Title

Alveolar fibroblast lineage orchestrates lung inflammation and fibrosis

Permalink

https://escholarship.org/uc/item/8t25p17p

Journal

Nature, 631(8021)

ISSN

0028-0836

Authors

Tsukui, Tatsuya

Wolters, Paul J

Sheppard, Dean

Publication Date

2024-07-10

DOI

10.1038/s41586-024-07660-1

Copyright Information

Peer reviewed

Alveolar fibroblast lineage orchestrates lung inflammation and fibrosis

Tatsuya Tsukui[1], Paul J. Wolters[1], and Dean Sheppard[1*]

[1]Division of Pulmonary, Critical Care, Allergy and Sleep, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

*Address correspondence to: Dean Sheppard

University of California, San Francisco, 555 Mission Bay Blvd South, San Francisco, CA 94158-9001.

E-mail: Dean.Sheppard@ucsf.edu

Tel: +1-415-514-4269

Abstract:

Fibroblasts are present throughout the body to maintain tissue homeostasis. Recent studies have identified diverse fibroblast subsets in healthy and injured tissues[1,2], but the origin(s) and functional roles of injury-induced fibroblast lineages remain unclear. Here we show that lung-specialized alveolar fibroblasts take on multiple molecular states with distinct roles in facilitating responses to fibrotic lung injury. We generate a genetic tool that uniquely targets alveolar fibroblasts to demonstrate their role in providing niches for alveolar stem cells in homeostasis and show that loss of this niche leads to exaggerated responses to acute lung injury. Lineage tracing identifies alveolar fibroblasts as the dominant origin for multiple emergent fibroblast subsets sequentially driven by inflammatory and pro-fibrotic signals after injury. We identify similar, but not completely identical fibroblast lineages in human pulmonary fibrosis. TGF-β negatively regulates an inflammatory fibroblast subset that emerges early after injury and stimulates the differentiation into fibrotic fibroblasts to elicit intra-alveolar fibrosis. Blocking the induction of fibrotic fibroblasts in the alveolar fibroblast lineage abrogates fibrosis but exacerbates lung inflammation. These results demonstrate the multifaceted roles of the alveolar fibroblast lineage in maintaining normal alveolar homeostasis and orchestrating sequential responses to lung injury.

Main

Fibroblasts provide structural support to every organ by maintaining extracellular matrix (ECM) architecture[3]. In response to tissue injury, new subsets of fibroblasts, including cells that produce large amounts of ECM, emerge at injured sites. These cells play an important role in normal tissue repair but have been suggested to contribute to pathologic fibrosis in the setting of chronic diseases[4,5]. Pro-fibrotic fibroblasts have been historically described as myofibroblasts based on increased expression of alpha-smooth muscle actin (α-SMA). We and others have identified Cthrc1 (collagen triple helix repeat containing 1) as a more specific marker of the small subset of fibroblasts that produce the highest levels of ECM proteins in pulmonary fibrosis and other fibrotic diseases[1,2,6,7]. However, the origin of the cells that drive fibrotic pathology remains controversial. Although trans-differentiation from other cell types such as hematopoietic, epithelial, perivascular, and endothelial cells has been suggested[8,9], there is little evidence from extensive scRNA sequencing (scRNA-seq) of fibrotic tissue to support these non-fibroblast sources. Recent scRNA-seq studies identified diverse fibroblast subsets in healthy tissues with distinct transcriptional profiles and anatomical locations[2,10]. Using computational lineage inference, one group recently proposed that pro-fibrotic fibroblasts likely universally emerge from adventitial fibroblasts, marked by high expression of the gene encoding peptidase inhibitor 16 (Pi16)[1], while our previous computational lineage analysis suggested that pro-fibrotic fibroblasts in response to alveolar lung injury arise from alveolar fibroblasts[2], a fibroblast subset uniquely present in the alveolar region of the lung. Elucidating the trajectory for pro-fibrotic fibroblast development in response to lung injury could lead to new insights into therapeutic targets for pulmonary fibrosis.

Alveolar fibroblasts in homeostasis

3

We previously showed that healthy lungs contain multiple fibroblast subsets characterized by distinct anatomic localization in alveolar, adventitial, or peribronchial regions[2]. Previously reported tools to label lung fibroblasts do not adequately distinguish among these populations. For example, *Pdgfra* and *Tcf21* are broadly expressed in fibroblasts and inadequate to distinguish fibroblast subsets[1,2]. To specifically label alveolar fibroblasts, we generated Scube2-CreER mice (Fig. 1a). In the mouse, *Scube2* is expressed in the clusters of fibroblasts we previously showed localized to the alveolar region but not in adventitial fibroblasts, peribronchial fibroblasts, pericytes, or smooth muscle cells (Extended Data Fig. 1a, b)[2]. We crossed Scube2-CreER mice with Rosa26-tdTomato mice and injected tamoxifen in the steady state (Fig. 1a). Flow cytometry showed that 93.80 ± 0.73 % (n = 3, $\pm$ SEM) of tdTomato+ cells were negative for lineage markers (CD31, CD45, EpCAM, Ter119 and Mcam), and that lineage− tdTomato+ cells were essentially all Sca1−, CD9−, and Pdgfra+, which is consistent with our previous immunophenotyping of alveolar fibroblasts (Fig. 1b, c, Extended Data Fig. 1c-f)[2]. After two weeks of tamoxifen, labeling efficiency for lineage−, Sca1−, CD9−, and Pdgfra+ cells was 77.63 ± 3.34 % (n =3, $\pm$ SEM) (Extended Data Fig. 1g, h). We crossed Scube2-tdTomato reporter mice with Col1a1-GFP (Col-GFP) reporter mice, where all fibroblasts express GFP. By whole lung imaging, tdTomato+ signals were found diffusely throughout the alveolar region, while Col-GFP signals prominently highlighted bronchovascular bundles due to the higher cellular density of fibroblasts in adventitial and peribronchial locations (Fig. 1d). Imaging thick sections also confirmed the alveolar localization of tdTomato+ cells (Fig. 1e). Col-GFP+ cells in adventitial cuff spaces and peribronchial areas co-localized with Pi16 staining and were not labeled by tdTomato (Extended Data Fig. 1i). Scube2-CreER thus labels alveolar fibroblasts but no other fibroblast subsets.

Previous work showed that Pdgfra+ lipofibroblasts in alveolar walls are closely associated with alveolar type 2 epithelial (AT2) cells and can support the growth of AT2 cells ex vivo[11,12]. Our previous study showed that alveolar fibroblasts that express *Scube2* also express a lipid droplet marker, *Plin2*[2]. We asked if Scube2-CreER-labeled alveolar fibroblasts might directly contact AT2 cells and support their maintenance in uninjured lungs. Pro-surfactant protein C (proSP-C)+ AT2 cells closely localized with the cell bodies of tdTomato+ alveolar fibroblasts (Fig. 1f, Supplementary Video 1), which extended projections around AT2 cells (Extended Data Fig. 1j). We crossed Scube2-CreER mice with Rosa26-DTA mice to ablate alveolar fibroblasts (Fig. 1g). Three days after 10-day tamoxifen treatment, the frequency of proSP-C+ cells decreased on histology (Fig. 1h, i). Flow cytometry confirmed that the number of alveolar fibroblasts decreased by $30-40\%$ after ablation and that AT2 cells identified as EpCAM+ MHC class II (MHCII)+ also decreased by 30-40% (Fig. 1j, k, Extended Data Fig. 2a-c)[13]. The decrease of alveolar fibroblast and AT2 cell markers was also confirmed by whole lung quantitative PCR (qPCR) (Fig. 1l), although no structural abnormality was observed in the absence of injury (Extended Data Fig. 2d, e). We then asked if the decreased number of alveolar fibroblasts affects the response to lung injury induced by intratracheal bleomycin (Fig. 2m). Alveolar fibroblast ablation significantly increased weight loss and mortality with increased IgM in bronchoalveolar lavage fluid (BAL) (Fig. 2n) and neutrophil infiltration (Fig. 2o), suggesting increased vascular permeability and neutrophilic inflammation after bleomycin injury. Whole lung qPCR showed that *Il17a* was dramatically up-regulated after bleomycin in alveolar fibroblast-ablated lungs (Extended Data Fig. 2f). We examined γδ T cells, CD4 T cells, and innate lymphoid cells (ILC), major cell types previously reported to produce IL-17a in the lung[14] (Extended Data Fig. 2g-j), and found increased γδ T cells and up-regulated Il17a expression in γδ T cells in the alveolar fibroblast-ablated group (Fig. 2p, q, Extended Data Fig. 2i). IL-17a neutralizing antibody abrogated the

exaggerated weight loss and mortality in alveolar fibroblast-ablated mice (Fig. 2r, s). Together, these

data provide in vivo evidence of a role for alveolar fibroblasts in maintaining a supportive niche for

AT2 cells in the steady state and show that loss of this niche leads to lethal IL-17a-mediated

inflammation after lung injury.

Alveolar fibroblast response to injury

To investigate the fate of alveolar fibroblasts after lung injury, we performed scRNA-seq in Scube2-

CreER/Rosa26-tdTomato mice (Fig. 2a). We collected lungs on days 0 (untreated), 7, 14, and 21 with

3 biological replicates at each time point (Fig. 2a). Mesenchymal cells were purified for scRNA-seq

(Fig. 2a, Extended Data Fig. 3a)[2]. We identified 11 clusters from 47,476 cells (Fig. 2b-e, Extended

Data Fig. 3b-f). In addition to the mesenchymal subsets we found in our previous study, we found 4

distinct clusters that emerged in response to lung injury. We labeled these as fibrotic, inflammatory,

stress-activated, and proliferating based on the patterns of gene expression (Extended Data Fig. 3c-

f). Fibrotic fibroblasts were characterized by the expression of *Cthrc1* and high expression of *Col1a1*

and other pathologic ECM genes (Fig. 2d, e). Inflammatory fibroblasts expressed chemokines such

as *Cxcl12* and were marked by specific expression of serum amyloid A3 (*Saa3*), lipocalin 2 (*Lcn2*),

and interferon-responsive genes (Fig. 2d, e). Gene ontology (GO) enrichment analysis of

differentially expressed genes in inflammatory fibroblasts suggested responses to inflammatory

cytokines including interferons and interleukin 1 (IL-1) (Extended Data Fig. 3d, f), similar to the

inflammatory fibroblast subsets reported in arthritis or cancer[15,16]. Stress-activated fibroblasts were

characterized by the expression of the cell cycle arrest marker p21 (*Cdkn1a*), translation-related

genes, and stress-related genes (Fig. 2d, e, Extended Data Fig. 3e, f). There were very few cells in

any of these 4 emergent subsets in the absence of injury (Fig. 2c, Extended Data Fig. 3b). The

123   frequency of inflammatory and proliferating fibroblasts peaked at day 7 and decreased at later time

124   points (Fig. 2c). Fibrotic fibroblasts started to emerge on day 7 but their frequency increased on day

125   14 and 21 (Fig. 2c). To assess the contribution of alveolar fibroblasts to each of these subsets, we

126   analyzed tdTomato expression. Although almost all cells showed at least low-level tdTomato

127   expression due to the baseline leak at the Rosa26 locus in the tdTomato reporter line, cells that

128   underwent CreER-mediated recombination showed much higher tdTomato expression (Extended

129   Data Fig. 4a, b). We defined cells with normalized tdTomato levels above 3.5 as tdTomato+ cells

130   (Fig. 2f, Extended Data Fig. 4b) and quantified tdTomato+ cells in each cluster (Fig. 2g, h). In the

131   steady state, 70 – 80% of alveolar fibroblasts were tdTomato+ while the other subsets contained very

132   low numbers of tdTomato+ cells (Fig. 2g). The tdTomato+ frequency for each of the 4 emergent

133   subsets was virtually identical to the tdTomato+ frequency of alveolar fibroblasts (Fig. 2h). This

134   pattern was consistent across all replicates (Extended Data Fig. 4c). Moreover, when we compared

135   tdTomato+ frequencies for alveolar fibroblasts and Cthrc1+ fibrotic fibroblasts in individual

136   replicates, all were close to the line of identity (Extended Data Fig. 4d). These data suggest that

137   alveolar fibroblasts are the dominant origin of all the emergent fibroblasts after injury, although we

138   cannot exclude small contributions from other fibroblast subsets.

139

140   We next evaluated the emergence of fibrotic and inflammatory fibroblasts by histology and flow

141   cytometry. Whole lung imaging showed an accumulation of Scube2-CreER-labeled cells in

142   aggregates in alveolar regions 14 days after bleomycin treatment (Extended Data Fig. 5a, b). We

143   previously showed that Cthrc1+ fibroblasts are enriched in a CD9+ fraction[2]. Some tdTomato+ cells

144   up-regulated CD9 on day 21 after bleomycin treatment, while tdTomato+ cells from untreated mice

145   were mostly CD9− (Extended Data Fig. 5c, d). Purified tdTomato+ CD9+ cells on day 21 expressed

higher fibrotic fibroblast markers and lower alveolar fibroblast markers compared to all tdTomato+ cells or tdTomato+ cells from untreated mice (Extended Data Fig. 5e), suggesting that Scube2-CreER-labeled alveolar fibroblasts differentiated into Cthrc1+ CD9+ fibrotic fibroblasts after lung injury. We also confirmed that some Scube2-CreER-labeled cells became inflammatory fibroblasts by staining for Saa3 (Extended Data Fig. 5f, g).

To see if the alveolar fibroblast origin of emergent fibroblasts is conserved in another model of lung fibrosis, we administered intratracheal silica to Scube2-CreER/Rosa26-tdTomato mice. After intratracheal silica, tdTomato+ silicotic nodules with collagen deposition emerged in alveolar regions (Extended Data Fig. 6a-e). Immunohistochemistry revealed that approximately 80% of the fibroblasts in silicotic nodules were tdTomato+ (Extended Data Fig. 6f, g). We also observed up-regulation of fibrotic and inflammatory markers in Scube2-CreER-labeled fibroblasts (Extended Data Fig. 6h-k). These data suggest that alveolar fibroblasts are the major origin of fibroblasts that form silicotic nodules and that markers of inflammatory and fibrotic fibroblasts are induced in Scube2-CreER-labeled cells in the silicosis model as well.

Although the presence of multiple emergent fibroblast subsets has been reported in other pathologies[15–18], the lineage relationships among these subsets remain unclear. To address this question, we focused on alveolar fibroblasts and the 3 largest emergent populations, inflammatory, stress-activated, and fibrotic fibroblasts, and performed pseudotime analysis (Extended Data Fig. 7a)[19]. This analysis suggested that stress-activated fibroblasts arose from inflammatory fibroblasts and were potentially in a terminal state (Extended Data Fig. 7a). Since fibrotic fibroblasts seemed to be the other terminal state and a trajectory went through inflammatory fibroblasts, we further focused

on alveolar, inflammatory, and fibrotic fibroblasts (Fig. 3a, b). Changes in representative markers along the pseudotime showed that alveolar fibroblast markers gradually decreased along the pseudotime towards fibrotic fibroblasts (Fig. 3c, d, Extended Data Fig. 7c). Inflammatory fibroblast markers increased in the middle of the pseudotime projection, but decreased later, along with increases of fibrotic fibroblast markers (Fig. 3c, d, Extended Data Fig. 7c). Overlay of some of these markers on UMAP plots indicated heterogeneity within fibrotic fibroblasts, with *Spp1* potentially expressed prior to *Cthrc1* along the pseudotime (Fig. 3d, Extended Data Fig. 7c). Although *Cthrc1+* cells expressed the highest levels of collagen genes, broad *Col1a1* expression in *Cthrc1−* cells in the fibrotic fibroblast cluster suggested a potential contribution of these cells to fibrosis (Extended Data Fig. 7c). Together, these data suggest that inflammatory fibroblasts are induced early after injury and fibrotic fibroblasts emerge later. Although one trajectory through which inflammatory fibroblasts could serve as an intermediate for the eventual emergence of fibrotic fibroblasts was suggested, confidence in such a model will require additional direct experimental evidence.

Since the differentially expressed genes of inflammatory fibroblasts indicated activation by inflammatory cytokines, we tested if inflammatory cytokines could induce inflammatory fibroblast markers in vitro (Fig. 3e). IL-1β and TNF-α are inflammatory cytokines produced in the early phase of bleomycin injury[20], while TGF-β1 is a profibrotic cytokine up-regulated in the later phase[21]. We stimulated freshly isolated murine alveolar fibroblasts with IL-1β, TNF-α or TGF-β1 (Fig. 3e, f). IL-1β dramatically up-regulated *Saa3* and *Lcn2* expression, while TGF-β1 up-regulated *Col1a1* and *Cthrc1* as previously described (Fig. 3f)[2]. TGF-β1 treatment up-regulated *Col1a1* and *Cthrc1* within 24 hours regardless of prior stimulation by IL-1β (Extended Data Fig. 7d, e). In contrast, IL-1β-induced *Saa3* and *Lcn2* were down-regulated when cells were subsequently stimulated with TGF-β1

192 (Extended Data Fig. 7e). These data suggest that TGF-β1 antagonizes expression of inflammatory

193 fibroblast markers and is a potent driver of the fibrotic fibroblast signature.

194

Alveolar fibroblast lineage in humans

196 We next evaluated whether a similar lineage of alveolar, inflammatory, and fibrotic fibroblasts could

197 be inferred from scRNA-seq data from human lungs. We first used our previously reported scRNA-

198 seq data[2]. We focused on alveolar and pathologic fibroblast clusters and re-clustered these to get

199 higher resolution (Fig. 3g-j, Extended Data Fig. 8a-d). We identified two previously unrecognized

200 clusters expressing inflammatory chemokines, which were enriched in idiopathic pulmonary fibrosis

201 (IPF) and scleroderma samples (Fig. 3h, i, j, Extended Data Fig. 8a, b). GO enrichment analysis

202 suggested that inflammatory cluster 1 was potentially induced by IL-1 and/or TNF, while

203 inflammatory cluster 2 was potentially induced by interferons (Extended Data Fig. 8c, d).

204 Transcriptomic comparison to mouse emergent clusters also supported the similarity between the

205 mouse inflammatory cluster and human inflammatory cluster 1, as well as the similarity of fibrotic

206 clusters and alveolar clusters from both species (Extended Data Fig. 8e). Although we did not observe

207 a distinct cluster of stress-activated fibroblasts in humans, each of these inflammatory clusters was

208 also enriched with GO terms related to stress responses (Extended Data Fig. 8c, d). Analysis of unique

209 GO terms showed features for antigen presentation in inflammatory cluster 2 (Extended Data Fig.

210 8f), suggesting potential immune-modulating roles as has been recently described for a subset of

211 cancer-associated fibroblasts[22]. Pseudotime analysis showed trajectories from alveolar fibroblasts

212 through each inflammatory cluster toward the fibrotic cluster (Fig. 3k), consistent with the results

213 from our mouse model. We next examined if this lineage was observed in other data sets with larger

214 sample sizes for human pulmonary fibrosis[23,24]. We extracted alveolar and pathologic subsets from

215 the publicly available data sets and merged these cells with our alveolar and pathologic fibroblast

216 clusters (Extended Data Fig. 9a-h). We confirmed the presence of two inflammatory clusters and a

217 fibrotic cluster in multiple lung samples from Adams et al. and Habermann et al. (Extended Data Fig.

218 9e-h).

219

220 We then asked if the localization of these fibroblast subsets in human pulmonary fibrosis shows any

221 association with pathology. We selected 2 or 3 markers that can jointly identify each fibroblast subset

222 and performed in situ hybridization on sequential IPF sections (Fig. 3l-n). We selected *COL1A1* and

223 *CTHRC1* for fibrotic fibroblast, *SFRP2* and *CCL2* for inflammatory fibroblast 1, and *SFRP4* and

224 *CXCL14* for inflammatory fibroblast 2 (Fig. 3n). Although some adventitial fibroblasts also express

225 *SFRP2* or *SFRP4*, expression of *ITGA8* can exclude adventitial fibroblasts (Fig. 3m). IPF is

226 characterized by the presence of fibroblastic foci as sites of ongoing fibrogenesis and adjacent fibrous

227 areas accompanied by mild inflammation[25]. Consistent with previous reports, *COL1A1*[high] and

228 *CTHRC1*+ fibrotic fibroblasts were enriched inside fibroblastic foci (Fig. 3o, p)[2]. *SFRP2*+, *CCL2*+,

229 and *ITGA8*+ inflammatory fibroblasts 1 were mostly localized adjacent to fibroblastic foci (Fig. 3o,

230 p). *SFRP4*+, *CXCL14*+, and *ITGA8*+ inflammatory fibroblasts 2 were present in fibroblastic foci but

231 biased to the interstitial side (Fig. 3o, p). These data suggest that the two inflammatory fibroblast

232 subsets are closely localized to regions of active fibrosis.

233

Fibrogenesis by Cthrc1+ fibroblasts

235 Cthrc1+ fibroblasts uniquely emerge after lung injury and localize at the leading edge of fibrogenesis

236 in IPF[2,26]. To evaluate the pro-fibrotic function of Cthrc1+ fibroblasts, we generated Cthrc1-CreER

237 mice by knocking a P2A-CreERT2-T2A-GFP sequence into the last exon of the *Cthrc1* gene (Fig.

238   4a). Because GFP generated by this construct was not detectable by flow cytometry or tissue

239   microscopy, we crossed Cthrc1-CreER mice with Rosa26-tdTomato mice and injected tamoxifen on

240   days 8 – 12 after injury (Fig. 4b). tdTomato+ cells emerged and formed aggregates in bleomycin-

241   treated lungs (Fig. 4c). Flow cytometry showed the emergence of tdTomato+ cells among lineage

242   (CD31, CD45, EpCAM, and Ter119)− cells in bleomycin-treated lungs but not in saline-treated lungs

243   (Extended Data Fig. 10a-c). qPCR showed that the expression of fibrotic genes was highly enriched

244   in tdTomato+ cells compared to all mesenchymal (lineage−) cells or all lung cells (Fig. 4e). Some

245   tdTomato+ fibroblasts showed intermediate CD9 expression on day 14, consistent with our previous

246   study[2], and the majority of tdTomato+ cells became CD9+ on day 21 (Extended Data Fig. 10d-f).

247   Saa3 immunostaining on day 14 showed that fibrotic and inflammatory fibroblasts appeared to

248   aggregate in adjacent but not overlapping regions as we observed in human pulmonary fibrosis

249   (Extended Data Fig. 10g, h). To examine the association of Cthrc1+ fibroblasts with de novo

250   fibrogenesis, we stained sequential sections for collagen 1 or Pi16. In uninjured lungs, collagen 1+

251   areas were mostly around Pi16+ bronchovascular bundles, reflecting normal collagen deposition in

252   bronchovascular cuffs (Extended Data Fig. 10i). After bleomycin, de novo collagen 1+ areas emerged

253   in alveolar areas (Extended Data Fig. 10i). Cthrc1+ fibroblasts showed close localization to collagen

254   1 as illustrated by mean distances of tdTomato to collagen 1, while there was no geographical

255   association between Cthrc1+ fibroblasts and Pi16+ bronchovascular cuffs (Extended Data Fig. 10j,

256   k). These data suggest that Cthrc1-CreER successfully targets Cthrc1+ fibroblasts at sites of de novo

257   fibrogenesis, which are anatomically distant from adventitial cuff spaces.

258

259   To micro-anatomically characterize the emergence of Cthrc1+ fibroblasts, we stained lung sections

260   for collagen 1 and collagen 4. Previous work suggested that fibroblasts migrate across the basal

261   lamina into alveolar airspaces to form fibroblastic foci in human pulmonary fibrosis[27]. After

262   bleomycin, we observed some Scube2-CreER-labeled cells within airspace lumens. Newly

263   synthesized collagen 1 in airspace lumens, which was not present in uninjured lungs, was adjacent to

264   these cells (Fig. 4e). The same analysis with Cthrc1-CreER mice revealed that the majority of Cthrc1-

265   CreER-labeled cells were within alveolar lumens and associated with intra-alveolar collagen 1 (Fig.

266   4e-g). These results are consistent with our previous demonstration that Cthrc1+ fibroblasts are highly

267   migratory and suggest that Cthrc1+ fibroblasts are a principal source of intraluminal collagen in

268   pulmonary fibrosis[2].

269

270   We crossed Cthrc1-CreER to Rosa26-lox-stop-lox-DTA mice and asked if ablation of Cthrc1+ cells

271   reduces fibrosis (Fig. 4h). Ablation efficiency assessed by loss of tdTomato+ cells and *Cthrc1* mRNA

272   was approximately 50% (Fig. 4i, j). Despite this limited ablation, fibrosis measured by

273   hydroxyproline content was significantly reduced by Cthrc1+ cell ablation (Fig. 4k). Collagen 1+

274   area also decreased in the ablated group (Extended Data Fig. 10l, m). Although the technical inability

275   to more efficiently delete Cthrc1+ fibroblasts resulted in only partial protection from pulmonary

276   fibrosis, these data demonstrate that Cthrc1+ fibroblasts do contribute to de novo fibrogenesis after

277   alveolar injury.

278

279   **Alveolar fibroblast deletion of Tgfbr2**

280   To further evaluate the alveolar fibroblast origin of Cthrc1+ fibrotic fibroblasts and to examine the

281   role of TGF-β signaling in the emergence of these cells, we conditionally deleted *Tgfbr2* from alveolar

282   fibroblasts (Fig. 5a). We observed nearly complete inhibition of pulmonary fibrosis assessed by

283   hydroxyproline content and a dramatic reduction in intra-alveolar collagen 1 in *Tgfbr2* conditional

284 knockout (cKO) mice (Fig. 5 b, c, d Extended Data Fig. 11a). However, *Tgfbr2* cKO mice showed

285 more severe body weight loss and increased mortality in the inflammatory phase after bleomycin

286 treatment (Fig. 5e, Extended Data Fig. 11b). qPCR of Scube2-CreER-labeled tdTomato+ cells

287 showed a substantial decrease of the fibrotic fibroblast markers in *Tgfbr2* cKO but an increase of

288 inflammatory fibroblast markers (Fig. 5f, Extended Data Fig. 11c). This increase in inflammatory

289 fibroblast markers was associated with increased alveolar permeability as evidenced by increased

290 IgM and red blood cells in BAL (Fig. 5g, h, Extended Data Fig. 11d, e), suggesting that exaggerated

291 lung inflammation and more severe acute lung injury in *Tgfbr2* cKO mice might contribute to the

292 body weight loss and increased mortality. *Tgfbr2* cKO mice also had increased monocytes in BAL,

293 consistent with increased expression of the monocyte chemoattractants, *Ccl2* and *Ccl7* (Fig. 5h,

294 Extended Data Fig. 11d). The presence of increased numbers of inflammatory fibroblasts (detected

295 by Saa3) and myeloid cell accumulation (detected by CD68) is also confirmed by

296 immunohistochemistry (Extended Data Fig. 11f-h). These data suggest that, in addition to being a

297 central driver of fibrosis, TGF-β signaling in fibroblasts plays a critical role in shutting off fibroblast-

298 driven inflammation. These findings support the conclusion that, in response to injury, alveolar

299 fibroblasts undergo a temporal progression to inflammatory fibroblasts and then fibrotic fibroblasts,

300 and that fibrotic fibroblasts derived from alveolar fibroblasts are the central drivers of pulmonary

301 fibrosis (Fig. 5i).

302

303 **Discussion**

304 In this study, we show that alveolar fibroblasts, which maintain alveolar homeostasis in the steady

305 state by providing niches for AT2 cells, are the dominant source of emergent fibroblast subsets after

306 fibrotic lung injury. Our results suggest that alveolar fibroblasts are likely induced to differentiate

307 into inflammatory fibroblasts by inflammatory cytokines during the initial phase of injury and that

308 fibrotic fibroblasts are later induced by pro-fibrotic cytokines like TGF-β1. *Tgfbr2* cKO using

309 Scube2-CreER demonstrates that this sequential lineage transition is critical not only for fibrosis but

310 also for terminating fibroblast enhancement of inflammatory responses to lung injury. The

311 exaggerated inflammation after injury in *Tgfbr2* cKO mice is consistent with the evolutionarily

312 conserved role of TGF-β in the orderly resolution of tissue injury[28]. We also demonstrate that Cthrc1+

313 fibroblasts, which have been described to emerge in IPF[2], scleroderma-associated pulmonary

314 fibrosis[2,29], SARS-CoV-2-associated lethal pneumonia[7], myocardial infarction[6], and cancer[1,30] are

315 significant contributors to fibrosis, although there are conflicting reports about the molecular function

316 of Cthrc1 in acute injury and fibrosis[31,32]. One important limitation of our study is the partial reduction

317 of fibrosis after Cthrc1+ fibroblast ablation. One possible explanation for this finding is the limited

318 recombination efficiency in the Rosa26-DTA locus, which has a longer distance between the two

319 loxp sites compared to Rosa26-tdTomato[33,34]. However, our scRNA-seq data also show that some

320 cells in the fibrotic fibroblast cluster do not express high levels of *Cthrc1*. It thus seems likely that

321 *Cthrc1−* fibrotic fibroblasts also contribute to fibrosis. Further study is required to determine whether

322 Cthrc1 is simply a late marker of the differentiation of fibrotic fibroblasts or whether the Cthrc1−

323 fibroblasts make unique contributions to the development of pulmonary fibrosis.

324

325 Previous efforts to trace the pro-fibrotic fibroblast lineage in pulmonary fibrosis were limited by a

326 lack of understanding of the fibroblast subsets in the normal lung and those that emerge in lung injury

327 and fibrosis. In this study, we developed a new mouse line, Scube2-CreER, that specifically

328 distinguishes alveolar fibroblasts from other cells present in the normal lung. Using this line, we now

329 clarify that previously proposed progenitors such as adventitial fibroblasts, pericytes, epithelial cells,

endothelial cells, and hematopoietic cells are not major sources of the new fibroblast subsets that emerge after fibrotic injury in the lung[8,9], but rather that all of these emergent subsets are principally derived from alveolar fibroblasts. However, our results do not exclude the possibility that other resting fibroblasts contribute to pathologic fibroblasts in response to injuries at other anatomic locations, such as perivascular or peri-airway fibrosis. Furthermore, whether fibrotic fibroblasts in other organs similarly arise from tissue-specific fibroblasts or other sources will need to be addressed with other tissue-specific tools. Further investigation of the mechanisms underlying the induction of these emergent populations and the functional contributions of each to progressive fibrosis or repair should lead to new therapeutic strategies targeting a wide array of diseases characterized by tissue inflammation and fibrosis.

Reference

1.   Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).

2.   Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun* **11**, 1920 (2020).

3.   Plikus, M. V. *et al.* Fibroblasts: Origins, definitions, and functions in health and disease. *Cell* **184**, 3852–3872 (2021).

4.   Hinz, B. & Lagares, D. Evasion of apoptosis by myofibroblasts: a hallmark of fibrotic diseases. *Nat Rev Rheumatol* **16**, 11–31 (2020).

5.   Henderson, N. C., Rieder, F. & Wynn, T. A. Fibrosis: from mechanisms to medicines. *Nature* **587**, 555–566 (2020).

6.   Ruiz-Villalba, A. *et al.* Single-Cell RNA Sequencing Analysis Reveals a Crucial Role for CTHRC1 (Collagen Triple Helix Repeat Containing 1) Cardiac Fibroblasts After Myocardial Infarction. *Circulation* **142**, 1831–1847 (2020).

355   7.   Melms, J. C. *et al.* A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).

356   8.   Pakshir, P. *et al.* The myofibroblast at a glance. *Journal of Cell Science* **133**, jcs227900 (2020).

357   9.   Friedman, S. L., Sheppard, D., Duffield, J. S. & Violette, S. Therapy for Fibrotic Diseases: Nearing the

358   Starting Line. *Science Translational Medicine* **5**, 167sr1-167sr1 (2013).

359   10. Narvaez del Pilar, O., Gacha Garay, M. J. & Chen, J. Three-axis classification of mouse lung

360   mesenchymal cells reveals two populations of myofibroblasts. *Development* **149**, dev200081 (2022).

361   11. McGowan, S. E. & Torday, J. S. The pulmonary lipofibroblast (lipid interstitial cell) and its

362   contributions to alveolar development. *Annu Rev Physiol* **59**, 43–62 (1997).

363   12. Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J Clin Invest* **123**, 3025–3036

364   (2013).

365   13. Hasegawa, K. *et al.* Fraction of MHCII and EpCAM expression characterizes distal lung epithelial cells

366   for alveolar type 2 cell isolation. *Respiratory Research* **18**, 150 (2017).

367   14. Gurczynski, S. J. & Moore, B. B. IL-17 in the lung: the good, the bad, and the ugly. *Am J Physiol Lung*

368   *Cell Mol Physiol* **314**, L6–L16 (2018).

369   15. Croft, A. P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**,

370   246–251 (2019).

371   16. Öhlund, D. *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic

372   cancer. *Journal of Experimental Medicine* **214**, 579–596 (2017).

373   17. Boyd, D. F. *et al.* Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* **587**,

374   466–471 (2020).

375   18. Korsunsky, I. *et al.* Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast

376   phenotypes in four chronic inflammatory diseases. *Med* **0**, (2022).

377   19. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–

378   502 (2019).

379   20. Liu, W. *et al.* Antiflammin-1 attenuates bleomycin-induced pulmonary fibrosis in mice. *Respiratory*

380   *Research* **14**, 101 (2013).

381   21. Strobel, B. *et al.* Time and phenotype-dependent transcriptome analysis in AAV-TGFβ1 and Bleomycin-

382   induced lung fibrosis models. *Sci Rep* **12**, 12190 (2022).

383   22. Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals

384   Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discovery* **9**, 1102–1123 (2019).

385   23. Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in

386   idiopathic pulmonary fibrosis. *Science Advances* **6**, eaba1983 (2020).

387   24. Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and

388   mesenchymal lineages in pulmonary fibrosis. *Science Advances* **6**, eaba1972 (2020).

389   25. Katzenstein, A.-L. A. & Myers, J. L. Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* **157**,

390   1301–1315 (1998).

391   26. Barratt, S. L., Creamer, A., Hayton, C. & Chaudhuri, N. Idiopathic Pulmonary Fibrosis (IPF): An

392   Overview. *J Clin Med* **7**, 201 (2018).

393   27. Kuhn, C. *et al.* An immunohistochemical study of architectural remodeling and connective tissue

394   synthesis in pulmonary fibrosis. *Am Rev Respir Dis* **140**, 1693–1703 (1989).

395   28. Massagué, J. & Sheppard, D. TGF-β signaling in health and disease. *Cell* **186**, 4007–4037 (2023).

396   29. Valenzi, E. *et al.* Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic

397   sclerosis-associated interstitial lung disease. *Annals of the Rheumatic Diseases* **78**, 1379–1387 (2019).

398   30. Dominguez, C. X. *et al.* Single-Cell RNA Sequencing Reveals Stromal Evolution into LRRC15+

399   Myofibroblasts as a Determinant of Patient Response to Cancer Immunotherapy. *Cancer Discovery* **10**,

400   232–253 (2020).

401   31. Binks, A. P., Beyer, M., Miller, R. & LeClair, R. J. Cthrc1 lowers pulmonary collagen associated with

402   bleomycin-induced fibrosis and protects lung function. *Physiol Rep* **5**, e13115 (2017).

403   32. Li, J. *et al.* Autocrine CTHRC1 activates hepatic stellate cells and promotes liver fibrosis by activating

404   TGF-β signaling. *EBioMedicine* **40**, 43–55 (2019).

405   33. Voehringer, D., Liang, H.-E. & Locksley, R. M. Homeostasis and effector function of lymphopenia-

406   induced 'memory-like' T cells in constitutively T cell-depleted mice. *J Immunol* **180**, 4742–4753 (2008).

34. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133–140 (2010).

Figure Legends

Fig. 1. **Scube2-CreER specifically labels alveolar fibroblasts and ablation of these cells leads to the loss of alveolar stem cell niches.**

(a) Schematic of Scube2-CreER mouse generation and experiments. (b) Flow cytometric quantification of tdTomato+ cells in each fraction. Pdgfra+, CD9+, or Sca1+ fractions were pre-gated on lineage− cells. n = 3 mice. (c) Flow cytometric analysis of lineage− Sca1− cells. (d) Maximum projection of whole lung imaging. (e) Maximum projection of 32 z-stack images with step size 0.9 μm. aw, airway. bv, blood vessel. al, alveoli. cuff, cuff space. (f) 28 z-stack images with step size 0.5 μm shown as maximum projection (left), color-coded projection to the depth from the top (middle) or the bottom (right). (g) Time course of tamoxifen treatment. (h) Representative lung sections with proSP-C staining. (i) Histological quantification of proSP-C+ cells. n = 5 mice. (j, k) Flow cytometric counting of alveolar fibroblasts (j) or AT2 cells (k). (l) qPCR analysis of all lung cells. n = 4 (vehicle) or 5 (Tamoxifen) mice (j, k). (m) Time course of tamoxifen and bleomycin treatment. (n) IgM in BAL measured by ELISA. n = 4 (saline) or 5 (bleomycin) mice. (o) Flow cytometric quantification of neutrophils in bleomycin-treated lungs. (p) qPCR for *Il17a* in purified populations. (q) Flow cytometric quantification of IL-17a+ cells in bleomycin-treated lungs. (o-q) n = 4 (Rosa26-DTA/DTA) or 7 (Rosa26-WT/WT) mice. (r) Percent body weight change after bleomycin treatment. (s) Survival after bleomycin treatment. n = 10 mice (r, s). Scale bars, 1 mm (d), 100 μm (e), 10 μm (f), 50 μm (h). Data are mean ± SEM. Data are representative of at least two independent experiments. Statistical analysis was performed using unpaired two-tailed t-test (i, j, k, o), unpaired two-tailed t-

433   test followed by Holm–Sidak's multiple-comparisons adjustment (l), two-tailed Mann-Whitney test

434   (n), or two-way analysis of variance (ANOVA) followed by Sidak's multiple comparison test (p, q).

435

436

437   **Fig. 2**. **Lineage tracing by scRNA-seq reveals alveolar fibroblasts as the origin of multiple**

438   **emergent fibroblast subsets**

439   (a) Schematic of scRNA-seq experiment design. (b) UMAP plot of the scRNA-seq data. Clusters

440   were shown with different colors. (c) Frequency of each cluster on different time points. (d) Dot plot

441   showing representative markers for each cluster. (e) Dot plot showing markers for subsets that emerge

442   after injury. (f) UMAP plot showing tdTomato+ and tdTomato-negative cells. (g) Percent tdTomato+

443   of each subset that was present in normal lungs. (h) Percent tdTomato+ of each subset that emerged

444   after injury. n = 3 mice (g, h). Data are mean ± SEM.

445

446

447   **Fig. 3. Alveolar fibroblasts sequentially differentiate into inflammatory and fibrotic fibroblasts**

448   **in mouse and human pulmonary fibrosis**

449   (a) UMAP plot of scRNA-seq data subsetted into alveolar, inflammatory, and fibrotic fibroblasts. (b)

450   UMAP plot overlaid with pseudotime. (c) Scaled expression of representative markers in

451   pseudospace. (d) Heat map with cells arranged in pseudotemporal order showing changes in

452   representative markers. Cluster annotations for each cell were shown above the heat map. (e, f) in

453   vitro cytokine stimulation of primary alveolar fibroblasts. (e) Schematic of the experiment. (f) qPCR

454   analysis for representative genes. n = 3 wells. Data are mean ± SEM. Data are representative of three

455   experiments. (g) Schematic of re-analysis of our previous human scRNA-seq data. (h) UMAP plot

after subsetting and re-clustering alveolar and pathologic clusters. (i) UMAP plots for cells from control (n = 3), scleroderma (n = 2) or IPF (n = 3) lungs. (j) Dot plot for representative markers for each subset. (k) UMAP plot overlaid with pseudotime. (l) Schematic of in situ hybridization experiment. (m) Dot plot showing *ITGA8* expression can distinguish inflammatory fibroblasts from adventitial fibroblasts. (n) Joint density plots showing two markers that can highlight fibroblast subsets. (o) In situ hybridization on sequential sections from an IPF lung. Yellow squares in the left images are magnified in the right images. Arrows indicate cells expressing the markers for each subset. Yellow dashed circles indicate a fibroblastic focus. Scale bars, 50 μm (merge), 10 μm (magnified). Images are representative of 3 IPF patients. (p) Schematic of localization for fibrotic and two inflammatory fibroblast subsets.

Fig. 4. Cthrc1-CreER mouse demonstrates the pro-fibrotic function of Cthrc1+ fibroblasts

(a) Schematic of Cthrc1-CreER mouse generation. (b) Time course of bleomycin and tamoxifen treatment for day 14 analysis. (c) Maximum projection of whole lung imaging. (d) qPCR analysis of all lung cells, lineage− cells, and tdTomato+ cells. n = 3 mice. (e) Collagen 1 and collagen 4 staining on day 14. Scube2-CreER mice were treated with tamoxifen as shown in Fig. 2a. Arrowheads indicate fibroblasts inside the basal lamina. Arrows indicate fibroblasts outside of the basal lamina. (f) Image quantification of mean distances between tdTomato and collagen 4. (g) Image quantification of tdTomato+ cells that are directly associated with intra-alveolar collagen 1. n= 4 mice (f, g). (h) Time course of bleomycin and tamoxifen treatment for day 28 analysis. (i, j) Ablation efficiency assessed by lineage− tdTomato+ cell number in left lobes (i) or by whole lung qPCR for *Cthrc1* (j) from Cthrc1-CreER+/- Rosa26-tdTomato/WT (n = 11 mice) or Cthrc1-CreER+/- Rosa26-tdTomato/DTA

479   (n = 12 mice). (k) Hydroxyproline assay on day 28 of Cthrc1-CreER+/- Rosa26-WT/WT (n = 16

480   mice for saline, n = 26 mice for bleomycin) or Cthrc1-CreER+/- Rosa26-DTA/DTA (n = 17 mice for

481   saline, n = 31 mice for bleomycin). Scale bars, 1 mm (c), 20 μm (e). Data are representative of at least

482   two independent experiments except (k), which is a pool from two independent experiments. Data

483   are mean ± SEM. Statistical analysis was performed using Tukey's multiple comparisons test after

484   one-way ANOVA (f), unpaired two-tailed t-test (g), or two-tailed Mann-Whitney test (i, j, k).

485

486

487   **Fig. 5. Tgfbr2 conditional knockout in alveolar fibroblasts abrogates fibrosis but exacerbates**

488   **inflammation**

489   (a) Schematic of *Tgfbr2* conditional knockout experiments. (b) Sections after bleomycin treatment

490   were stained for collagen 1 and collagen 4. Arrows indicate intra-alveolar collagen 1. Scale bars, 100

491   μm. (c) Quantification of Collagen 1+ area of whole sections. n = 3 (saline) or 6 (bleomycin) mice.

492   (d) Hydroxyproline assays of left lobes. n = 5 (saline), 15 (bleomycin, control), or 13 (bleomycin,

493   *Tgfbr2* fl/fl) mice. (e) Percent body weight change after bleomycin. n = 15 (bleomycin, control) or 13

494   (bleomycin, *Tgfbr2* fl/fl) mice. (f) qPCR of purified tdTomato+ cells after bleomycin showing the

495   relative expression to the control group. n = 13 (bleomycin, control) or 11 (bleomycin, *Tgfbr2* fl/fl)

496   mice. (g) ELISA of BAL for IgM. n = 5 (saline), 13 (bleomycin, control), or 11 (bleomycin, Tgfbr2

497   fl/fl) mice. (h) Flow cytometric counting of myeloid populations in BAL. n = 13 (bleomycin, control),

498   or 10 (bleomycin, *Tgfbr2* fl/fl) mice. (i) Schematic of sequential differentiation of alveolar fibroblast

499   lineage after injury. Data are representative of at least two independent experiments. Data are mean

500   ± SEM. Statistical analysis was performed using two-tailed Mann-Whitney test (c, d, g, h) or unpaired

501 two-tailed t-test (f). P-values were adjusted using Holm–Sidak's multiple-comparisons adjustment (f,

502 h).

503
504

505

506 **Methods**

507 Mice

508 Rosa26-lox-stop-lox-tdTomato (Stock No. 007914), Rosa26-lox-stop-lox-DTA (Stock No. 009669),

509 and *Tgfbr2* fl/fl (Stock No. 012603) mice were obtained from the Jackson Laboratory. Col-GFP mice

510 were obtained from Dr. David Brenner at University of California, San Diego[35]. Mice between the

511 ages of 8 and 16 weeks old were used for the experiments. Male mice were used for the Scube2-

512 CreER scRNA-seq experiment. Both male and female mice were used in the other experiments.

513 Heterozygous Cthrc1-CreER mice were used for experiments to avoid potential impact on fibrosis by

514 altered Cthrc1 expression[31]. Homozygous Scube2-CreER mice were used for experiments to achieve

515 higher recombination efficiency. No obvious phenotype of lung structure or fibrosis was observed in

516 homozygous Scube2-CreER mice. Mice with homozygous Rosa26-tdTomato or Rosa26-DTA alleles

517 were used for experiments unless specified. For fibrosis induction, mice were treated with bleomycin

518 in 75 µl saline by oropharyngeal aspiration. Since male mice develop more severe fibrosis[36], we used

519 2.5 U/kg bleomycin for male mice and 3 U/kg bleomycin for female mice, which were determined

520 by induction of 7-9% body weight loss on day 7 and approximately 10% mortality rate. Male and

521 female mice showed similar degrees of fibrosis measured by hydroxyproline with these doses. For

522 Scube2-CreER *Tgfbr2* fl/fl mice, we used 1.7 U/kg for male mice and 2 U/kg for female mice to avoid

523 mortality due to the exaggerated inflammation during the early phase after bleomycin treatment. For

524 the silica-induced lung fibrosis model, silica (MIN-U-SIL5, US Silica) was heated in 1N hydrochloric

23

525 acid at 110 °C for 1 hour. Silica was then washed with sterile water twice, followed by drying at

526 110 °C overnight. Dried silica was resuspended in saline and 400 mg/kg body weight silica was

527 intratracheally instilled by oropharyngeal aspiration. Tamoxifen (Millipore Sigma) was dissolved in

528 olive oil (Millipore Sigma) at 20 mg/ml, and 2 mg was intraperitoneally injected once a day. For

529 labeling Cthrc1+ cells, tamoxifen was injected on days 8-12 after bleomycin treatment in most

530 experiments. For ablating Cthrc1+ cells, tamoxifen was injected on days 8, 9, 11, 12, 15, 16, 18, 19,

531 22, 23, 25, and 26 after bleomycin treatment. Scube2-CreER mice were treated with tamoxifen for 2

532 weeks and used for experiments at least 1 week after the last tamoxifen injection unless specified.

533 BAL was collected with 800 μl PBS from right lungs after dissecting left lungs for dissociation. IgM

534 concentration in BAL was measured using the Mouse IgM ELISA Kit (Millipore Sigma). For IL-17

535 blocking experiments, mice intraperitoneally received 15 μg of anti-IL-17a neutralizing antibody

536 (R&D, MAB421) on day 1 after bleomycin treatment and subsequently received 7.5 μg of anti-IL-

537 17a neutralizing antibody on days 3, 5, 7, 9, and 11 after bleomycin treatment. Mice were maintained

538 in the UCSF specific pathogen-free animal facility in accordance with guidelines established by the

539 Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. All animal

540 experiments were in accordance with protocols approved by the University of California, San

541 Francisco Institutional Animal Care and Use Committee.

542

543 Generation of Cthrc1-CreER and Scube2-CreER mice

544 The Cthrc1-CreER mouse strain was generated by homology-directed repair at the endogenous

545 Cthrc1 locus aided by CRISPR/Cas9 endonuclease activity in C57BL/6 mice. Briefly, target sequence

546 (5'-atatattggaatgccattac-3'), which had an adjacent PAM sequence, for guide RNA was selected to

547 induce   double-strand   breaks   within   the   3'UTR,   and   crRNA   with   input   sequence

548 GTAATGGCATTCCAATATAT and tracrRNA were obtained from IDT. A 2.38kb 5'homology arm

549 was amplified from C57BL/6 mouse genomic DNA with forward primer 5'-

550 GAGCTGAATGTTCAGGACCTCTTC-3' and reverse primer 5'-

551 TTTCGGTAGTTCTTCAATGATGAT-3'. A 2.15kb 3' homology arm was amplified with forward

552 primer 5'-CATTACAGTATTTAGTATTTCCTTCT-3' and reverse primer 5'-

553 ATTTGTTTGTTCCTAGGAGCTCTATAC-3'. A targeting vector with P2A-CreERT2-T2A-GFP-

554 stop codon-rabbit beta globin polyA sequence flanked by 5' and 3' homology arms was generated

555 using NEBuilder HiFi DNA Assembly (NEB) and cloned into a pKO2 backbone plasmid. The

556 targeting vector was linearized at SalI (NEB) and NotI (NEB) sites flanking the donor DNA sequence

557 and the linearized donor DNA was purified by agarose gel electrophoresis with GeneJet Gel

558 Extraction kit (Thermo Fisher). Linearized donor DNA and CRISPR/Cas9 complex were injected

559 into C57BL/6 fertilized zygotes, which were then implanted into the oviducts of pseudopregnant

560 female mice. 215 embryos were implanted and 17 pups were born. Three founders were identified by

561 genotyping. We used one founder to expand the colony. The Scube2-CreER mouse strain was

562 generated by a similar process, inserting P2A-CreERT2-rabbit beta globin polyA into the 3'UTR of

563 endogenous *Scube2* locus. The target sequence of guide RNA for endogenous *Scube2* 3'UTR locus

564 was 5'-GTGACTCGTCAGAGTTCAGT-3' and the input sequence for crRNA was

565 ACTGAACTCTGACGAGTCAC. A 2.79 kb 5'homology arm was amplified with forward primer

566 5'-TGGCCTTGACTGTGTACACTTACATTA-3' and reverse primer 5'-

567 TTTGTAAGGCCTCAGAAACCTTGACACTTT-3'. A 2.24kb 3'homology arm was amplified with

568 forward primer 5'-TTTTATAGACAATACAGATATCTTGA-3' and reverse primer 5'-

569 TGTGTGAGAATACATGTGTACCACA-3'. A targeting vector with P2A-CreERT2-stop codon-

570 rabbit beta globin polyA sequence flanked by 5' and 3' homology arms was generated and linearized

571 for injection. 220 embryos were implanted and 22 pups were born. 5 founders were identified and we

572 used one of them to expand the colony.

573

574 Tissue dissociation

575 Mouse lungs were harvested after perfusion through the right ventricle with 5 ml PBS. After mincing

576 with scissors, the tissue was suspended in protease solution [0.25 % Collagenase A (Millipore Sigma),

577 1 U/ml Dispase II (Millipore Sigma), 2000 U/ml Dnase I (Millipore Sigma) in Hanks' Balanced Salt

578 Solution (Thermo Fisher)], except the experiments for analyzing IL-17 production from lymphocytes

579 where lungs were dissociated without Dispase II to avoid degradation of some surface markers for

580 immune cells. The suspension was incubated at 37°C for 60 min with trituration by micropipette

581 every 20 min. Then the cells were passed through a 70 μm cell strainer (BD Biosciences), washed

582 with PBS, and suspended in PBS with 0.5 % bovine serum albumin (BSA) (Fisher BioReagents).

583

584 Flow cytometry

585 After tissue dissociation, 1 x 10^6 cells were used for flow cytometry. Cells were resuspended in PBS

586 with 0.5% BSA-containing antibodies. For identifying lineage+ cells, cells were first stained with

587 biotin-labeled antibodies for lineage markers, followed by washing and staining with other antibodies

588 and streptavidin-A488 or APC/Cy7. DAPI (Thermo Fisher) was used at 0.1 μg/ml to identify dead

589 cells. Flow cytometric cell count was performed using CountBright Plus Absolute Counting Beads

590 (Invitrogen). The following antibodies were used at 1:200 unless specified: anti-CD9 (clone MZ3,

591 FITC, APC/Fire750, biotin; BioLegend), anti-CD31 (clone 390, A488, biotin; BioLegend), anti-

592 CD45 (clone 30F-11, PE/Cy7, BV421, biotin; BioLegend), anti-Mcam (clone ME-9F1, biotin;

593　BioLegend), anti-Pdgfra (clone APA5, APC; BioLegend), anti-Epcam (clone G8.8, PE, APC/Cy7,

594　biotin; BioLegend), anti-I-A/I-E (MHC class II) (clone M5/114.15.2, APC/Cy7; BioLegend), anti-

595　Sca1 (clone D7, PE/Cy7, biotin; BioLegend), anti-CD3 (clone 17A2, PE/Cy7; BioLegend), anti-CD4

596　(clone GK1.5, FITC; BioLegend), anti-B220 (clone RA3-6B2, FITC; BioLegend), anti-CD11b

597　(clone M1/70, Alexa 488; BD Pharmingen), anti-CD11c (clone N418, BV786; BD Bioscience), anti-

598　Ly6c (HK1.4, APC; BioLegend), anti-Ly6g (clone 1A8, APC/Cy7, BV605; BioLegend), anti-SiglecF

599　(clone S17007L, PE/Cy7; BioLegend) anti-CD90.2 (clone 30-H12, BV605; BioLegend), anti-TCR β

600　chain (clone H57-597, BV650; BioLegend), anti-TCR γ/δ (clone GL3, APC/Fire 750; BioLegend),

601　anti-IL-17a (clone TC11-18H10.1, APC; BioLegend), streptavidin-APC/Cy7 (BioLegend),

602　streptavidin-A488 (1:1000, Thermo Fisher). Data acquisition or cell sorting was performed using

603　FACS Aria III or Aria Fusion (BD Biosciences) using BD FACSDIVA Sfotware Version 8.0.1.1.

604　Flow cytometry data were analyzed using FlowJo v10 (Becton Dickinson).

605

606　Hydroxyproline assay

607　Fibrosis after bleomycin treatment was assessed by hydroxyproline assay of tissue lysates as

608　described previously[37]. Briefly, left lobes were homogenized and precipitated with trichloroacetic

609　acid. Following baking at 110°C overnight in HCl, samples were reconstituted in 1 ml water, and

610　hydroxyproline content was measured by a colorimetric chloramine T assay.

611

612　scRNA-seq library preparation and sequencing

613　Scube2-CreER, Rosa26-tdTomato double homozygous mice were treated with tamoxifen for two

614　weeks. Bleomycin treatment was performed 2 weeks after the last tamoxifen treatment. Three

27

biological replicates from day 0 (non-bleomycin-treated), 7, 14, and 21 samples were collected on the same day, and tamoxifen/bleomycin treatment was scheduled accordingly. After harvesting and dissociating left lobes, mesenchymal cells were enriched by magnetic negative selection with anti-CD31, CD45, EpCAM, and Ter119-biotin antibodies (1:200) and Dynabeads MyOne Streptavidin T1 (40 μl / sample, Invitrogen). After magnetic negative selection, cells were stained with Streptavidin-APC/Cy7 (1:200) and DAPI (0.1 μg/ml). Approximately 2 x 10^5 lineage-APC/Cy7-negative cells were sorted for each sample. The sorted cells were counted and labeled with oligonucleotide tags for multiplexing using 10x Genomics 3' CellPlex Kit Set A. Tag assignment was as follows; day 0 (301, 302, 303), day 7 (304, 305, 306), day 14 (307, 308, 309), and day 21 (310, 311, 312). All 12 samples were pooled and 30,000 cells/lane were loaded onto 4 lanes of Chromium Next GEM Chip (10x Genomics). Chromium Single Cell 3' v3.1 (10x Genomics) reagents were used for library preparation according to the manufacturer's protocol. The libraries were sequenced on an Illumina NovaSeq 6000 S4 flow cell.

Sequencing data processing

Fastq files were uploaded to the 10x Genomics Cloud Analysis website (https://www.10xgenomics.com/products/cloud-analysis) and reads were aligned to a custom reference of mouse genome mm10 with tdTomato-WPRE-polyA transcript sequence using Cell Ranger version 6.1.1. tdTomato-WPRE-polyA sequence was obtained from the sequence of the targeting vector for the Ai9 mouse (Addgene plasmid #22799) since the Ai9 mouse shares the same sequence for tdTomato-WPRE-polyA with the Ai14 mouse used in this study[34]. The data were demultiplexed and multiplets identified by the presence of multiple oligonucleotide tags were removed using the 10x Genomics cloud analysis function with default parameters. Raw count

638   matrices were imported to the R package Seurat v4.1.0[38] and cells with fewer than 200 detected genes,

639   larger than 7500 detected genes, or larger than 15% percent mitochondria genes were excluded. We

640   used the DoubletFinder package[39] for individual samples to remove doublets that were not detected

641   upon alignment using an estimated multiple rate of 2%. We then merged all the sample objects,

642   identified the top variable genes using the Seurat implementation *FindVariableGenes*, and integrated

643   the samples using the *RunFastMNN*[40] function of the SeuratWrappers R package. For visualization,

644   the *RunUMAP* function of Seurat was performed using MNN dimensional reduction. Nineteen

645   clusters were initially identified using *FindNeighbors* and *FindClusters* functions of Seurat with

646   resolution = 0.8 from a total of 47,809 cells. Cluster 17 (168 cells) was a cluster mixed with a small

647   number of lineage (CD31, CD45, and EpCAM)+ cells that were not removed by FACS sorting.

648   Cluster 18 (165 cells) showed up in two different locations on the UMAP embedding, one close to

649   alveolar fibroblasts and the other close to peribronchial fibroblasts. Cluster 18 cells expressed both

650   alveolar and peribronchial fibroblast markers, suggesting that they were doublets that were not

651   removed by prior processing. We excluded clusters 17 and 18, and re-clustered the remaining 47,476

652   cells with *FindVariableGenes, RunFastMNN, RunUMAP, FindNeighbors,* and *FindClusters*

653   functions with clustering resolution = 0.3. Differentially expressed genes for each cluster were

654   identified using the *FindAllMarkers* function of Seurat focusing on genes expressed by more than

655   25% of cells (either within or outside of a cluster) and with a log fold change greater than 0.25.

656   tdTomato+ cells were defined by natural log-normalized tdTomato expression level greater than 3.5.

657   The metadata including cluster, sample, and tdTomato+ annotations was exported for quantifying the

658   tdTomato+ frequency in each cluster. Gene ontology enrichment analysis for the differentially

659   expressed genes was performed using DAVID (Database for Annotation, Visualization, and

660   Integrated Discovery) Bioinformatic Resources software version 2021, or using one-sided Fisher's

661 exact tests implemented in gsfisher R package (https://github.com/sansomlab/gsfisher/). We

662 performed pseudotime analysis on the UMAP embeddings using Monocle 3 v1.0.0[19], specifying cells

663 on day 0 as roots of the pseudotime. Scaled expression of representative markers along the pseudotime

664 was visualized using ggplot2 v3.3.6 (Fig. 3c). A heatmap with cells arranged in pseudotemporal

665 ordering (Fig. 3d) was generated using Slingshot v2.2.0[41] and ComplexHeatmap R package 2.10.0,

666 specifying starting cluster as "Alveolar" and ending cluster as "Fibrotic".

667

668 Human scRNA-seq data processing

669 We used our previous human scRNA-seq data set of pulmonary fibrosis (GSE132771)[2]. We subsetted

670 alveolar and pathologic fibroblast clusters from our mesenchymal cell data, and re-clustered them

671 using *FindVariableGenes, RunFastMNN, RunUMAP, FindNeighbors,* and *FindClusters* functions of

672 Seurat with clustering resolution = 0.3. Cluster markers were identified using the *FindAllMarkers*

673 function of Seurat with min.pct = 0.25 and logfc.threshold = 0.25. For comparison between human

674 and mouse emergent clusters, the average expression of the clusters was exported from scaled data

675 of Seurat objects, and human genes were converted to mouse orthologs using the biomaRt R package,

676 followed by calculation of Spearman's correlation coefficient by *cor* function of R. Pseudotime

677 analysis was performed on the UMAP embeddings using Monocle 3, specifying cells from control

678 lungs as roots of the pseudotime. Joint density plots were generated using the scCustomize package

679 [42]. For integrating alveolar and pathologic fibroblasts from Adams et al.[23] and Habermann et al.[24], we

680 obtained their data sets from GSE147066 and GSE135893, respectively. The raw count matrix of

681 mesenchymal cells of control and IPF lungs from Adams et al. was batch-corrected using the

682 *RunFastMNN* function of Seurat, and visualized by *RunUMAP, FindNeighbors,* and *FindClusters*

683 functions of Seurat. Alveolar and pathologic fibroblast clusters were identified by examining markers

30

684 such as *INMT, NPNT, TCF21, CTHRC1, COL1A1*, and *POSTN*. For Habermann et al., mesenchymal

685 cells annotated by the original authors were subsetted from the Seurat object containing all cells. By

686 examining markers for alveolar and pathologic fibroblasts, a cluster the authors annotated as

687 myofibroblasts was identified as cells containing both alveolar and pathologic fibroblasts. We

688 subsetted those alveolar and pathologic fibroblast clusters from Adams et al. and Habermann et al.

689 data sets, merged all with our alveolar and pathologic fibroblast clusters, and integrated these data

690 sets using the *RunFastMNN* function of Seurat by splitting the object by individual patients or donors.

691 After UMAP visualization and clustering, there were two minor clusters of which cells originated

692 only from Adams et al.. One of these clusters was characterized by unusually high numbers of genes

693 and read counts. The other cluster was characterized by high mitochondrial gene proportions. Since

694 these two clusters were only seen in Adams et al. and seemed to be driven by technical artifacts but

695 the other clusters from Adams et al. merged well with the other two data sets, we excluded these two

696 clusters. We re-clustered the remaining cells and annotated the clusters based on the overlap with

697 cells from our data set, which had transferred cluster annotations as shown in Fig. 3h.

698

699 Histology, immunohistochemistry, and imaging

700 For histology, lungs were inflated with 4% PFA and immersed in 4% PFA overnight at 4°C. The

701 lungs were then immersed in 30% sucrose for 24 hours at room temperature, and then embedded in

702 OCT. 12 μm sections for thin section histology or 100 μm sections for thick section histology were

703 made using a cryostat CM1850 (Leica). Thin sections were attached to Superfrost Plus microscope

704 slides (Fisher). For Sirius Red staining, sections were incubated with 0.1% Sirius Red in Saturated

705 Picric Acid (Electron Microscopy Sciences) with 0.125% w/v Fast Green FCF (Fisher) for 1 hour,

706 except Extended Data Fig. 6c, d, which were stained with Sirius red without Fast Green FCF. Thick

707 sections were processed as floating sections in buffers. Thick sections were cleared using a CUBIC

708 method[43]. Sections were stained with anti-Pi16 (5 µg/ml, R&D, AF4929), anti-proSP-C (1:5000,

709 Sigma-Aldrich, AB3786) anti-collagen 1 (1:200, Southern Biotech, 1310-01), anti-collagen 4 (1:5000,

710 LSL, LSL-LB-1403), anti-Pdgfra (5 µg/ml, R&D, AF1062), anti-CD68 (5 µg /ml, R&D, MAB10114)

711 or anti-Saa3 (1:100, Abcam, JOR110A) followed by donkey anti-rabbit IgG-Alexa 488 or 647

712 (1:1000, Thermo Fisher, A-21206, A-21245), donkey anti-goat IgG-Alexa 647 (1:1000, Thermo

713 Fisher, A-21447), or donkey anti-rat IgG-Alexa 647 (1:1000, Thermo Fisher, A78947). Thick

714 sections were then treated with CUBIC-R+(M) (TCI), placed in a well of glass bottom plate with

715 sections covered with CUBIC-R+(M), and imaged by an inverted Crest LFOV spinning disk confocal

716 microscope (Nikon Ti2). Images were processed using Image J version 1.53q. Distance to collagen 1

717 or Pi16 from tdTomato was measured by the "Co-localization by Cross-Correlation" plugin of Image

718 J. For collagen 1+ area quantification, we first iteratively selected two auto threshold methods of

719 Image J for the collagen 1 channel, which can distinguish background autofluorescence and antibody-

720 specific collagen 1 signal. An auto threshold method "Mean" highlighted all of the tissue areas

721 including autofluorescence of tissue. We generated a binary image with the "Mean" method to mask

722 all of the tissue areas that excluded airspaces and empty spaces for quantification. We then applied

723 an auto threshold method "Triangle", which highlighted antibody-specific collagen 1+ area, to the

724 masked original images to measure % collagen 1+ area. We measured the % collagen 1+ area from

725 images of whole sections for each mouse. For Saa3 and CD68 area quantification, approximately 1

726 mm x 1 mm alveolar areas excluding large airways and visceral pleura were selected. % area of the

727 signal was measured after applying the "Triangle" auto threshold on Image J. 3D reconstruction of z-

728 stack images was performed using Icy version 2.0. For whole lung imaging, 4% PFA-fixed lungs

729 were cleared with CUBIC-L and treated with CUBIC-R+(M), followed by imaging with Mounting

730   Solution (RI 1.520, TCI) using a Nikon AZ100 microscope configured for light sheet microscopy.

731   Autofluorescence signal in the GFP channel was used to visualize the lung structure except Fig. 1d.

732   Maximum projection images were generated using Image J.

733

734   Human Lung Tissue

735   The studies described in this paper were conducted according to the principles of the Declaration of

736   Helsinki. Written informed consent was obtained from all subjects, and the study was approved by

737   the University of California, San Francisco Institutional Review Board. Fibrotic lung tissues were

738   obtained at the time of lung transplantation from patients with a diagnosis of usual interstitial

739   pneumonia.

740

741   In situ hybridization

742   In situ hybridization was performed by using the RNAscope Multiplex Fluorescent Reagent kit v2

743   (ACD). Explanted IPF tissues were fixed with 4% PFA overnight and cryoprotected with 30% sucrose

744   for 24 hours, followed by embedding in OCT for frozen blocks. 12 μm sections were attached to

745   Superfrost Plus microscope slides. Target retrieval was performed by heating in a steamer for 15 min.

746   After target retrieval, we photochemically bleached autofluorescence by exposing the sections to LED

747   light as previously described[44]. Sections were then treated with Protease III (ACD), followed by in

748   situ hybridization according to the manufacturer's protocol. Probe channels used are as follows:

749   COL1A1-C1 (TSA Vivid 650), CTHRC1-C3 (TSA Vivid 570), SFRP2-C1 (TSA Vivid 650), CCL2-

750   C2 (TSA Vivid 570), SFRP4-C1 (TSA Vivid 650), CXCL14-C2 (TSA Vivid 570), ITGA8-C3 (TSA

751   Vivid 520).

752

Quantitative Real-Time PCR analysis

Approximately 2000 cells were directly sorted into TRIzol reagent (Thermo Fisher), and RNA was

isolated according to the manufacturer's protocol. The RNA was reverse-transcribed using a Super

Script IV VILO Master Mix with ezDNase Enzyme kit (Thermo Fisher). Quantitative Real-Time

PCR was performed using PowerUp SYBR Green Master Mix (Thermo Fisher) with a Quant Studio

4 (Applied Biosystems). Primer sequences are listed in Supplementary Table 1.

759

In vitro stimulation of primary alveolar fibroblasts

Alveolar fibroblasts were isolated by magnetic negative selection for CD31, CD45, EpCAM, Ter119,

Sca1, and CD9. 2 x 10^5 cells were seeded into 48 well plates and initially cultured in DMEM

(Corning) with 2% FBS (Gibco) and 1% penicillin-streptomycin (Gibco) for 24 hr. Then the medium

was changed to serum-free DMEM with 1% penicillin/streptomycin for 24 hr. After serum starvation,

the medium was changed to serum-free DMEM with 1% penicillin/streptomycin, containing 1 ng/ml

IL-1β (R&D, 401-ML), 1 ng/ml TGF-β1 (R&D, 7754-BH), or 10 ng/ml TNF-α (R&D, 210-TA). For

sequential stimulation, the medium was changed to serum-free DMEM with 1%

penicillin/streptomycin containing 1 ng/ml IL-1β, 1 ng/ml TGF-β1, or both for 24 hr. After the

cytokine stimulations, cells were lysed by directly adding 400 μl Trizol into the wells. Cell culture

was performed under standard conditions (37 °C, 5% CO_2).

771

Data analysis

773　The mean linear intercept was quantified as described previously[45]. scRNA-seq data analysis was

774　performed in R version 4.1.3. Statistical tests were performed in GraphPad Prism version 9.4.0.

775

776　Data availability

777　The scRNA-seq data generated in this study are deposited in Gene Expression Omnibus (GEO)

778　(GSE210341). Human scRNA-seq data analyses were performed using publicly available data,

779　GSE132771, GSE147066, and GSE135893.

780

781　Code availability

782　The codes used in the scRNA-seq analysis are available on GitHub

783　(https://github.com/TatsuyaTsukui/AlveolarLineage)

784
785

786　**Method Reference**

787　35. Yata, Y. *et al.* DNase I-hypersensitive sites enhance alpha1(I) collagen gene expression in hepatic

788　　stellate cells. *Hepatology* **37**, 267–276 (2003).

789　36. Redente, E. F. *et al.* Age and sex dimorphisms contribute to the severity of bleomycin-induced lung

790　　injury and fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **301**,

791　　L510–L518 (2011).

792　37. Henderson, N. C. *et al.* Targeting of αv integrin identifies a core molecular pathway that regulates

793　　fibrosis in several organs. *Nat Med* **19**, 1617–1624 (2013).

794　38. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

795　39. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA

796　　Sequencing Data Using Artificial Nearest Neighbors. *cels* **8**, 329-337.e4 (2019).

797   40. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-

798   sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36**, 421–427

799   (2018).

800   41. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC*

801   *Genomics* **19**, 477 (2018).

802   42. Marsh, S., Salmon, M. & Hoffman, P. samuel-marsh/scCustomize: Version 1.1.1. Zenodo

803   https://doi.org/10.5281/zenodo.7534950 (2023).

804   43. Susaki, E. A. *et al.* Versatile whole-organ/body staining and imaging based on electrolyte-gel properties

805   of biological tissues. *Nat Commun* **11**, 1982 (2020).

806   44. Meeker, A. K., Heaphy, C. M., Davis, C. M., Roy, S. & Platz, E. A. Photochemical pre-bleaching of

807   formalin-fixed archival prostate tissues significantly reduces autofluorescence to facilitate multiplex

808   immunofluorescence staining. 2021.11.09.467916 Preprint at https://doi.org/10.1101/2021.11.09.467916

809   (2021).

810   45. Crowley, G. *et al.* Quantitative lung morphology: semi-automated measurement of mean linear intercept.

811   *BMC Pulm Med* **19**, 206 (2019).

812

Acknowledgments

821

826

827   **Author contributions**

828   T.T. and D.S. conceived the study, interpreted the data, and wrote the manuscript. T.T. performed

829   and analyzed the experiments. P.J.W. procured human samples. D.S. supervised the study.

830

831   **Competing interests**

832   D.S. is a founder of Pliant Therapeutics and has received research funding from Abbvie, Pfizer, and

833   Pliant Therapeutics. D.S. serves on the Scientific Review Board for Genentech and on the

834   Inflammation Scientific Advisory Board for Amgen. P.J.W. received research funding from

835   Boehringer Ingelheim, Pliant Therapeutics, and Genentech.

836

837

838   **Extended Data legend**

839   **Extended Data Fig. 1. Scube2-CreER specifically labels alveolar fibroblasts, which provide a**

840   **niche to support AT2 cells**

841   (a) UMAP plots of all lung cells from scRNA-seq data of Tsukui et al. 2020, for cell types (left) and

842   Scube2 expression (right). (b) UMAP plots of Col1a1+ cells from scRNA-seq data of Tsukui et al.

843   2020, for cell types (left) and Scube2 expression (right). (c-f) Gating strategy to evaluate the

844   specificity of tdTomato+ cells for lineage (CD31, CD45, EpCAM, Mcam, Ter119), Sca1, and CD9.

845   (g) Gating strategy for alveolar fibroblasts. (h) Flow cytometric quantification for percent tdTomato+

846   of alveolar fibroblasts. n = 3 mice. (i) Pi16 staining of a lung section from Scube2-CreER/Rosa26-

847   tdTomato Col-GFP mouse. tdTomato is shown in magenta. Col-GFP is shown in green. Pi16 is shown

848   in grey. DAPI is shown in blue. aw, airway. (j) Three representative planes from z-stack images with

849   z-positions are shown below the images. tdTomato is shown in red. proSP-C is shown in cyan.

850   Asterisks indicate the same proSP-C+ cell. Arrows point to projections extending from alveolar

851   fibroblasts. Scale bars, 200 μm (i) or 5 μm (j). Data are mean ± SEM. Data are representative of at

852   least two independent experiments.

853

854   **Extended Data Fig. 2. Scube2-CreER-labeled alveolar fibroblasts provide a niche to support**

855   **AT2 cells**

856   (a, b) Gating strategy for alveolar fibroblast-ablation experiments with Scube2-CreER/Rosa26-DTA

857   mice. (a) Gating strategy for EpCAM+ subpopulations. (b) Gating strategy for fibroblast subsets.

858   Lineage markers include CD31, CD45, EpCAM, and Ter119. (c) Flow cytometric cell count for each

859   population, normalized to means of vehicle groups. n = 4 (tamoxifen) or 5 (vehicle) mice. (d)

860   Representative images of H&E staining of lung sections from alveolar fibroblast-ablation

861   experiments. Scale bars, 200 μm. (e) Quantification of mean linear intercept of alveolar regions. n =

862   4 (tamoxifen) or 5 (vehicle) mice. (f) qPCR analysis of whole lung cells 6 days after bleomycin

863   treatment. n = 5 mice. (g) Gating strategy for CD4 and γδ T cells. (h) Gating strategy for ILCs. (i)

864 Flow cytometric quantification in bleomycin-treated lungs. n = 4 (Rosa26-DTA/DTA) or 7 (Rosa26-

865 WT/WT) mice. (j) Gating strategy for IL-17a+ cells. Data are mean ± SEM. Data are representative

866 of at least two experiments. Statistical analysis was performed using unpaired two-tailed t-test

867 followed by Holm–Sidak's multiple-comparisons adjustment (f) or two-way ANOVA followed by

868 Sidak's multiple comparison test (i).

869

870 **Extended Data Fig. 3. Longitudinal scRNA-seq reveals multiple fibroblast subsets that emerge**

871 **after lung injury**

872 (a) Gating strategy for purifying lineage (CD31, CD45, EpCAM, Ter119)− mesenchymal cells for

873 scRNA-seq. (b) UMAP plots for cells obtained before (day 0) or at various time points after

874 bleomycin treatment. (c) Dot plots showing the top differentially expressed genes for each cluster.

875 (d) GO enrichment analysis by DAVID for differentially expressed genes of inflammatory fibroblasts.

876 (e) GO enrichment analysis by DAVID for differentially expressed genes of stress-activated

877 fibroblasts. (f) GO over-representation analysis with the Fisher test for all clusters.

878

879 **Extended Data Fig. 4. Lineage tracing by scRNA-seq reveals alveolar fibroblasts as the origin**

880 **of multiple pathologic fibroblast subsets**

881 (a) UMAP plot with tdTomato expression. (b) Violin plot for tdTomato shows peaks for tdTomato^{low}

882 or tdTomato^{high} cells. The threshold for tdTomato+ cells was defined as an expression level > 3.5. (c)

883 UMAP plots with tdTomato expression split by biological replicates. (d) Plot showing percent

884 tdTomato+ of alveolar fibroblasts (x-axis) versus percent tdTomato+ of fibrotic fibroblasts (y-axis)

885 for each biological replicate.

886

Extended Data Fig. 5. Scube2-CreER-labeled alveolar fibroblasts differentiate into fibrotic or inflammatory fibroblasts after lung injury

(a) Maximum projection of whole lung imaging for untreated or bleomycin day 14 Scube2-CreER/Rosa26-tdTomato mice. (b) Representative optical sections from whole lung imaging. tdTomato is shown in magenta. Autofluorescence is shown in grey (a, b). (c) Flow cytometry plots showing the increase of CD9+ cells among Scube2-CreER-labeled (tdTomato+) cells on day 21 after bleomycin treatment. (d) Flow cytometric quantification of percent CD9+ of tdTomato+ cells. n = 4 mice. Statistical analysis was performed using unpaired two-tailed t-test. (e) qPCR analysis of sorted cells from Scube2-CreER/Rosa26-tdTomato mice. All lineage (CD31, CD45, EpCAM, Ter119)− tdTomato+ cells (untreated or bleomycin day 21) or lineage− tdTomato+ CD9+ cells (bleomycin day 21) were sorted. The y-axis is the relative expression level to the housekeeping gene Rps3. n = 3 mice. (f) Saa3 staining in sections from Scube2-CreER/Rosa26-tdTomato mice 7 days after bleomycin treatment. Arrows indicate tdTomato and Saa3 double-positive cells. (g) Quantification of percent Saa3+ of tdTomato+ cells. n = 3 mice. tdTomato is shown in red. Saa3 is shown in cyan. DAPI is shown in blue. Collagen 4 is shown in grey. Data are mean ± SEM. Data are representative of at least two independent experiments. Scale bars, 1mm (a, b), 20 μm (f).

903

904

Extended Data Fig. 6. Alveolar fibroblasts up-regulated activation markers and formed silicotic nodules in the silicosis model

(a) Time course of tamoxifen and silica treatment. (b) Representative lung sections of saline or silica-treated mice. (c, d) Sirius red staining of silica-treated lung section imaged as bright field (c) or polarized light (d). (e) Fluorescence imaging of a sequential section of (c, d). (f) Representative images of silicotic nodules. (g) Histological quantification of % tdTomato+ cells of Pdgfra+ cells inside silicotic nodules. n = 5 mice. (h) Flow cytometric analysis of Pdgfra and CD9 on tdTomato+ cells from saline or silica-treated lungs. (i) Flow cytometric quantification of %CD9+ Pdgfra-low cells of tdTomato+ cells. n = 5 mice. (j, k) qPCR analysis of purified populations. The y-axis is the relative expression level to the housekeeping gene Rps3. n = 5 mice. Scale bars, 200 μm (b), 500 μm (c-e), 50 μm (f). Data are mean ± SEM. Data are representative of at least two independent experiments. Statistical analysis was performed using unpaired two-tailed t-test (i, k) or Tukey's multiple comparisons test after one-way ANOVA (j).

Extended Data Fig. 7. TGF-β1 antagonizes inflammatory marker expression induced by IL-1β and induces fibrotic markers

(a) Pseudotime analysis of tdTomato+ clusters suggests that both stress-activated fibroblasts and fibrotic fibroblasts can emerge from inflammatory fibroblasts. (b) UMAP plots of re-clustered alveolar, inflammatory, and fibrotic lineage split by days after bleomycin treatment. (c) Expression of selected markers on UMAP plots. (d) Schematic of sequential cytokine stimulations for primary alveolar fibroblasts. (e) qPCR analysis after sequential cytokine stimulations. Group names indicate (first stimulation) → (second stimulation). DMEM means medium-only control. The y-axis is the relative expression level to the housekeeping gene Rps3. n = 3 wells. Data are mean ± SEM. Data are representative of two independent experiments.

930

931

Extended Data Fig. 8. Re-analysis of our publicly available scRNA-seq data from human pulmonary fibrosis reveals inflammatory and fibrotic clusters

(a) UMAP plots of re-clustered pathologic and alveolar clusters shown for individual patients or donors (control). (b) Dot plot showing top differentially expressed genes for each cluster. (c) GO enrichment analysis by DAVID for differentially expressed genes in inflammatory fibroblast 1 cluster. (d) GO enrichment analysis by DAVID for differentially expressed genes in inflammatory fibroblast 2 cluster. (e) Heat map showing Spearman's correlation coefficients of average gene expression from mouse and human emergent clusters. (f) GO over-representation analysis with the Fisher test for all clusters.

941

942

943

Extended Data Fig. 9. Combined analysis of publicly available scRNA-seq data sets from human pulmonary fibrosis from 3 groups reveals conserved inflammatory and fibrotic clusters

(a) UMAP plot of mesenchymal cells from Adams et al. (b) Expression levels of selected genes on UMAP plot of Adams et al. mesenchymal cells show alveolar and pathologic fibroblast clusters. (c) UMAP plot of mesenchymal cells from Habermann et al. (d) Expression levels of selected genes on UMAP plot of Habermann et al. mesenchymal cells show that "Myofibroblasts" cluster contains alveolar and pathologic fibroblasts. (e) UMAP plot after combining alveolar and pathologic fibroblasts from Adams et al., Habermann et al., and Tsukui et al. (f) UMAP plots of combined data

952 split by original data set. (g) UMAP plots shown for each data set and colored by samples. (h)

953 Expression levels of selected markers for fibrotic (COL1A1, CTHRC1), inflammatory1 (SFRP2,

954 CXCL12), inflammatory2 (SFRP4, CXCL14) and alveolar fibroblasts (NPNT, TCF21) on UMAP

955 plots.

956

957

958 **Extended Data Fig. 10. Cthrc1-CreER mouse demonstrates the pro-fibrotic function of Cthrc1+**

959 **fibroblasts**

960 (a) Gating strategy for cell size, singlet, and live cells. (b) Flow cytometry plots show an increase in

961 lineage− tdTomato+ cells on day 14 after bleomycin treatment. (c) Flow cytometric cell count of

962 lineage (CD31, CD45, EpCAM, Ter119)− tdTomato+ cells on day 14. n = 3 (Saline-Vehicle) or 5

963 (Saline-Tamoxifen, Bleo-Tamoxifen) mice. (d) Flow cytometry plots show CD9 expression on

964 tdTomato+ lineage− cells increases between day 14 and 21. (e) Flow cytometric quantification of

965 percent CD9+ of tdTomato+ lineage− cells. n = 5 mice. (f) Mean fluorescence intensity (MFI) of

966 CD9 on tdTomato+ lineage− cells. n = 5 mice. (g) Saa3 staining in sections from Cthrc1-

967 CreER/Rosa26-tdTomato mice 14 days after bleomycin treatment with tamoxifen injected on days 8

968 -12. tdTomato is shown in red. Saa3 is shown in cyan. Collagen 4 is shown in grey. DAPI is shown

969 in blue. (h) Schematic for localization of inflammatory and fibrotic fibroblasts. (i) Representative

970 images of sequential lung sections from Cthrc1-CreER/Rosa26-tdTomato mice stained for collagen

971 1 or Pi16 (shown in grey). tdTomato is shown in red. DAPI is shown in blue. (j, k) Image

972 quantifications of the mean distance to collagen 1 or Pi16 from tdTomato or DAPI on the sections. n

973 = 4 mice. (l) Representative images of lung sections stained for collagen 1 (shown in grey) from

974 ablation experiments using Cthrc1-CreER/Rosa26-DTA mice. DAPI is shown in blue. (m) Image

975　quantification of % collagen 1+ area on the sections. n = 3 (saline), 8 (bleomycin, Rosa26-WT/WT),

976　or 10 (bleomycin, Rosa26-DTA/DTA) mice. Data are mean ± SEM. Data are representative of at least

977　two independent experiments. Scale bars, 20 μm (g), 1 mm (i, l). Statistical analysis was performed

978　using unpaired two-tailed t-test (j, k) or two-way ANOVA followed by Sidak's multiple comparison

979　test (m).

980

981

982　**Extended Data Fig. 11. Tgfbr2 conditional knockout in alveolar fibroblasts abrogates fibrosis**

983　**but exacerbates inflammation**

984　(a) Representative images of whole sections stained for collagen 1 (magenta) and collagen 4 (green).

985　Arrows indicate regions of intra-alveolar collagen 1. (b) Survival after bleomycin treatment. n = 15

986　mice. (c) qPCR of purified tdTomato+ cells from saline-treated mice showed no difference for fibrotic

987　and inflammatory genes between control and Tgfbr2 cKO. (d) Flow cytometric counting of myeloid

988　populations in BAL from saline-treated mice showed no difference between control and Tgfbr2 cKO.

989　n = 5 mice (c, b). (e) Gating strategy for myeloid populations in BAL. (f) Representative images of

990　sections from bleomycin-treated lungs stained for Saa3 (magenta) and CD68 (green). tdTomato is

991　shown in blue. Magnified single-channel images of yellow rectangles are shown on the right. (g, h)

992　Image quantification of Saa3 (g), or CD68 (h). n = 5 (control) or 6 (Tgfbr2 fl/fl) mice. Data are

993　representative of at least two independent experiments. Data are mean ± SEM. Statistical analysis

994　was performed using unpaired two-tailed t-test (g, h). Scale bars, 1 mm (a), 100 μm (f, wide), or 20

995　μm (f, magnified).

996

997

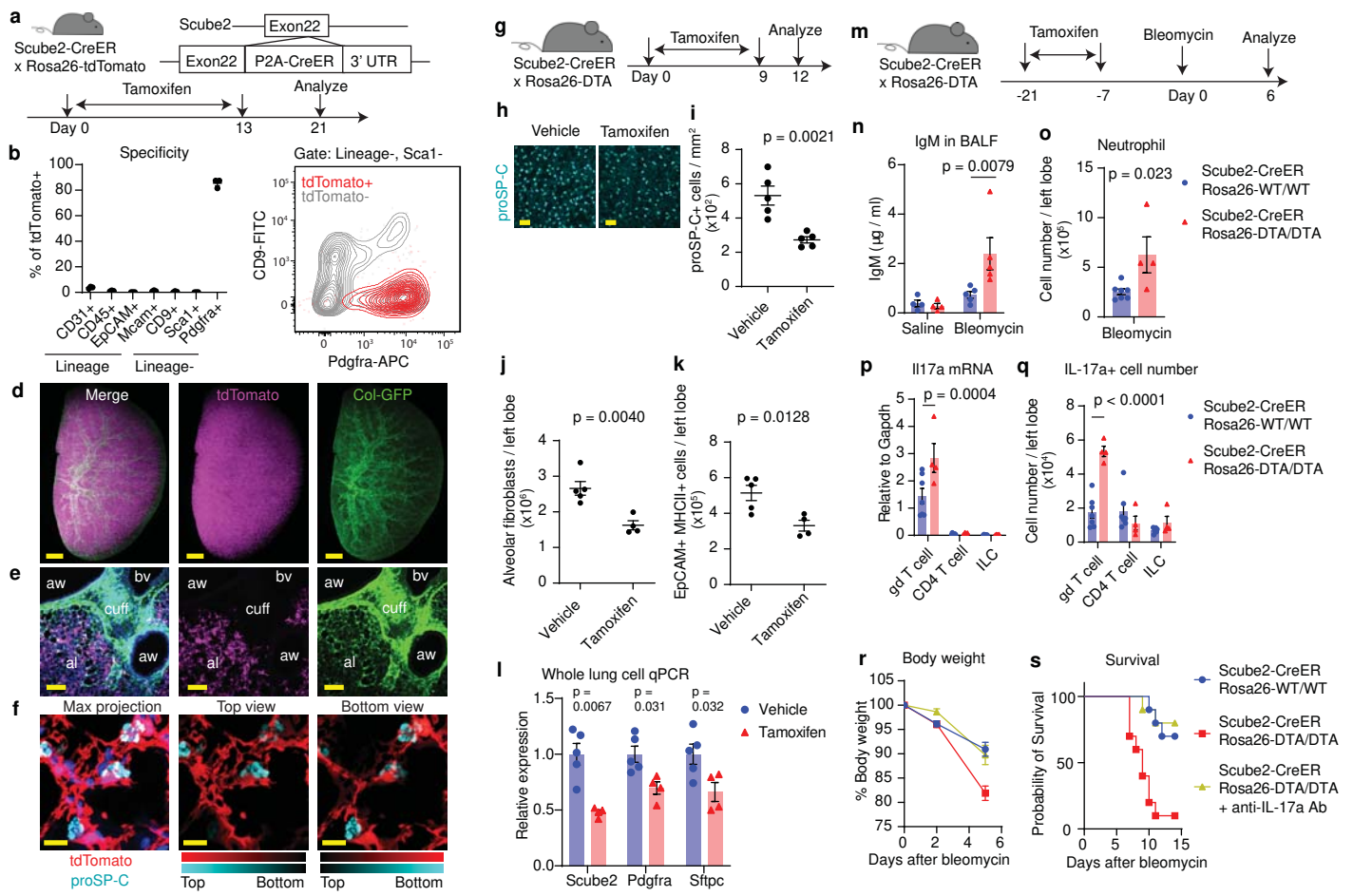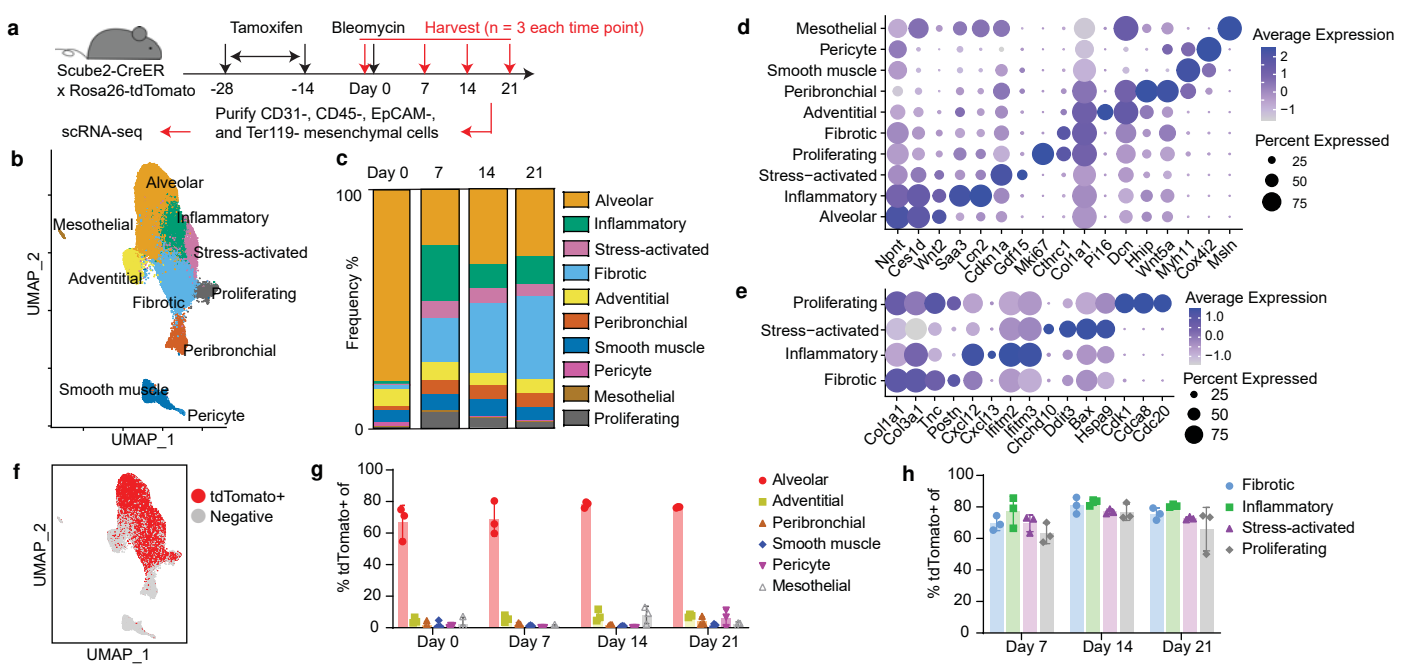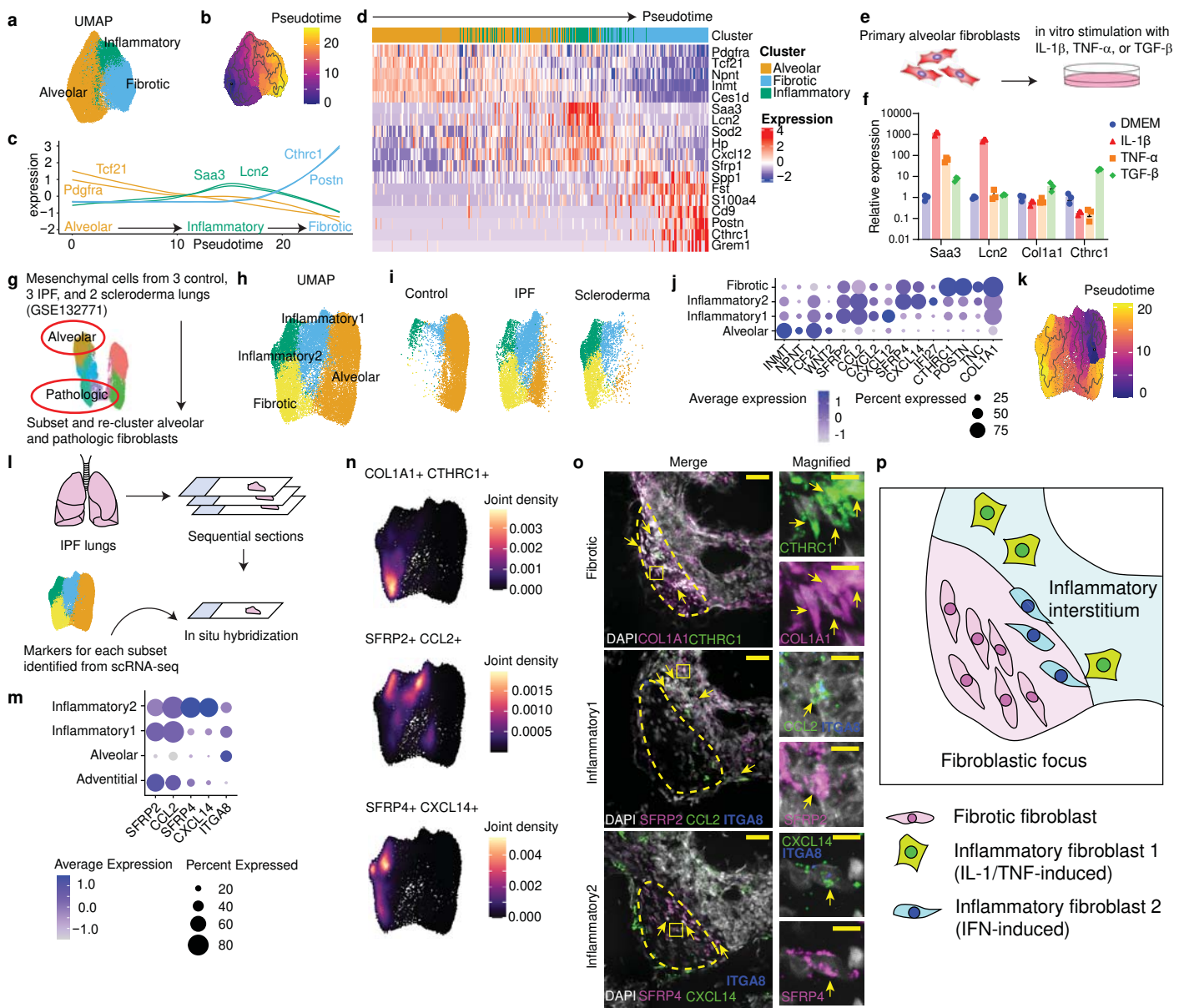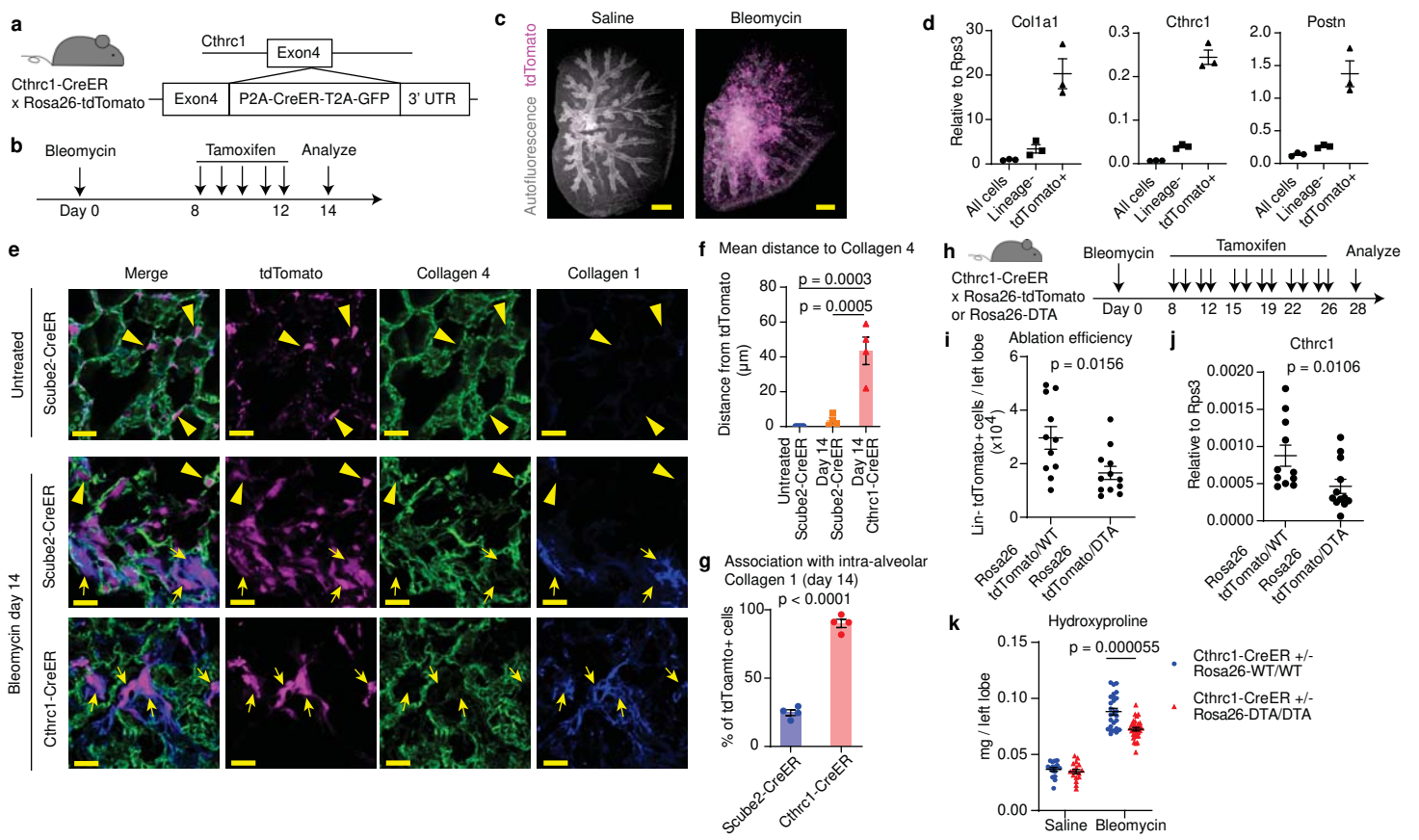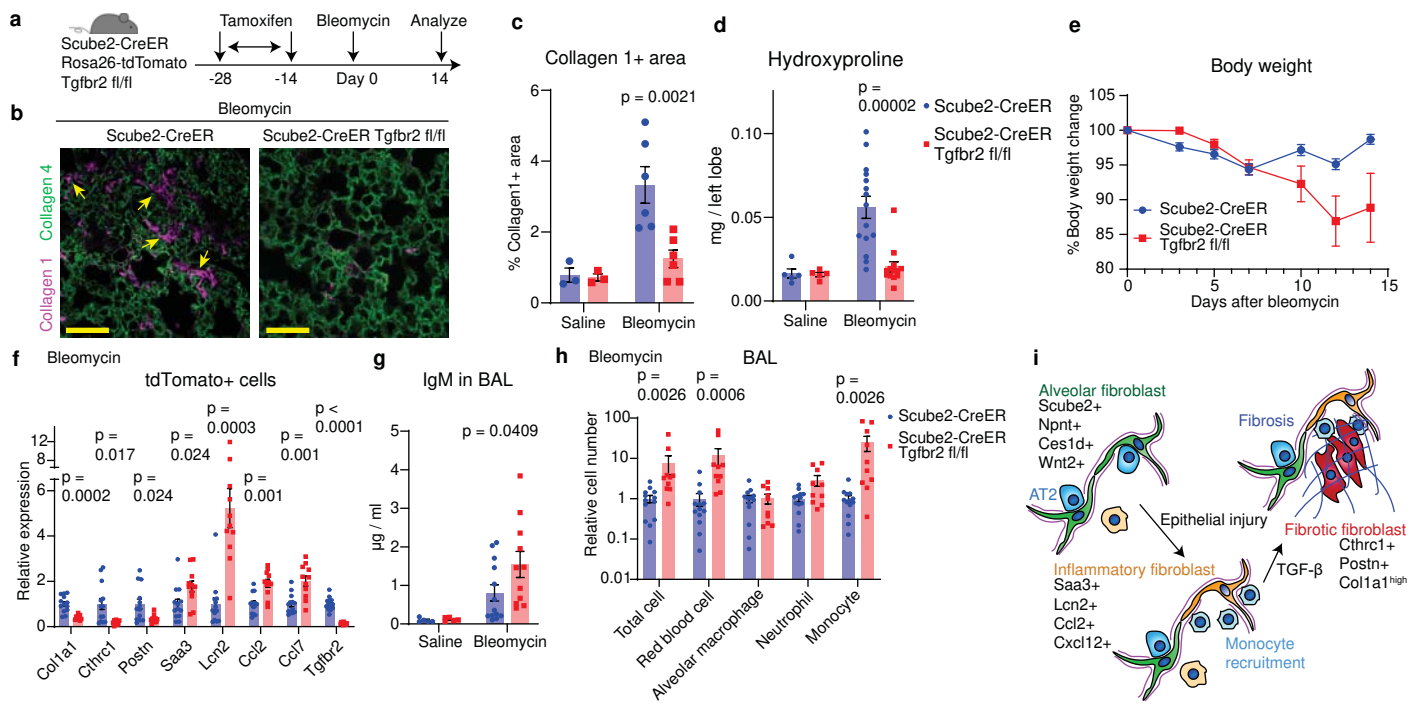**Figure 1**

Figure 2

Figure 3

Figure 4

Figure 5