Ytbuvq_lIhWV2K_draxpHA
anti-Endomucin (Santa Cruz, sc-65495, 1:200) https://www.scbt.com/p/endomucin-antibody-v-7c7?srsltid=AfmBOoqXQ8F9_FDkULuY-gQb9aWSdj5RC2cgS212NxGvsJF-GjFABRe4
anti-Runx2 (Cell Signaling Technology, 12556S, 1:200) https://www.cellsignal.com/products/primary-antibodies/runx2-d1l7f-rabbit-mab/12556?srsltid=AfmBOooAZq10o2rm0fUZVjpe05Z5Hi7kbrcDNazEHzkpkJZWoU7CDIPb
anti-Ki67 (Cell Signaling Technology, 9129S, 1:200) https://www.cellsignal.com/products/primary-antibodies/ki-67-d3b5-rabbit-mab/9129
anti-NG2 (Millipore Sigma, AB5320, 1:200) https://www.sigmaaldrich.com/US/en/product/mm/ab5320?srsltid=AfmBOoprr5KkHP6lX5YrRlBoGB6PaUuEWg1SmkQh88rhojuTuX86zzT9
anti-ProSPC (Abcam, ab211326, 1:500) https://www.abcam.com/en-us/products/primary-antibodies/prosurfactant-protein-c-antibody-epr19839-ab211326?srsltid=AfmBOoqnPBGAkx2jarRkoluDN_Jv6mUYEZ4E1w1BAoUN3H8VsYIxpaKk#
Secondary antibodies:
Alexa Fluor 488 Donkey Anti-Mouse (Jackson Immuno Research, 715-545-151, 1:500) https://www.jacksonimmuno.com/catalog/products/715-545-151
Cy3 Donkey Anti-Goat (Jackson Immuno Research, 705-165-147, 1:500) https://www.jacksonimmuno.com/catalog/products/705-165-147
Alexa Fluor 488 Donkey Anti-Rabbit (Jackson Immuno Research, 711-545-152, 1:500) https://www.jacksonimmuno.com/catalog/products/711-545-152
Alexa Fluor 488 Donkey Anti-Rat (Jackson Immuno Research, 712-545-150, 1:500) https://www.jacksonimmuno.com/catalog/products/712-545-150

For FACS, antibodies:
PE-Cy7-CD45 (Biolegend, 103114, 1:100) https://www.biolegend.com/en-us/products/pe-cyanine7-anti-mouse-cd45-antibody-1903
APC-EpCAM (Biolegend, 118214, 1:100) https://www.biolegend.com/en-us/products/apc-anti-mouse-cd326-ep-cam-antibody-4974
BV711-CD31 (Biolegend, 102449, 1:100) https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd31-antibody-19258
BV510-CD45 (Biolegend, 103138, 1:100) https://www.biolegend.com/en-us/products/brilliant-violet-510-anti-mouse-cd45-antibody-7995
PE-Epcam (Biolegend, 118206, 1:100) https://www.biolegend.com/en-us/products/pe-anti-mouse-cd326-ep-cam-antibody-4726
PE-CD31 (Biolegend, 102408, 1:100) https://www.biolegend.com/en-us/products/pe-anti-mouse-cd31-antibody-122
APC-PDGFRa (Biolegend, 135908, 1:100) https://www.biolegend.com/en-us/products/apc-anti-mouse-cd140a-antibody-6439
Alexa Fluor 488-a-SMA (eBioscience, 53-9760-82, 1:100) https://www.thermofisher.com/antibody/product/Alpha-Smooth-Muscle-Actin-Antibody-clone-1A4-Monoclonal/53-9760-82
eFluor 450-CD45 (eBioscience, 48-0451-82, 1:100) https://www.thermofisher.com/antibody/product/CD45-Antibody-clone-30-F11-Monoclonal/48-0451-82
Alexa Fluor 700-Ly6G (Biolegend, 127622, 1:100) https://www.biolegend.com/en-us/products/alexa-fluor-700-anti-mouse-ly-6g-antibody-6754
Alexa Fluor 488-Ly6C (Biolegend, 128022, 1:100) https://www.biolegend.com/en-us/products/alexa-fluor-488-anti-mouse-ly-6c-antibody-6756

PE-Cy7-CD11b (eBioscience, 25-0112-82, 1:100) https://www.thermofisher.com/antibody/product/CD11b-Antibody-clone-M1-70-Monoclonal/25-0112-82

Validation

All the antibodies used in this study were validated for their specificity and specific application by the manufacturer and validation methods and data are described in the provided links to the manufacturer's websites.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Primary human lung fibroblasts were isolated from normal and IPF patient lungs. Primary mouse lung fibroblasts were isolated from Lepr-Cre;R26-tdtomato;Col1a1-EGFP mice.
Authentication	The cells have not been authenticated.
Mycoplasma contamination	All the primary fibroblasts have been tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study according to ICLAC register version II.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Lepr-Cre (#008320), Postn-MerCreMer (#029645), Rosa26-tdtomato (#007914), Rosa26-DTA (#009669) and ACTB-Flpe (#005703) were imported from Jackson Laboratory. Lepr-CreERT2 and Higd1b-CreERT2 mouse strains were generated in the Que lab. Scube2-CreERT2 mice were a kind gift from Dr. Dean Sheppard (University of California, San Francisco, CA). Runx2flox/flox (CDB0832K, RBBC10433) mice were a kind gift from Dr. Takeshi Takarada and imported from RIKEN BRC, Japan. Col1a1-EGFP mice were a kind gift from Dr. David Brenner (University of California, San Francisco, CA). All the mice between the ages of 8 and 12 weeks old maintained on C57BL/6 background (both male and female) were used in the experiments. Mice were housed with a 12h light/dark cycle at 18-23 degree and 40-60% humidity in the animal facility at Columbia University Medical Center.
Wild animals	No wild animals were used in this study.
Reporting on sex	Both male and female were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All mouse experiments and care were conducted in accordance with the procedures approved by the Institutional Animal Care and Use Committee at Columbia University (protocol # AABM6565).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To obtain single cell suspension, lung tissues were dissected, washed with PBS and minced with a razor blade followed by digesting in a digestion buffer (2 mg/ml collagenase-IV, 2 mg/ml dispase II and 10 U/ml DNase I) for 30 min at 37°C. DMEM containing 10% FBS was added to stop digestion and the cells were filtered through 100-µm and 40-µm strainers. After centrifuged at 300g for 5 min, cell pellet was resuspended in red blood cell lysis buffer (Sigma), incubated for 2 min at 37°C, followed by washing with HBSS containing 10% FBS and centrifuged at 300g for 5 min. Flow cytometry analysis was performed as we previously described. In brief, cells were incubated with PE-Cy7-CD45 (Biolegend, 103114), APC-EpCAM (Biolegend, 118214) and BV711-CD31 (Biolegend, 102449) antibodies in FACS buffer (5% FBS with 0.5mM EDTA in PBS) for one hour at 4°C, and then incubated with Live/Dead stain dye for 10 min at room temperature to exclude dead cells. After washing with FACS buffer, cells were fixed and permeabilized by using Fixation/Permeabilization buffer and then incubated with AF488-α-SMA antibody at 4°C for one hour. BD LSRII and FlowJo V10 software were used for obtaining data and analysis, respectively. Live CD45-EpCAM-CD31- mesenchymal cells were gated for further analysis. For Single cell RNA sequencing (scRNA-seq), single cell suspension was applied to sort live Tdtomato+ cells from Lepr-Cre;R26tdT mice or live Tdtomato +CD45-EpCAM-CD31- mesenchymal cells from Lepr-CreERT2;R26tdT mice by using BD Influx cell Sorter. For bone marrow

monocytes and neutrophils analysis, the single cell suspensions were obtained from femur bone marrow as previously described. Bone marrow cells were incubated with eFluor 450-CD45 (eBioscience, 48-0451), Alexa Fluor 700-Ly6G (Biolegend, 127622), Alexa Fluor 488-Ly6C (Biolegend, 128022) and PE-Cy7-CD11b (eBioscience, 25-0112) antibodies in FACS buffer for one hour at 4°C. The flow cytometry data were collected on BD LSRII and analyzed using FlowJo V10 software. CD45+Ly6G-CD11b+Ly6C+ monocytes and CD45+Ly6G+CD11b+Ly6C+ neutrophils were gated for analysis.

Instrument

BD LSRII equipped with four lasers (405nm, 488nm, 561nm and 635nm)

Software

BD FACSDiva v9.0 and FlowJo v10.0.7

Cell population abundance

The proportion of a-SMA+Tdtomato+ cells in Live CD45-EpCAM-CD31- mesenchymal cells varied from 2.21 to 13.6% depending on the saline and bleomycin treatment.
The proportion of CD45+Ly6G-CD11b+Ly6C+ monocytes among CD45+ immune cells varied from 15.23 to 22.81% in bone marrow from Lepr-Cre;Runx2+/+ and Lepr-Cre;Runx2f/f mice.
The proportion of CD45+Ly6G+CD11b+Ly6C+ neutrophils among CD45+ immune cells varied from 38.11 to 49.79% in bone marrow from Lepr-Cre;Runx2+/+ and Lepr-Cre;Runx2f/f mice.

Gating strategy

Live cells were gated by forward scatter, side scatter, doublets discrimination and by Live/Dead stain dye exclusion.
Mesenchymal cells were selected by CD45-, EpCAM- and CD31-. a-SMA+Tdtomato+ cells were gated based on the expression of a-SMA and Tdtomato.
Monocytes were gated by forward scatter, side scatter, doublets discrimination, CD45+, Ly6G-, CD11b+ and Ly6C+.
Neutrophils were gated by forward scatter, side scatter, doublets discrimination, CD45+, Ly6G+, CD11b+ and Ly6C+.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.