Persistence of a novel regeneration-associated transitional cell state in pulmonary fibrosis

Yoshihiko Kobayashi[1,#], Aleksandra Tata[1,#], Arvind Konkimalla[1,2], Hiroaki Katsura[1], Rebecca, F. Lee[1], Jianhong Ou[1], Nicholas E. Banovich[3], Jonathan A. Kropski[4,5,6] and Purushothama Rao Tata[1,7,8,*].

1 Department of Cell Biology, Duke University School of Medicine, Durham, NC 27710, USA

2 Medical Scientist Training Program, Duke University School of Medicine, Durham, NC 27710, USA

3 Translational Genomics Research Institute, Phoenix, AZ 85004, USA

4 Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37212, USA

5 Department of Veterans Affairs Medical Center, Nashville, TN 37212, USA

6 Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37212, USA

7 Duke Cancer Institute, Duke University School of Medicine, Durham, NC 27710, USA

8 Regeneration Next, Duke University, Durham, NC 27710, USA

# Denotes co-first author

* Corresponding author: Purushothama Rao Tata, purushothamarao.tata@duke.edu

Abstract

Stem cell senescence is often seen as an age associated pathological state in which cells acquire an abnormal and irreversible state. Here, we show that alveolar stem cell differentiation during lung regeneration involves a unique previously uncharacterized transitional state that exhibits cardinal features normally associated with cell senescence. Specifically, using organoid cultures, multiple *in vivo* injury models coupled with single cell transcriptomics and lineage tracing analysis, we find that alveolar stem cell differentiation involves a novel, pre-alveolar type-1 transitional state (PATS) en route to their terminal maturation. PATS can be distinguished based on their unique transcriptional signatures, including enrichment for TP53, TGFβ, and DNA damage repair signaling, and cellular senescence in both *in vivo* and *ex vivo* regenerating tissues. Significantly, PATS undergo extensive cell stretching, which makes them vulnerable to DNA damage, a feature commonly associated with most degenerative lung diseases. Importantly, we find enrichment of PATS-like state in human fibrotic lung tissues, suggesting that persistence of such transitional states underlies the pathogenesis of pulmonary fibrosis. Our study thus redefines senescence as a state that can occur as part of a normal tissue maintenance program, and can be derailed in human disease, notably fibrosis.

Introduction

Adult stem cells undergo dynamic changes in phenotype in response to tissue damage [1,2]. These changes include resurgence from a quiescent or poised state, onset of proliferation, activation of new programs of gene expression and a return to homeostasis [3–8]. In many cases, repair also involves changes in epithelial cell shape, for example through transient stretching and expansion to cover areas of damage or denudation, or more permanent changes in cell phenotype. Studies of stem cells during regeneration usually focus on understanding how cells select new differentiation programs in response to signals from the niche. However, much less is known about the significance of changes in parameters such as cell shape and spreading and whether they involve the transient and tightly controlled expression of genes usually associated with pathologic states such as DNA repair and senescence. Here, we explore this question in relation to epithelial repair in the distal gas-exchange alveolar region of the mammalian lung.

In the lung, the maintenance of the alveolar epithelium at homeostasis and its regeneration after injury are fueled by surfactant-producing, cuboidal, type-2 alveolar epithelial cell (AEC2), which can self-renew and differentiate into very large, thin, type-1 alveolar epithelial cells (AEC1), specialized for gas exchange [1,2,9–17]. Recent studies have identified a subset of AEC2s that are enriched for active Wnt signaling and have higher "stemness" compared to Wnt-inactive AEC2s. Such differences in alveolar progenitor cell subsets have been attributed to differences in microenvironmental signals [9,12,18]; in this case, to the vicinity of PDGFRα-expressing alveolar fibroblasts which produce ligands to activate Wnt signaling in AEC2s. Recent studies have also implicated other intercellular signaling pathways, including BMP, Notch, TGFβ, YAP, and NF-κB, in the proliferation

and differentiation of AEC2s, both at steady state and in response to alveolar injury [15,19–21]. However, the precise mechanisms by which the cuboidal AEC2s orchestrate their dramatic changes in cell shape, structure and mechanical properties as they convert into thin, flat AEC1s, remain elusive. In addition, the cellular mechanisms that drive AEC2s to express genes associated with cell senescence, a feature commonly observed in most progressive pulmonary diseases, remain unknown.

Here, using organoid cultures and single-cell transcriptome studies, we uncovered novel, distinct cell states encompassing the transition between AEC2s and AEC1. Moreover, murine lineage tracing, coupled with injury repair models, have revealed the existence of similar transition states *in vivo*. Our study reveals signaling pathways that control these transition states. We also discovered that these novel transitional states exhibit DNA damage responses, and express senescence-related genes en route to AEC1. Importantly, these transitional states correlate with abnormal epithelial cells that are associated with defective fibrotic foci in lungs of human patients with progressive pulmonary fibrosis.

Results

Single-cell transcriptomics revealed novel alveolar epithelial cell states in *ex vivo* organoids.

Recent studies have shown that in response to lung injury, AEC2s proliferate and give rise to AEC1s [11]. Moreover, in alveolar organoid culture AEC2s spontaneously generate AEC1s [10]. However, the molecular mechanisms and transitional cell states underlying the differentiation of AEC2 into AEC1 is still poorly understood. To understand the

mechanisms associated with AEC2 differentiation to AEC1, we performed single-cell transcriptome analysis on cells isolated from alveolar organoids. Purified AEC2 were mixed with fibroblasts and cultured for 10 days to grow organoids for scRNA-seq analysis (Fig. 1a). Uniform Manifold Approximation and Projection (UMAP) identified two major clusters consisting of *Epcam*$^{+}$ epithelial cells and *Vim*$^{+}$/*Pdgfra*$^{+}$ fibroblasts (Supplementary Fig. 1a,b). Next, we further deconvoluted and visualized epithelial cell populations. Within the epithelial cells, we observed multiple sub-clusters: cells expressing high levels of *Sftpc,* a marker for AEC2s, cells expressing high levels of *Ager* (*Rage*), a marker for AEC1s, and *Sftpc*$^{+}$*/Mki67*$^{+}$ proliferating AEC2s (Fig. 1b). These data indicate that under these culture conditions, purified lineage labeled AEC2s proliferate and spontaneously generate AEC1s. In addition, we identified a novel population of alveolar epithelial cells expressing *Cldn4*, *Krt19*, and *Sfn* (Fig. 1b,c). The marker genes unique to this cluster showed two distinct patterns when visualized in UMAP and volcano plots (Fig.1 c,d). One subset (*Ctgf*$^{+}$ cells) is enriched for *Ctgf, Clu, Sox4 and Actn1* while the other (*Lgals3*$^{+}$ cells) is enriched for *Lgals3, Csro1, S100a14, and Cldn18* (Fig.1 c,d). Additional transcripts that are enriched in the *Lgals3*$^{+}$ sub-cluster include *Ager, Emp2,* and *Hopx,* markers of AEC1, suggesting resemblances and a potential lineage hierarchy between *Lgals3*$^{+}$ cells and AEC1 (Fig. 1c,d). These data suggest that the newly identified *Cldn4/Krt19/Sfn*$^{+}$ population is an intermediate between AEC2 and AEC1. We therefore termed this population "pre-alveolar type-1 transitional cell state" (in short, PATS). To validate our single cell data, we performed immunofluorescence analysis for PATS markers on alveolar organoids (Fig. 1e). Immunostaining analysis confirmed the presence of cells expressing PATS-specific markers, including CLDN4, LGALS3, and SOX4 in

alveolospheres (Fig. 1f). Taken together, these data identified unique and novel cell states during alveolar epithelial stem cell differentiation in organoid cultures.

PATS emerge *in vivo* after alveolar injury

We then asked whether PATS can be observed *in vivo* in homeostatic and regenerating alveolar tissues. To test this, first we analyzed a publicly available scRNA-seq data set from alveolar epithelial cells isolated from mice exposed to lipopolysaccharide (LPS), a bacterial endotoxin that causes lung injury [22]. We rendered scRNA-seq data from LPS and control lungs in UMAP plots and found a population that is unique to LPS treated lungs but is not in controls. Significantly, this population is enriched for genes expressed in PATS, including *Cldn4, Sox4, Lgals3,* and *Fn1* (Supplementary Fig. 2). To further characterize PATS, we utilized *Ctgf-GFP* mice*,* a fluorescence reporter transgenic mouse line in which green fluorescent protein is driven by the promoter for *Ctgf*, a gene highly enriched in PATS [23]. We then exposed these mice to bleomycin, a drug that damages the alveolar region causing transient fibrosis. We collected lungs on day 12 post injury for analysis (Fig. 2a). Immunofluorescence analysis for GFP expression in control, uninjured, *Ctgf-GFP* mice revealed GFP signal specifically in alveolar fibroblasts and not in alveolar epithelial cells (Fig. 2b,c and Supplementary Fig. 3). By contrast, in bleomycin injured lungs, we found GFP expression in a significant number of epithelial cells co-labeled by the expression of the AEC2 marker, SFTPC (Fig. 2b,c and Supplementary Fig. 3). These data corroborate our organoid scRNA-seq data and suggest that PATS arise from AEC2s after alveolar injury. Next, we performed co-immunofluorescence analysis for other PATS markers including CLDN4, LGALS3 and SFN. Our analysis revealed that *Ctgf*-GFP

expression in epithelial cells overlaps with that of other PATS markers in bleomycin-injured mice but not in control lungs (Fig. 2 b,c). Interestingly, we noticed that many cells that are positive for *Ctgf*-GFP and other PATS markers show an elongated cell shape as compared to cuboidal AEC2, suggesting that they are en route to differentiation into AEC1 (Fig. 2b,c). Flow cytometric analysis further revealed a significant number of *Ctgf*-GFP^{+} cells and LGALS3^{+} cells in bleomycin treated lungs compared to controls (Fig. 2d). A recent study found elevated levels of KRT8 expression in alveolar epithelial cells after bleomycin-induced injury [24]. Therefore, we tested the expression of KRT8 in alveoli after bleomycin-induced lung injury. We found a significant increase in overlap between KRT8 and *Ctgf*-GFP-expressing cells (Supplementary Fig. 3). Taken together, our data reveal that cells with a PATS phenotype emerge in alveoli after bleomycin injury *in vivo*.

We then asked whether PATS cells are specific to bleomycin injury or appear in other injury models. To test this we employed *AGER-CreER;R26R-DTR* mice in which administration of tamoxifen activates the expression of diphtheria toxin receptor in AGER-Expressing AEC1 cells. This receptor then binds to exogenously administered diphtheria toxin resulting in selective ablation of AEC1 cells (Fig. 2a). We collected lung samples at day 6 after DT injection and performed immunostaining analysis for PATS markers (Fig. 2e). Interestingly, we found cells expressing CLDN4 and LGALS3 that appear to show elongated morphology in AEC1 ablated lungs but not in controls (Fig. 2f and Supplementary Fig. 4). Further quantification revealed a significant number of LGALS3^{+} cells after AEC1 ablation (1.647$\pm$0.1041 after ablation vs 0.274$\pm$ 0.02086 in control) (Fig. 2g). These data suggest that emergence of PATS can be a general mechanism in alveolar regeneration.

Lineage tracing revealed PATS cells originate from AEC2

Our above data from *ex vivo* organoid cultures and *in vivo* injury models suggested that PATS cells originate from AEC2s (Fig. 1). To empirically test this hypothesis, we used a *Sftpc-creER;R26R-tdTomato* mouse model, in which tamoxifen administration permanently induces the expression of tdTomato fluorescent protein specifically in AEC2 and their descendants (Fig. 2h) [25]. On day-10 post bleomycin administration, we observed lineage labeled (tdTomato$^+$) cells co-expressing LGALS3 (Fig. 2i), SFN (Fig. 2j) and CLDN4 (Fig. 2k), and high levels of KRT8 (Supplementary Fig. 5) in damaged regions of the alveoli. Of note, LGALS3, SFN and CLDN4 expression was not observed in tdTomato+ cells in the control lungs (Fig. 2i-k). Interestingly, some lineage labeled SFN$^+$ and CLDN4$^+$ cells showed flat, thin, and elongated morphology and co-expressed AGER, implying that they are in transition to AEC1. Taken together, our data provide evidence that PATS cells emerge *in vivo* from AEC2s after alveolar injury and generate AEC1 (Fig. 2).

Lineage tracing analysis revealed PATS cells generate AEC1

Multiple lines of data from our above experiments suggest that PATS are en route to AEC1 (Fig. 1 and Fig. 2). To test this possibility, we first used an algorithm tool called Velocyto[26], which allows the prediction of cell differentiation trajectories based on ratios between spliced and unspliced mRNA. Using this algorithm, we observed that strong RNA velocities (trajectory) originating from the newly identified *Ctgf*$^+$ population to AEC1 through the *Lgals3*$^+$ population (Fig. 3a-c). These data further strengthened our above

findings that PATS cells are intermediate between AEC2 and AEC1. We then used *Krt19-CreEr an allele of* a gene specifically expressed in PATS, in combination wit*h R26R-tdTomato* (in short, *Krt19-tdTomato*) to carry out lineage tracing in the bleomycin lung injury model to test whether the same trajectory occurs in regenerating alveoli *in vivo*. First, mice were exposed to bleomycin to induce alveolar injury, then tamoxifen was subsequently administered on day 7 after the exposure to label *Krt19*-expressing cells and their progeny by tdTomato (Fig. 3d). In control lungs, we observed a small number of tdTomato labeled cells co-expressing AGER^{+} (1.39%±0.31%) and LGALS3^{hi} macrophages (Fig. 3g and Supplementary Fig. 6a,c). However, we did not find any labeling in AEC2 cells (0%±0) (Supplementary Fig. 6a,c). In contrast, we observed a significant number of tdTomato^{+} cells co-expressing low levels of SFTPC in damaged regions in bleomycin injured lungs (47.22%±6.81%) (Supplementary Fig. 6 a-c). Of note, we did not find tdTomato^{+} cells co-expressing SFTPC in uninjured regions from the same lungs, indicating that *Krt19* expression is specifically activated in response to injury in the damaged regions. These data also indicate that the *Krt19-tdTomato* mouse line and the conditions we tested here are well suited for labeling and tracing PATS. Furthermore, co-immunofluorescence analysis revealed lineage labeled cells co-expressed PATS markers in bleomycin treated lungs but not in control lungs (Fig. 3e-g and Supplementary Fig. 6). Specifically, we observed SFN expression in tdTomato^{+} cells that appear cuboidal (white arrows) as well as in elongated cells (yellow arrowheads) (Fig. 3e-g, right). By contrast, CLDN4 and LGALS3 expression was present mostly in elongating cells (yellow arrowhead and white arrow, respectively)) (Fig. 3e-g right). Importantly, on day-12 post bleomycin administration, we found numerous *Krt19*-tdTomato^{+} cells co-expressing

AGER, a marker for AEC1 (Fig. 3e,f and Supplementary Fig. 6). Thus, our combined RNA velocity and lineage tracing analysis reveal that the newly identified PATS traverse between AEC2 and AEC1.

Conserved Transcriptional programs and pathways in PATS

Differentiation of AEC2 to AEC1 is associated with a dramatic change in the shape and structure of the cells, from a cuboidal to flat and thin morphology. Such a transition is typically accompanied by many changes in the expression of signaling and structural proteins [27,28]. We therefore hypothesized that these changes occur in PATS. To test our hypothesis, we analyzed our scRNA-seq data from *ex vivo* organoid cultures and *in vivo* alveolar injury models (LPS-induced lung injury), focusing on genes that were commonly enriched in both datasets. We found numerous genes that are conserved among PATS (Fig. 4a). Pathway enrichment analysis revealed activation of TP53, TNF/NF-κB, ErbB, HIF1, cell cycle arrest, cytoskeletal dynamics, tight junction signaling and TGFβ signaling in PATS cells compared to other populations (Fig. 4 b,c). Previous studies have implicated all these pathways in alveolar regeneration following injury, highlighting the importance of these pathways in PATS [20,22,29–31]. Unexpectedly, we also found significant enrichment for genes representative of senescence and DNA damage response pathways (Fig. 4).

PATS cells naturally exhibit DNA damage and senescence during alveolar regeneration *in vivo*

Our above data indicated that DNA damage and senescence associated genes are highly enriched in PATS. To validate these findings, we assayed lung tissue sections from bleomycin or PBS treated mice for *β-galactosidase* activity, that serves as a biomarker for senescent cells, as well as γH2AX, a marker for DNA damage signaling. Interestingly, we detected numerous *β-galactosidase* active cells, as assessed based on X-gal derived blue color deposition, in *Sftpc*-tdt lineage derived cells in bleomycin injured alveoli but not in control alveoli (Fig. 5a,b). Similarly, we also observed accumulation of γH2AX puncta in the nuclei of elongated *Sftpc*-tdt$^+$ cells (Fig. 5c,d). Moreover, quantification revealed that 15.09%±1.52% *Sftpc*-tdt$^+$ cells expressed γH2AX in bleomycin injured lungs compared to 0.84%±0.25% in controls (Fig. 5d). Furthermore, we found that γH2AX$^+$ cells also co-expressed LGALS3, indicating that PATS undergo DNA damage during alveolar regeneration.

Bleomycin is known to induce DNA damage in cells. However, both our scRNA-seq analysis and immunostaining revealed a strong enrichment for DNA damage repair signaling in PATS but not in other cell types. To avoid any potential effects from bleomycin on DNA damage, we performed γH2AX staining on lungs from the AEC1-specific cell ablation model. In this model, 37.37%±6.197 LGALS3$^+$ cells showed γH2AX$^+$ nuclear speckles in AEC1 cells compared to 0%±0% in control lungs (Fig. 5e-h). Of note, γH2AX$^+$ cells were not observed in experimental lungs after recovery of AEC1 cells, suggesting that the DNA damage observed in PATS cells is transient and repaired as the cells progress to AEC1 (data not shown).

PATS cells are vulnerable to mechanical stretch-induced DNA damage

Previous studies have shown that ionizing radiation, oxidative stress, and the mechanical stretching that occurs during cell migration and cell stretching can all cause DNA damage[32,33]. Interestingly, we observed an enrichment for genes involved in cytoskeletal changes but not oxidative stress in our pathway analysis. Since, AEC2 differentiation into AEC1 requires extensive cell stretching and cytoskeletal dynamics (Fig. 4) we hypothesized that PATS cells experience significant mechanical stretching that can lead to DNA damage. To test this hypothesis, we purified AEC2s from *SFTPC-GFP* mice and cultured them on a plastic surface (2D cultures), conditions under which AEC2s stretch and differentiate into AEC1s [34]. Within 5 days after plating, most AEC2s stretched and either lost (GFP^-) or downregulated (GFP^{lo}) GFP expression. These GFP^- and GFP^{lo} cells increased their surface area through stretching and spreading and began to express AGER, a marker for AEC1. Interestingly, these AGER+ cells are also positive for the DNA damage marker, γH2AX (Fig. 5i,k). Of note, we did not see cell stretching and DNA damage markers in cells that we collected on day-2 after plating (Fig. 5k). Taken together, our data show that cuboidal AEC2s that differentiate into large and thin AEC1 during alveolar regeneration naturally experience DNA damage and repair and undergo transient senescence.

Ectopic activation of TP53 signaling promotes AEC2 to AEC1 differentiation via PATS during alveolar injury-repair

Our data suggest that activation of TP53 signaling is associated with PATS during AEC2 differentiation into AEC1 after lung injury (Fig. 4). Furthermore, previous studies in other tissues have suggested that TP53 signaling promotes differentiation and suppresses

stem cell self-renewal [35]. This led us to hypothesize that enhanced activation of TP53 signaling increases AEC2 differentiation into AEC1 involving PATS after bleomycin-induced lung injury. To test this hypothesis, we utilized the *Sftpc-CreER;R26-tdTomato* mouse model, in which tamoxifen administration is followed by bleomycin or PBS. These mice were then administered Nutlin-3a (an activator of the TP53 pathway) or DMSO (control) daily starting on day-8 after bleomycin and tissues were collected on day-18 (Fig. 6a). Immunostaining analysis for the AEC1 cell marker, AGER, revealed that Nultin-3a treatment led to significantly greater differentiation of lineage labelled AEC2 cells into AEC1 compared to DMSO controls (Fig. 6b). We did not observe any AGER$^+$*Sftpc*-tdt$^+$ cells in uninjured lungs that received Nutlin-3a, suggesting that in the absence of injury, Nutlin3a-induced TP53 activation alone is not sufficient to induce AEC2 differentiation into AEC1 (Fig. 6b). Our quantitative analysis revealed that 67.25%±1.94% AGER$^+$ cells are *Sftpc*-tdt$^+$ after bleomycin injury and Nutlin-3a treatment compare to 54.55%±3.71% in bleomycin and DMSO treated animals. Additionally, we also tested the effect of Nutlin-3a on AEC2s in alveolar organoid cultures. Organoid cultures were treated with Nutlin-3a or DMSO starting on day-7 until harvest on day 15 (Fig. 6d). Immunostaining analysis revealed that organoids treated with Nutlin-3a showed both an increase in the number and the expression intensity of AGER$^+$ cells and a decrease in the number of proliferating cells as demonstrated by Ki67 immunostaining compare to controls (Fig. 6d,e). These data further support our hypothesis that TP53 signaling enhances AEC2 differentiation both *in vivo* regenerating tissues and *ex vivo* organoid cultures.

Transcriptional control of PATS by transcription factor TP53

To identify the transcription factor modules that are potentially active in PATS, we performed chromatin immunoprecipitation analysis for Histone H3 lysine 4 tri-methylation (H3K4me3), Histone H3 lysine 36 tri-methylation (H3K36me3), and Histone H3 lysine 27 acetylation (H3K27ac), to identify active promoters, transcribing genes, and enhancers, respectively. We isolated and sorted PATS from *Ctgf-GFP* mice after bleomycin administration and AEC2s from control mice. Sorted cells were used for profiling the above-mentioned histone marks using Mint-ChIP, a recently described method that requires fewer cell numbers (Fig. 6f) [36]. As expected, we found H3K4me3 enrichment in genes corresponding to cell type specific promoters (for example, *Sftpc in* AEC2s and *Fn1* in PATS) (Supplementary Fig. 7a,b). Further analysis revealed enrichment of numerous H3K4me3 peaks overlapping with transcriptional start sites (TSS) and promoters of transcripts specific to PATS from bleomycin exposed lungs compared to AEC2s from controls (Fig. 6g-h). Motif analysis was performed to predict enrichment of binding sites for transcription factors (TFs) in H3K4me3 and H3K27ac peaks. We found enrichment of binding sites for TFs including TP53, ETS1, NF1, ATF3, and SOX4, all of which have been implicated in PATS enriched pathways such as cell cycle arrest, senescence, DNA damage repair, and cytoskeletal control (Fig. 6g,h) [37–40]. For example, we found predicted binding sites for TP53 in the *Mdm2* enhancer and *Fas* promoter, two well-known direct targets of TP53 (Fig. 6i,j) [41]. Similarly, we found predicted binding sites for NF1, ETS1, and ATF3 in the promoter regions of PATS specific genes including *Lgals3 and Sox4* (Supplementary Fig. 7c,d). Taken together, our data implicate a direct role for TP53 in transcriptional control of PATS-specific genes.

PATS associated gene expression signatures and signaling pathways are enriched in human fibrotic lungs

Numerous studies have suggested that abnormal ECM remodeling and chronic inflammation and stress pathways are associated with defective regeneration in multiple organs, including the lung [42–45]. Indeed, recent studies have suggested that alveolar epithelial cells that line fibrotic foci in idiopathic pulmonary fibrosis (IPF) show features of senescence, growth arrest, and differentiation blockade [46–48]. Thus, we hypothesized that in response to a non-permissive pathologic microenvironment, alveolar progenitors stall their differentiation process in a PATS-like state. To address this hypothesis, first we analyzed a recently reported scRNA-seq dataset from human IPF lungs [49]. First, we segregated alveolar epithelial cells by excluding all non-epithelial cells and airway cells from further analysis. We observed numerous AEC2s and AEC1s in both healthy and IPF lungs, with apparent overlap in UMAP plots (Fig. 7a-c). By contrast, we found a distinct cell cluster that is highly enriched in IPF samples and did not overlap with either AEC1 or AEC2 (Fig. 7a,b). Previously, these cells were annotated by their marker gene expression as $KRT5^{-}/KRT17^{+}$[49]. Significantly, differential gene expression analysis revealed a striking resemblance of transcripts, including *Sfn, Sox4* and *Fn1,* between the IPF-enriched cell cluster and PATS from our *ex vivo* organoids and LPS-induced murine lung injury model *in vivo* (Fig. 1, Fig. 4, Fig. 7a-e, and Supplementary Fig. 8a,b). We therefore named this human IPF-specific cluster as ‘PATS-like’ cells (Fig. 7a-c). Other transcripts that are highly enriched in the PATS-like population are *KRT17, CALD1, PRSS2, MMP7,* and *S100A2.* Gene ontology and pathway analysis revealed an enrichment for components of p53 signaling (*CDKN1A, CDKN2A*, and *MDM2*), DNA-damage checkpoint (*RPS27L*

and *PLK3*) and cellular senescence (*TGFB2 and HIPK2*) (Fig. 7d-f and Supplementary Fig. 8). Moreover, the PATS-like cluster is enriched for components of focal adhesion, tight junction, and regulation of actin cytoskeleton, indicating a remarkable resemblance between PATS-like cells in IPF and those in organoids and regenerating alveoli *in vivo* (Fig. 4 and Supplementary Fig. 8c). We also found a significant enrichment for *AREG*, *TGFB1*, *TGFB2* and *TIMP1*, known regulators of fibrosis, in PATS-like cells (Fig. 7f). To validate single cell transcriptome data, we performed immunofluorescence analysis on human lung sections from healthy and IPF samples (Supplementary Fig. 9a). Co-immunofluorescence analysis for markers of PATS-like cells (SFN, CLDN4, LGALS3 and KRT17), AEC2 (HTII-280), AEC1 (AGER) and myofibroblasts (ACTA2) revealed the expression of PATS-like markers specifically in IPF lungs but not in healthy controls and healthy appearing regions in IPF lungs (Fig. 7g-i and Supplementary Fig. 9). Interestingly, KRT17 has been recently shown to be expressed in basal cells of the normal lung and basaloid-like cells in pulmonary fibrotic lungs [49,50]. In addition, we also found similar expression pattern of TP63, another marker for basal cells (Supplementary Fig. 9). We find that markers of PATS-like cells are almost exclusively present in fibrotic regions as assessed based on high levels of ECM deposition and accumulation of myofibroblasts (Fig. 7g-i and Supplementary Fig. 9). Quantitative analysis for SFN^+ cells in total $HTII\text{-}280^+$ cells further supported these findings (91.07%±3% in IPF samples vs 4.16% ±1.02% in healthy human) (Fig. 7h). Moreover, we observed that SFN^+, $HTII\text{-}280^+$ cells lost their cuboidal shape and acquired an elongated morphology in IPF lungs as opposed to cuboidal AEC2 or extremely thin and elongated AEC1 in healthy lungs (Fig. 7g,i and Supplementary Fig. 9b,c). Of note, AGER expression is absent in PATS-like cells in IPF

lungs (Fig. 7i). Next, we analyzed markers of cell senescence (β-galactosidase activity and CDKN1A/p21) and DNA damage response (γH2AX) in IPF and healthy lungs. Our data revealed specific expression of these markers in the PATS-like population in IPF but not in healthy lungs (Fig. 7 j,k and Supplementary Fig. 9 d,e). Taken together, our analysis revealed similarities between PATS from regenerating tissues and PATS-like cells that are specific to pathological fibrotic lungs.

Discussion

The work reported here has uncovered a previously unknown transitional state (PATS) that traverses between AEC2 and AEC1 and has a specific gene expression signature in both *ex vivo* organoid cultures and *in vivo* regenerating tissues. These findings further highlight the resemblance between *ex vivo* organoid models and their *in vivo* correlates. Our fate mapping studies using *Sftpc*-driven and *Krt19*-driven CreER alleles demonstrate a clear linear lineage relationship between AEC2 to PATS to AEC1. Therefore, with the addition of this novel transitional cell state, our study revises the alveolar epithelial cell hierarchy. Based on distinct gene expression signatures, our study indicates that PATS cells are not merely undergoing a gradual loss of AEC2 characteristics but represent a unique transitional population. These cells are identified by their expression of many pathways, including, TP53, NF-κB, YAP, TGFβ, and HIF1, previously shown to be important for lung regeneration [20,22,30,51,52]. In addition, we observed enrichment for transcripts associated with cell cycle arrest, senescence TP53 and TGFβ signaling, and the transcription factor SOX4, a known regulator of epithelial-mesenchymal transition and cell adhesion in other tissues [53–55]. These transcriptional signatures are consistent with

391 the dramatic morphological changes that occur during AEC2 to AEC1 differentiation.
392 Through pharmacological modulation, our data demonstrate that TP53 signaling
393 promotes AEC2 differentiation into AEC1 involving PATS in both *ex vivo* organoid cultures
394 and *in vivo* regenerating tissues. Our analysis of human lungs identified PATS-like cells
395 that are specifically present in alveoli of lungs with pulmonary fibrosis. Similar to murine
396 cells, PATS-like cells in human lungs are also characterized by enrichment for genes
397 associated with cellular senescence and TP53 signaling, as well as TGFβ regulated
398 genes, all of which have been implicated in the pathogenesis of fibrosis multiple organs,
399 including the lung [56,57]. In contrast, we also found some differences in gene expression
400 signatures between murine PATS and human IPF-specific PATS-like cells. Some of them
401 include, TP63, KRT17, and COL1A1, which are found only in human PATS-like cells.
402 Interestingly, these genes are characteristic markers of basal cells of the normal airway.
403 Our RNA velocity projections from human scRNA-seq data suggest that these
404 KRt17^{+}/TP63^{+} cells originate from AEC2. Indeed, our immunofluorescence studies for
405 KRT17 and TP63 further suggested that these cells are surrounded by cells that express
406 markers of PATS within the same alveoli. All together, these data suggest that
407 KRT17^{+}/TP63^{+} cells originate from AEC2.

408 Another significant finding from this study is that PATS cells undergo extensive
409 stretching and spreading, which makes them vulnerable to DNA damage, a feature
410 commonly associated with most degenerative lung diseases, notably pulmonary fibrosis
411 and cancers [58–62]. Of note, we did not find cells with DNA damage in AEC1 of lungs that
412 had fully regenerated following bleomycin-induced injury, suggesting that alveolar cells
413 can efficiently repair the damage DNA (data not shown). However, it is also possible that

414 tissue may eliminate such “unfit cells” (with DNA damage) through cell extrusion or cell
415 death mechanisms. Nevertheless, our study uncovered a novel transitional cell state that
416 is vulnerable to DNA damage during AEC2 differentiation into AEC1. Previous studies
417 have revealed that cell stretching causes DNA damage when they migrate or squeeze
418 through narrow spaces [63]. Indeed, our *in vivo* AEC1 ablation and *ex vivo* 2D culture
419 models suggest similar mechanisms are at play when cuboidal AEC2 undergo extensive
420 stretching during their differentiation into thin and large AEC1. Therefore, the novel
421 transitional cell state has clinical implications as most degenerative lung diseases are
422 accompanied by the presence of cells with DNA damage accumulation [59,60]. Interestingly,
423 the PATS population is enriched for SOX4, a transcription factor known to regulate
424 cytoskeletal genes, which is induced following DNA damage and has a crucial role in
425 TP53 stabilization and function[40]. Altogether, these data support a model in which, cells
426 evolved co-transcriptional programs to combat DNA damage that can occur when cells
427 undergo extreme stretching. In addition, genome-wide association studies have identified
428 mutations in DNA damage repair pathways components, such as *XRCC* family genes,
429 *LIG4*, *TERC*, PARP, and *RTEL1* with emphysema and pulmonary fibrosis[58,64]. Thus, the
430 novel transitional state we have identified here implicates cell shape changes and
431 associated vulnerabilities accompanying alveolar stem cell differentiation in the
432 pathogenesis of some lung diseases (Fig. 8).
433 Senescence is often seen as an age-associated pathological state in which cells acquire
434 an abnormal and irreversible state[48,62,65,66]. Here, we find that alveolar stem cell
435 differentiation involves a novel transitional state which exhibits cardinal features of
436 senescence in normal tissue regeneration. Indeed, prior studies have found senescent

cells in embryos in developing limb bud tissues[67]. Therefore, we propose that senescence may not necessarily occur exclusively in aged tissues but can be a reversible transient state accompanying tissue repair or regeneration. Our study thus redefines senescence as a state that can occur as part of normal tissue maintenance programs and can be derailed in human diseases.

In conclusion, using alveolar organoid cultures and *in vivo* injury-repair models, we have identified a novel pre-AEC1 transitional state in lung regeneration. This novel and unique state is associated with cellular senescence and enrichment for pathways known to be associated with defective alveolar regeneration. These results strongly suggest that prolonged senescence and stress mediated pathways in transitional cell states can lead to diseases such as fibrosis.

Online Methods

Mice

Sftpc$^{tm1(cre/ERT2)Blh}$ (*Sftpc-CreER*) [25], *Krt19*$^{tm1(cre/ERT)Ggu}$*/J* (*Krt19-CreER*) [68], *Rosa26R-CAG-lsl-tdTomato* [69] (crossed with *Sftpc-CreER*), *B6.Cg-Gt(ROSA)26Sor*$^{tm14(CAG-tdTomato)Hze}$*/J* (*R26R-tdTomato*) [70] (crossed with *Krt19-CreER*), *Tg(SFTPC-GFP)#Heat (Sftpc-GFP)* [71], *B6.Cg-Ager*$^{tm2.1(cre/ERT2)Blh}$*/J* (*Ager-CreER*) [72], *C57BL/6-Gt(ROSA)26Sor*$^{tm1(HBEGF)Awai}$*/J* (*R26R-DTR*) [73], *Mki67*$^{tm1.1Cle}$*/J* (*Mki67RFP*) [74] and *Tg(Ctgf-EGFP)FX156Gsat* (*Ctgf-GFP*) [23] were utilized for experiments. For lineage tracing with *Sftpc-CreER;R26R-tdTomato*, 0.2 mg/g body weight Tamoxifen (Tmx, Sigma-Aldrich) was given via oral gavage or intraperitoneal injection. For lineage tracing using *Krt19-CreER;R26R-tdTomato*, one

460 dose of 0.1 mg/g body weight tamoxifen was given via intraperitoneal injection seven
461 days after bleomycin injury or PBS administration. Animal experiments were approved
462 by the Duke University Institutional Animal Care and Use Committee.
463
464 **Bleomycin injury**
465 For bleomycin-induced lung injury, 2.5 U/Kg bleomycin was administered intranasally at
466 two weeks after tamoxifen injection and mice were monitored daily. PBS administered
467 mice served as controls. Mice were sacrificed at different time points after bleomycin
468 injury.
469
470 **Diphtheria toxin (DT) administration**
471 Two weeks before DT administration, *Ager-CreER;R26R-DTR* received tamoxifen via IP
472 injection. One dose of 3 μg diphtheria toxin (Millipore #322326) was administer via
473 intraperitoneal injection and mice were sacrificed six days later for tissue collection and
474 analysis.
475
476 **Mouse lung dissociation and fluorescence assisted cell sorting**
477 Lung dissociation and FACS were performed as described previously[19]. Briefly, lungs
478 were intratracheally inflated with 1 ml of enzyme solution (Dispase (5 U/ml, Corning
479 #354235), DNase I (0.33 U/ml) and collagenase type I (450 U/ml, Gibco #17100-017)) in
480 DMEM/F12. Separated lung lobes were diced and incubated with 3 ml enzyme solution
481 for 25 min at 37ºC with rotation. The reaction was quenched with an equal amount of
482 medium containing 10% Fetal bovine serum (FBS) and filtered through a 100 μm strainer.

The cell pellet was resuspended in red blood cell lysis buffer (155 mM NH_4Cl, 12 mM $NaHCO_3$, 0.1 mM EDTA) and incubated for 2 min, then filtered through a 40 μm strainer. The cell pellet was resuspended in DMEM/F12 + 2% BSA and stained with following antibodies: EpCAM (eBioscience, G8.8), PDGFRα (BioLegend, APA5) and Lysotracker (Thermo Fisher, L7526) as described previously[29]. Sorting was performed using a BD FACS Vantage SE, SONY SH800S or Beckman Coulter MoFlo Astrios EQ.

Alveolar organoid culture

Alveolar organoid culture was performed as described previously [10]. Briefly, lineage labeled AEC2s (1-3 × 10^3) from *SFTPC-GFP* or *Sftpc-CreER;R26R-lsl-tdTomato* mice treated with Tmx were FACS-sorted and PDGFRα$^+$ (5 × 10^4) fibroblasts were resuspended in MTEC/Plus and mixed with equal amount of growth factor-reduced Matrigel (Corning #354230). Medium was changed every other day.

Nutlin-3a treatment

For *in vivo* studies, *Sftpc-CreER;R26R-tdTomato* were injected with one dose of tamoxifen and rested for two weeks followed by bleomycin or PBS administration. Eight days after injury, Nutlin-3a (Selleckchem #S8059) or DMSO (control) was administered by intraperitoneal injection at concentration of 20mg/Kg/day for ten consecutive days and samples were collected on 20 days after bleomycin administration. For *Ex vivo* study, alveolar organoids were grown for seven days followed by Nutlin-3a (2 μM) treatment for 8 days before their harvest.

Droplet-based single-cell RNA sequencing (Drop-seq)

Organoids embedded in Matrigel were incubated with *Accutase* solution (sigma #A6964) at 37°C for 20 min followed by incubation with 0.25% Trypsin-EDTA at 37°C for 10 min. Trypsin was inactivated using DMEM/F-12 Ham supplemented with 10% FBS and cells were then resuspended in PBS supplemented with 0.01% BSA. After filtration through 40 µm strainer, cells at a concentration of 100 cells/µl were run through microfluidic channels at 3,000 µl/h, together with mRNA capture beads at 3,000 µl/h and droplet-generation oil at 13,000 µl/h. DNA polymerase for pre-amplification step (1 cycle of 95°C for 3 min, 15-17 cycles of 98°C for 15 sec, 65°C for 30 sec, 68°C for 4 min and 1 cycle of 72°C for 10 min) was replaced by Terra PCR Direct Polymerase (#639271, Takara). The other processes were performed as described in the original Drop-seq protocol [75]. Libraries were sequenced using HiSeq X with 150-bp paired end sequencing.

Computational analysis for scRNA-seq

scRNA-seq analysis of alveolar organoids was performed by processing FASTQ files using dropSeqPipe v0.3 (https://hoohm.github.io/dropSeqPipe) and mapped on the GRCm38 genome reference with annotation version 91. Unique molecular identifier (UMI) counts were then further analyzed using an R package Seurat v3.1.1 [76]. UMI count matrix of murine lungs treated with LPS (GSE130148) [22] was obtained from Gene Expression Omnibus (GEO). UMI counts were normalized using SCTransform. Cell barcodes for the clusters of interests were extracted and utilized for *velocyto run* command in velocyto.py v0.17.15 [77] as well as generating RNA velocity plots using velocyto.R v0.6 in combination with an R package SeuratWrappers v0.1.0

(https://github.com/satijalab/seurat-wrappers). Twenty-five nearest neighbors in slope calculation smoothing was used for *RunVelocity* command. After excluding duplets, specific cell clusters were isolated based on enrichment for *Sftpc, Sftpa1, Sftpa2, Sftpb, Lamp3, Abca3, Hopx, Ager, Akap5, Epcam, Cdh1, Krt7, Krt8, Krt18, Krt19, Scgb1a1* and *Scgb3a1* as well as negative expressions of *Vim, Acta2, Pdgfra* and *Pdgfrb* in UMAP plots. The Rds files for control and idiopathic pulmonary fibrosis (IPF) lungs were obtained from GEO (#GSE135893)[49]. Cell clusters of AEC2, AEC1, transitional AEC2 and *KRT5*$^{-}$/*KRT17*$^{+}$ were extracted and analyzed. Markers for each cluster (Supplementary Table 1) obtained using *FindAllMarkers* command in Seurat were utilized for identifying specific signaling pathways and gene ontology through Enrichr [78]. Z-scores were calculated based on combined score in Kyoto encyclopedia of genes and genomes (KEGG) to compare enrichment of signaling and ontology across different cell clusters. The results were displayed in heatmap format generated using an R package pheatmap v1.0.12. Scaled data in Seurat object were extracted and mean values of scaled score of gene members in each pathway were calculated and shown in UMAP as enrichment of signaling pathways. The gene member lists of utilized pathways were obtained from KEGG pathways [79] and AmiGO [80]. Log_2 fold changes and P-values in each gene extracted using *FindMarkers* command in Seurat with Wilcoxon rank sum test were shown in a volcano plot using an R package EnhancedVolcano v1.3.1 (https://github.com/kevinblighe/EnhancedVolcano) to show specific markers for *Ctgf*$^{+}$ cells and *Lgals3*$^{+}$ cells.

Mint-ChIP (Multiplexed indexed T7 ChIP-seq)

552 FACS-sorted CD31$^-$/CD45$^-$/CD140a$^-$/CD326$^+$/*CTGF*-GFP$^+$ cells (PATS) from *Ctgf-GFP*
553 mice treated with Bleomycin (d12) and CD31$^-$/CD45$^-$/CD326$^+$/Lysotracker$^+$/*Mki67*-RFP$^-$
554 cells (AEC2) from *Mki67-RFP* homeostatic mice were processed using Mint-ChIP
555 described previously[36]. Open-sourced newest protocol named Mint-ChIP3 in protocol.io
556 was used. Following cell lysis, chromatin was digested with 300 units reaction of MNase
557 (#M0247S, New England Biolabs) at 37°C for 10 min. T7 adapter ligation was performed
558 for 2 hrs and then the samples were split to have ~7,000 cells per antibody. The samples
559 were incubated with Histone H3 (H3) antibody (1 ul, #39763, Active Motif), Histone H3
560 lysine 36 trimethylation (H3K36me3) antibody (1 ul, #61101, Active Motif) or Histone H3
561 lysine 4 trimethylation (H3K4me3) antibody (1 ul, #ab8580, Abcam) and Histone H3 lysine
562 27 acetylation (H3K27ac) antibody (1 ul, #39133, Active Motif) at 4°C overnight. DNA was
563 purified followed by T7-driving *in vitro* transcription at 37°C for 3 hrs. Reverse transcription
564 was performed as described in original protocol followed by library preparation using
565 Terra Direct PCR polymerase (#639271, TaKaRa). Two experimental replications were
566 performed for each cell type. Libraries were sequenced using Hiseq X or NovaSeq 6000
567 with at least 5M reads of 150-bp paired end per sample.
568
569 **Computational analyses for Mint-ChIP**
570 FASTQ files were generated using Bcl2fastq. Additional demultiplexing for Mint-ChIP
571 FASTQ files were performed using Je [81]. Low quality reads were trimmed out from FASTQ
572 files using trimmomatic v0.38 [82]. Reads were mapped on mm10 genome reference using
573 BWA [83]. HOMER [84] was used for generating bedGraph files to visualize them in
574 Integrative Genomics Viewer (IGV)[85]. Peak calling for H3K4me3 was performed using

575 HOMER's *getDifferentialPeaksReplicates.pl -region -size 1000 -minDist 2000 -C 0 -L 50*
576 with normalization by H3. Motif analysis was performed using HOMER's
577 *findMotifsGenome.pl*. The packages were run through a pipeline called MintChIP
578 (https://github.com/jianhong/MintChIP). deepTools [86] was used for generating a chart of
579 called peaks of H3K4me3. Called peaks of H3K4me3 in both homeostatic AEC2 and
580 PATS after bleomycin-induced lung injury were merged in Fig. 6g using Affinity Designer.
581 Genes shown in Fig. 4a left panel were used to generate Fig. 6g.
582
583 **Human lung tissue**
584 Excised subtransplant-quality human lung tissues without preexisting chronic lung
585 diseases were obtained from the Marsico Lung Institute at the University of North Carolina
586 at Chapel Hill under a University of North Carolina Biomedical Institutional Review Board-
587 approved protocol (#03-1396). Informed consent was obtained from all participants where
588 necessary. Samples of explanted fibrotic human lungs were procured through the
589 BioRepository and Precision Pathology Center at Duke University in accordance with
590 institutional procedures (Duke University Pro00082379 – "Human Lung Stem Cells";
591 exempt research as described in 45 CFR 46.102(f), 21 CFR 56.102(e) and 21 CFR
592 812.3(p) which satisfies the Privacy Rule as described in 45CFR164.514). The diagnosis
593 of idiopathic pulmonary fibrosis (IPF) was evaluated by a surgical pathology team.
594 Specimens were washed thoroughly in PBS prior to inflation and immersion in 4% PFA
595 overnight at 4°C. Specimens were subsequently washed in PBS until the appearance of
596 blood was minimal followed by incubation in 30% sucrose at 4°C. Samples were then

incubated with 1:1 mixture of OCT for 1 hour at 4°C before embedding in OCT. 7-9 μm thick sections were used for histological analysis.

Immunostaining

Lungs and alveolar organoids were prepared as described previously. Briefly, tissues were fixed with 4% paraformaldehyde (PFA) at 4ºC for 4 h and at room temperature for 30 min then embedded in OCT or Paraffin. 10-μm sectioned samples were utilized for staining following incubation at 95°C for 10-15 min for antigen retrieval using 10 mM sodium citrate. Primary antibodies were as follows: Pro-surfactant protein C (Millipore, ab3786, 1:500), AGER (R&D systems, MAB1179, 1:250), KRT8 (DSHB, TROMA-I, 1:50), KRT17 (NSJ, V2176; 1:250), KRT19 (DSHB, TROMA-III, 1:50), tdTomato (ORIGENE, AB8181-200, 1:500), CLDN4 (Invitrogen, 36-4800, 1:200), GFP (Novos Biologicals, NB100-1770, 1:500), LGALS3 (Cedarlane, CL8942AP, 1:500); SOX4 (Invitrogen, MA5-31424, 1:250), SFN (Invitrogen, PA5-95056, 1:250 or Proteintech, 66251-1-Ig, 1:500), ACTA2 (Sigma, C6198, 1:500), and , gamma-H2AX (R&D, 4418-APC, 1:500).

β-Galactosidase (X-gal) staining and Hematoxylin & Eosin (H&E) staining

PFA-fixed frozen sections were incubated with X-gal staining buffer containing 1 mg/ml of X-gal (Thermo, R0941), 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.01% sodium deocycholate and 0.02% NP-40 at 37ºC overnight. Sections were washed 3 time in PBS and mounted. For H&E staining, 10-μm paraffin sections were submerged in Histo-clear and series of ethanol. Mayer's Hematoxylin was used to stain nuclei, followed by staining using 1% Eosin Y.

Image acquisition, processing and quantification

Images were captured using Olympus Confocal Microscope FV3000 using a 20×, 40× or 60× objective, a Zeiss wide-field fluorescence microscope (X-gal staining) and a Zeiss Axio Imager Widefield Fluorescence Microscope (H&E). Cells were manually counted based on IHC markers and DAPI. For determination of average intersects per linear distance, a mean linear intercept analysis was conducted as previously described[87] over the single channel immunofluorescence stain of interest. Images were processed using Olympus CellSens application or ImageJ and figures were prepared using Affinity Designer.

Statistics

Sample size was not predetermined. Data are presented as means with standard error (s.e.m) to indicate the variation within each experiment. Statistics analysis was performed in GraphPad Prism. A two-tailed Student's *t*-test was used for the comparison between two experimental conditions. We used Mann Whitney one tailed test for the comparison between two conditions that showed non-normal distributions.

Acknowledgements

We thank Brigid Hogan for advice and critical reading of the manuscript, Monica Fernandez de Soto for technical support and members of the Tata lab and Barkauskas lab for discussions. We thank Karen Lyons for providing *Ctgf*-GFP mice, Peter van Galen and Yu Xiang for help with Mint-ChIP analysis., the Duke Cancer Institute Flow Cytometry

643 Shared Resource for help with cell sorting, the Duke Sequencing and Genomic
644 Technologies Shared Resource for sequencing NGS-libraries, the Duke University Light
645 Microscopy Core Facility for imaging equipment and consultation, the Duke Biorepository
646 & Precision Pathology Center for providing human specimens, the Duke Compute Cluster
647 for help with sequencing data analysis. The TROMA-I (KRT8) antibody clone developed
648 by Brulet, P/Kemler was obtained from the Developmental Studies Hybridoma Bank,
649 created by the NICHD of the NIH and maintained at The University of Iowa, Department
650 of Biology, Iowa City, IA 52242. Y.K. is a Japan Society for the Promotion of Science
651 Overseas Research Fellow. A.K. is supported by a medical scientist training program
652 fellowship from NHLBI/NIH (F30HL143911). Human scRNA-seq was supported by
653 R01HL145372 (NEB/JAK), Doris Duke Charitable Foundation (JAK),
654 K08HL130595(JAK), and Boehringer Ingelheim Pharmaceuticals (JAK). This work was
655 supported by a Pathways to Independence award from NHLBI/NIH (R00HL127181) to
656 P.R.T. and funds from Regeneration NeXT and Kaganov- MEDx Pulmonary Initiative at
657 Duke University. This work was partially supported by funds from Whitehead foundation
658 and P.R.T. is a Whitehead Scholar.

659

660 **Author contributions**

661 Y.K. designed, conceived and performed NGS-related experiments and the
662 computational analysis, and co-wrote the manuscript; A.T. designed, conceived and
663 performed in vivo and ex vivo experiments and immunostaining, and co-wrote the
664 manuscript; A.K. designed and performed AEC1 ablation experiments; H.K. performed
665 organoid experiments. R.F.L. performed histological analysis; J.O. built a pipeline for

Mint-ChIP analysis. N.E.B. and J.A.K. provided scRNA-seq data from healthy and IPF human lungs. P.R.T. designed, conceived and supervised the work and co-wrote the manuscript. All authors reviewed and edited the manuscript.

Competing interests

Nothing to declare.

Additional information

Extended Data is available online

Data availability: The data described in this manuscript are available from the corresponding author upon request. All the NGS sequencing data in this manuscript will be available at NCBI GEO (accession no. GSE will be provided upon acceptance of the manuscript).

Correspondence and requests for materials

Correspondence and requests for materials should be addressed to P.R.T.

Figure legends

Figure 1. scRNA-seq reveals novel alveolar epithelial cell states in *ex vivo* organoids.

a, Schematic of alveolar organoid culture utilized for single-cell RNA-seq. b, Uniform manifold approximation and projection (UMAP) visualization of epithelial populations in

689 cultured alveolar organoids. AEC2 (green) – alveolar epithelial type-2 cells, AEC2-
690 proliferating (pink) – proliferating alveolar epithelial type-2 cells, AEC1 (yellow) – alveolar
691 epithelial type-1 cells, New cell states :*Lgals3*$^{+}$ cells (blue) and *Ctgf*$^{+}$ cells (red) c, UMAP
692 plots show the expression of indicated genes in epithelial populations in cultured alveolar
693 organoids. Dotted circles in b and c indicate the novel cell states. d, Volcano plot shows
694 specific genes to *Ctgf*$^{+}$ population and *Lgals3*$^{+}$ population. e, Schematic of alveolar
695 organoid culture using fibroblasts and AEC2 cells sorted from *SFTPC-GFP* mice. f,
696 Immunostaining for PATS markers in alveolar organoids. Co-staining of LGALS3 (green)
697 and SFTPC (red) is shown in left panel; localization of CLDN4 (red) - middle panel and
698 *SFTPC*-GFP (green) co-stained with SOX4 (red) and SFTPC (grey) (right panel). DAPI
699 stains nuclei (blue). Scale bars indicate 20 μm. Arrows indicate PATS.
700
701 **Figure 2. The novel alveolar cell states emerge transiently and originate from**
702 **alveolar stem cells after alveolar injury *in vivo*.**
703 a, Schematic of bleomycin-induced lung injury in *Ctgf-GFP* mice. b-c, Immunostaining for
704 PATS markers. b, co-staining for *Ctgf*-GFP (green), CLDN4 (red) and LGALS3 (grey), c,
705 Co-staining for *Ctgf*-GFP (green), SFN (red) and LGALS3 (grey) in control lung (upper)
706 and bleomycin-treated lungs collected on day 12 after injury (lower). d, Flowcytometric
707 analysis of PATS in *Ctgf*-GFP mouse model. The percentage of *Ctgf*-GFP (CD31^{-}/CD45^{-}
708 /EPCAM^{+}) (red line) and LGALS3 (CD31^{-}/CD45^{-}/EPCAM^{+}) (blue line) are shown as
709 indicated in control (upper panel) and bleomycin injures (lower panel). e, Experimental
710 design to ablate AEC1 cells in *Ager-CreER;R26R-DTR* mouse model. Mice were
711 administered with tamoxifen (Tmx) followed by diphtheria toxin (DT) and tissue collection

on day 6. f, Immunostaining for CLDN4 (green) and LGALS3 (grey) in control (left) and AEC1-ablated lungs (right). g, Mean linear intercept length analysis of LGALS3$^+$ cells in control and AEC1-ablated lungs. Asterisks indicate $p < 0.0002$ (un-paired student's t-test). Data are from three independent experiments and are presented as mean ± s.e.m. h, Experimental workflow for sequential administration of tamoxifen followed by bleomycin injury and tissue collection for analysis using *Sftpc-CreER;R26R-tdTomato* mice. i-k, Immunostaining for PATS markers in Sftpc-lineage labeled cell in control (upper panel) and bleomycin inured lungs (lower panel). i, SFTPC (green), *Sftpc*-tdt (red) and LGALS3 (grey), j, SFN (green), *Sftpc*-tdt (red) and AGER (grey), and k, CLDN4 (green), *Sftpc*-tdt (red) and AGER (grey). DAPI stains nuclei (blue). Scale bars indicate 30 μm. White line box in merge image indicates region of single channel images shown in left side.

Figure 3. Lineage tracing revealed that PATS generate AEC1.

a, UMAP of epithelial populations in cultured alveolar organoids. Arrow indicates selected cell populations in the oval-shaped circle are shown in panel b. b, UMAP plots show the expression of indicated genes in the selected populations (oval-shaped circle in panel a). c, RNA velocity analysis for PATS and AEC1. Arrows indicate predicted lineage trajectories. d, Schematic representation of experimental design to sequentially administer bleomycin (injury) or PBS (control) followed by tamoxifen (to label Krt19-expressing cells) in *Krt19-CreER;R26R-tdTomato* mice. e. Immunostaining for SFTPC (green), *Krt19*-tdt (red) and AGER (grey). White arrows indicate SFTPC$^+$, *Krt19*-tdt$^+$ cells. Yellow arrowhead indicates AGER$^+$, *Krt19*-tdt$^+$ cells. (Scale bar: 30 μm). f, Co-staining for SFN (green), *Krt19*-tdt (red) and AGER (grey). White arrows indicate SFN$^+$, *Krt19*-tdt$^+$

735 cells. Yellow arrowhead indicates SFN$^+$, *Krt19*-tdt$^+$, AGER$^+$ cell. (Scale bar: 50 μm). g.
736 Immunostaining for CLDN4 (green), *Krt19*-tdt (red) and LGALS3 (grey). White arrows
737 indicate CLDN4$^+$, *Krt19*-tdt$^+$ cells. Yellow arrowhead demonstrates LGALS3$^+$, *Krt19*-tdt$^+$
738 cell. (Scale bar: 30 μm). e-g Control lungs are shown in left panels and bleomycin day 12
739 injured lungs are shown in right panels. DAPI stains nuclei (blue). White line box in merge
740 image indicates region of single channel images shown in left side.
741
742 **Figure 4. Gene expression signatures and signaling pathways enriched in PATS**
743 a, Heatmap shows marker gene expression of each cell population in alveolar organoids
744 (left) and in LPS-treated murine lungs (right) (scale shows z-score). Table on the right
745 indicates genes enriched in the indicated pathways and cellular processes specifically in
746 PATS. b, KEGG pathways enriched in PATS. Scale shows $\log_2$ (combined score)
747 obtained from web-based tool - Enrichr. c, UMAP rendering of PATS enriched signaling
748 pathways and cellular processes in alveolar organoids (left) and LPS-treated murine
749 lungs (right).
750
751 **Figure 5. PATS undergo stretch-induced DNA damage *in vivo* and *ex vivo***
752 a, Schematic of bleomycin-induced lung injury in *Sftpc-CreER;R26RtdTomato* mice.
753 Immunostaining for b, β-galactosidase staining in bleomycin injured and control lungs. c,
754 γH2AX (green), *Sftpc*-tdt (red) and LGALS3 (grey) in control (upper) and bleomycin-
755 treated lung collected on 12 days after the injury (lower). Inset on the right side in c shows
756 the higher magnification of region indicated by dotted yellow box. DAPI stains nuclei
757 (blue). Scale bar: 30 μm. d, Quantification of γH2AX$^+$ cells in *Sftpc*-tdt$^+$ cells in control and

bleomycin day 10 injured mice. Asterisks indicate $p < 0.0008$ (un-paired student's t-test). Data are from three independent experiments and are presented as mean ± s.e.m. e, Schematic of AEC1 ablation using diphtheria toxin (DT). f, β-galactosidase staining in DT-treated and control lungs. g, Co-staining for γH2AX (red) and LGALS3 (grey) in control (left) and DT-treated lung collected on 6 days after the injury (right). White arrows indicate $γH2AX^+$, $LGALS3^+$ cells. Scale bar 20 μm h, Quantification of $LGALS^+/γH2AX^+$ cells in total $γH2AX^+$ cells in control and DT-treated lung. Asterisk indicates $p = 0.05$. Data are from three independent experiments and are presented as mean ± s.e.m. i, Schematic of alveolar organoid culture. j, Co-staining for *SFTPC*-GFP (green) and γH2AX (red) in alveolar organoids. Inset on the right side shows the higher magnification of region indicated by dotted yellow box. White arrowheads indicate $γH2AX^+$ cells Scale bar: 30 μm. k, Schematic of 2D culture of AEC2. l, Immunostaining for *SFTPC*-GFP (green), γH2AX (red) and AGER (grey) in 2D culture of AEC2.

Figure 6. Transcriptional control of PATS by TP53 signaling

a. Experimental workflow for sequential administration of tamoxifen followed by PBS or bleomycin (d0) administration and Nutlin-3a or DMSO treatment (d8-18) and tissue collection (d20) for analysis using *Sftpc-CreER;R26R-tdTomato* mice. b, Co-staining of SFTPC (green), *Sftpc*-tdt (red) and AGER (grey) in PBS+Nutlin-3a (left panel), bleomycin+DMSO (middle panel) and bleomycin+Nutlin-3a (right panel) treated mice. DAPI stains nuclei (blue). Scale bar: 100 μm. c, Quantification of $AGER^+tdt^+$ cells in total $AGER^+$ normalized to *Sftpc*-tdt labeling efficiency. Asterisk shows $p = 0.036$. Data are from three independent experiments and are presented as mean ± s.e.m. d. Schematic

781 of alveolar organoid culture treated with Nutlin-3a. e, Immunostaining for Ki67 (green)
782 and AGER (grey) in control or Nutlin-3a treated alveolar organoids. DAPI stains nuclei
783 (blue). Scale bar: 30 μm. f, Schematic of bleomycin-induced lung injury in *Ctgf-GFP* mice.
784 g, Distribution of H3K4me3 peaks in PATS marker gene loci in PATS (red line) and
785 homeostatic AEC2 (blue line). i,j, Transcriptional activity of known TP53 target genes
786 (*Mdm2* and *Fas*) in PATS compared to AEC2s. Arrowhead indicates location of predicted
787 TP53 binding motifs. Green-shade regions are promoter or enhancer.
788
789 **Figure 7. Enrichment of PATS-like states in IPF suggests persistence of this state**
790 **in pathological milieu**
791 a, UMAP shows scRNA-seq data from alveolar epithelial cells in healthy and IPF lungs.
792 b, RNA velocity analysis predicts lineage trajectories in alveolar epithelial cell populations.
793 Arrows indicates strong RNA velocities. c, UMAP plots show the expression of indicated
794 genes in healthy and IPF lung scRNA-seq data. d, UMAP plots show enrichment of
795 candidate signaling pathways healthy and IPF lung scRNA-seq data. e, Heatmap showing
796 expression of known target genes of indicated signaling pathways in AEC1, AEC2, and
797 PATS-like state. Scale indicates z-score where red is high, and blue is low. f, Violin plots
798 showing IPF-relevant gene expression in indicated cell types/cell states in control and IPF
799 lungs. g-h, Co-staining for PATS-like markers in human IPF lungs. g, Quadruple
800 immunostaining for SFN (green), HTII-280 (red), KRT17 (grey) and ACTA2 (blue) in IPF
801 lung. White arrows indicate SFN^+, $HTII\text{-}280^+$ cells. Yellow arrowheads demonstrate SFN^+,
802 KRT17, $HTII\text{-}280^+$ cells. h, Quantification of SFN^+ cells in total $HTII\text{-}280^+$ cells. Data are
803 from three independent experiments and are presented as mean ± s.e.m. Asterisks

804 indicate $p < 0.0001$. i, Left panel shows triple immunostaining for SFN (green), HTII-280
805 (red) and AGER (grey) and right panel shows triple immunostaining for SFN (green),
806 CLDN4 (red) and LGALS (grey). j, Quadruple immunostaining for senescence marker p21
807 (green), in combination with SFN (blue), ACTA2 (red) and KRT17 (grey). White arrows
808 show SFN^+, $KRT17^+$, $p21^+$ triple positive cells surrounded by ACTA2 (red) positive cells
809 (left panel). Quadruple immunostaining for γH2AX (green), SFN (blue), HTII-280 (grey)
810 and ACTA2 (red) in IPF lungs (right panel). Inset indicates region of single channel
811 images shown in left side. Scale bars in g-i indicate 100 μm. j, k, β-galactosidase staining
812 in IPF lung. Black arrows indicate X-gal staining in epithelial cells. See Supplementary
813 Fig. 9 for all corresponding immunostainings on control lungs. Scale bar indicates 100
814 μm.
815
816 **Figure 8. Schematic describing alveolar stem cell-mediated epithelial regeneration**
817 **and disease pathogenesis.**
818 Alveolar stem cells replicate in response to damage and generate a novel transitional cell
819 state which normally matures into functional alveolar type-1 epithelial cells. The newly
820 identified transitional state is vulnerable to DNA damage and undergoes a transient
821 senescent state. This novel state is enriched in human fibrotic lungs.
822
823
824 **Supplementary Figure legends**
825 **Supplementary Figure 1. scRNA-seq segregates distinct alveolar population with**
826 **specific markers.**

a, Pearson correlation plot visualizes the number of genes per cell (nGene) and unique molecular identifier (nUMI) in total cells derived from alveolar organoids. b, UMAP shows major cell populations including epithelial cells (green), fibroblasts (red), and some minor populations such as endothelial cells (blue) and macrophages (purple) in alveolar organoids. c, UMAP plots show the expression of indicated genes in epithelial cell populations derived from our alveolar organoid scRNA-seq datasets.

Supplementary Figure 2. scRNA-seq analysis revealed the emergence of a novel epithelial cell population after LPS-induced lung injury *in vivo*.

a, UMAP shows homeostatic (red) and LPS-treated (blue) lung alveolar epithelial cells. b, Distinct cell populations in alveolar epithelial cells in control and LPS-treated lungs. c, UMAP plots show the expression of indicated genes in alveolar epithelial cells in control and LPS-treated lungs.

Supplementary Figure 3. Specific markers are expressed in the novel alveolar population in the lungs following bleomycin treatment *in vivo*.

a, Schematic of bleomycin-induced lung injury in *Ctgf*-GFP mice. b, Immunostaining for *Ctgf*-GFP (green), KRT8 (red) and SFTPC (grey) in control lung (left) and bleomycin day 12 injured mice (right). Magnified single channel images are shown on the right. White line box indicates magnified region. c, Quantification of signal intensity of KRT8 immunostaining in control (black circles) and experimental lung 12 days after bleomycin injury (black rectangles); $p < 0.0001$; Wilcoxon rank sum test. d, Quantification of $KRT8^{lo}$, $KRT8^{hi}$ and $KRT8^{hi}/GFP^{+}$ subpopulations of $SFTPC^{+}$ cells in control lung (white bars) and

lung 12 days after bleomycin injury (gray bars). Error bars, mean ± s.e.m (n = 3). KRT8^{lo}- P=1 X 10^{-5}; KRT8^{hi}- P=1 X 10^{-4}; KRT8^{hi} *Ctgf*-GFP^{+}- P=2.6 X 10^{-3}. e, Immunostaining for *Ctgf*-GFP (green), KRT8 (red) and Ki67 (grey), e, for *Ctgf*-GFP (green), Ki67 (red) and SFTPC (grey) on bleomycin-treated lungs collected on 12 days post injury. DAPI stains nuclei (blue). Scale bars indicate 20 μm.

Supplementary Figure 4. Expression of markers specific to the novel alveolar population in AEC1-specific ablation mouse model.

a-d, Co-immunostaining for a) SFN (green) and LGALS3 (grey) or b) SFTPC (green) and LGALS3 (grey) or c) SFN (green) and KRT19 (red) or d) LGALS3 (green) and Ki67 (red) in control (left) and AEC1-ablated lungs (right). DAPI stains nuclei (blue). Scale bars indicate 20 μm.

Supplementary Figure 5. AEC2-derived cells are the cell-of-origin for the novel alveolar cell state following bleomycin-induced lung injury.

a, Schematic of AEC2 lineage tracing using *Sftpc-CreER;R26R-tdTomato* mice. Tamoxifen was given 2 weeks prior to PBS or bleomycin administration followed by tissue harvest. b, Immunostaining for KRT8 (green), *Sftpc*-tdt (red) and SFTPC (grey). c, Co-staning for KRT8 (green), *Sftpc*-tdt (red) and CLDN4 (grey) in control (upper) and bleomycin-treated lungs (lower). White boxed insets are shown on the right in each panel. DAPI stains nuclei (blue). Scale bars indicate 50 μm.

Supplementary Figure 6. Lineage tracing revealed that PATS differentiate into AEC1 following bleomycin-induced alveolar injury *in vivo*.

a, Immunostaining for SFN (green), *Krt19*-tdt (red) and AGER (grey) in control (upper) and bleomycin-treated lungs (lower). Scale bars 100 μm. b, Quantification of *Krt19*-tdt$^+$ cells in total SFTPC$^+$ cells in injured regions. Asterisk shows $p < 0.031$. (Mann-Whitney test) Data are from three independent experiments and are presented as mean ± s.e.m. c, Quantification of *Krt19*-tdt$^+$ cells in total AGER$^+$ cells. In left image the quantification strategy is shown. Cells were identified based on co-localization of *Krt19*-tdt, AGER and DAPI (yellow doted lines) or AGER and DAPI (green dotted lines). Graph shows the percentage of KRT19$^+$, AGER$^+$ cells in total AGER$^+$ cells in control and bleomycin injured mice on day-12. Data are from three independent experiments and are presented as mean ± s.e.m. Asterisk shows $p < 0.0001$. d, Immunostaining for CLU (green)/*Krt19*-tdTomato (red) in bleomycin-treated lungs. DAPI stains nuclei (blue). Scale bars indicate 30 μm.

Supplementary Figure 7. Histone marks profiling revealed active promoters and transcribing genomic loci in AEC2 and PATS following bleomycin-induced lung injury.

Genomic tracks show enrichment of H3K4me3 and H3K36me3 in a), AEC2-specific gene loci (*Sftpc*) and b), PATS-specific gene loci (*Fn1*). H3K4me3 distribution and predicted binding motifs of NF1, ETS1 and ATF3 in c, *Lgals3* and d, *Sox4* loci. Blue and Red tracks indicate homeostatic AEC2 and PATS in bleomycin-induced lung injury, respectively. *y*-axis in AEC2 and PATS in all panels are normalized.

Supplementary Figure 8. scRNA-seq analysis revealed the accumulation of PATS-like cells of human fibrotic lungs.

a, Heatmap shows expression of marker genes of each cell population in human lungs (scale shows z-score). b, UMAP plots show the expression of indicated genes in alveolar epithelial populations in heathy controls and fibrotic human lungs. c, KEGG pathway enrichment analysis shows signaling pathways highly represented in PATS-like cells in human lungs. Scale shows $\log_2$ (combined score) obtained from Enrichr (see methods section for details). d, Violin plots shows the relative gene expression levels of indicated genes and cell types in control and IPF lungs.

Supplementary Figure 9. Markers specific to PATS-like cells are specifically expressed in human fibrotic lungs but not in healthy controls.

a, Hematoxylin and Eosin staining of IPF lung section. Scale bar indicates 200 μm. b-e, Immunostaining for PATS-like markers in healthy human lungs. b, Co-staining for SFN (green), CLDN4 (red) and LGALS3 (grey), c, Immunostaining for SFN (green), HTII-280 (red) and AGER (grey), d, Co-staining for SFN (blue), p21 (green), ACTA2 (red) and KRT17 (grey) and e, Immunostaining for γH2AX (green), SFN (blue), ACTA2 (red) and HTII-280 (grey). White line box in merged images indicate region of single channel images shown on right. In f and g DAPI (blue) stains nuclei. f, Immunostaining for SFN (green), TP63 (blue), HTII-280 (grey) and ACTA2 (red) in healthy (left panel) and IPF lungs (right panel). Scale bars indicate 100 μm.

References

1. Hogan, B. L. M. *et al.* Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15**, 123–138 (2014).
2. Zepp, J. A. & Morrisey, E. E. Cellular crosstalk in the development and regeneration of the respiratory system. *Nat. Rev. Mol. Cell Biol.* (2019) doi:10.1038/s41580-019-0141-3.
3. Scadden, D. T. Nice neighborhood: emerging concepts of the stem cell niche. *Cell* **157**, 41–50 (2014).
4. Hsu, Y.-C., Pasolli, H. A. & Fuchs, E. Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell* **144**, 92–105 (2011).
5. van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).
6. Shivdasani, A. A. & Ingham, P. W. Regulation of stem cell maintenance and transit amplifying cell proliferation by tgf-beta signaling in Drosophila spermatogenesis. *Curr. Biol.* **13**, 2065–2072 (2003).
7. Rock, J. R. *et al.* Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell* **8**, 639–648 (2011).
8. Papayannopoulou, T. & Scadden, D. T. Stem-cell ecology and stem cells in motion. *Blood* **111**, 3923–3930 (2008).
9. Nabhan, A., 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* (2018) doi:10.1126/science.aam6603.

940 10. Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* **123**,
941 3025–3036 (2013).
942 11. Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in lung
943 development, renewal and cancer. *Nature* **507**, 190–194 (2014).
944 12. Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved
945 epithelial progenitor. *Nature* **555**, 251–255 (2018).
946 13. Zepp, J. A. *et al.* Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-
947 Renewal and Myofibrogenesis in the Lung. *Cell* **170**, 1134-1148.e10 (2017).
948 14. Tata, P. R. & Rajagopal, J. Plasticity in the lung: making and breaking cell identity.
949 *Development* **144**, 755–766 (2017).
950 15. Lange, A. W. *et al.* Hippo/Yap signaling controls epithelial progenitor cell proliferation and
951 differentiation in the embryonic and adult lung. *J Mol Cell Biol* **7**, 35–47 (2015).
952 16. Weibel, E. R. Lung morphometry: the link between structure and function. *Cell Tissue Res.*
953 **367**, 413–426 (2017).
954 17. Hogan, B. & Tata, P. R. Cellular organization and biology of the respiratory system. *Nat.*
955 *Cell Biol.* (2019) doi:10.1038/s41556-019-0357-7.
956 18. Frank, D. B. *et al.* Emergence of a Wave of Wnt Signaling that Regulates Lung
957 Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. *Cell Rep* **17**,
958 2312–2325 (2016).
959 19. Chung, M.-I., Bujnis, M., Barkauskas, C. E., Kobayashi, Y. & Hogan, B. L. M. Niche-
960 mediated BMP/SMAD signaling regulates lung alveolar stem cell proliferation and
961 differentiation. *Development* **145**, dev163014 (2018).

962 20. LaCanna, R. *et al.* Yap/Taz regulate alveolar regeneration and resolution of lung
963 inflammation. *J. Clin. Invest.* **129**, 2107–2122 (2019).
964 21. Finn, J. *et al.* Dlk1-Mediated Temporal Regulation of Notch Signaling Is Required for
965 Differentiation of Alveolar Type II to Type I Cells during Repair. *Cell Rep* **26**, 2942-2954.e5
966 (2019).
967 22. Riemondy, K. A. *et al.* Single cell RNA sequencing identifies TGFβ as a key regenerative
968 cue following LPS-induced lung injury. *JCI Insight* **5**, (2019).
969 23. Hall-Glenn, F. *et al.* CCN2/connective tissue growth factor is essential for pericyte adhesion
970 and endothelial basement membrane formation during angiogenesis. *PLoS ONE* **7**, e30562
971 (2012).
972 24. Strunz, M. *et al.* Longitudinal single cell transcriptomics reveals Krt8+ alveolar epithelial
973 progenitors in lung regeneration. *bioRxiv* 705244 (2019) doi:10.1101/705244.
974 25. Rock, J. R. *et al.* Multiple stromal populations contribute to pulmonary fibrosis without
975 evidence for epithelial to mesenchymal transition. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1475-
976 1483 (2011).
977 26. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
978 27. Ward, H. E. & Nicholas, T. E. Alveolar type I and type II cells. *Aust N Z J Med* **14**, 731–734
979 (1984).
980 28. Weibel, E. R. On the tricks alveolar epithelial cells play to make a good lung. *Am. J. Respir.*
981 *Crit. Care Med.* **191**, 504–513 (2015).
982 29. Katsura, H., Kobayashi, Y., Tata, P. R. & Hogan, B. L. M. IL-1 and TNFα Contribute to the
983 Inflammatory Niche to Enhance Alveolar Regeneration. *Stem Cell Reports* **12**, 657–666
984 (2019).

985 30. Xi, Y. *et al.* Local lung hypoxia determines epithelial fate decisions during alveolar
986 regeneration. *Nat. Cell Biol.* **19**, 904–914 (2017).
987 31. Antony, V. B. & Thannickal, V. J. Cellular Senescence in Chronic Obstructive Pulmonary
988 Disease: Multifaceted and Multifunctional. *Am. J. Respir. Cell Mol. Biol.* **59**, 135–136
989 (2018).
990 32. Mandal, P. K., Blanpain, C. & Rossi, D. J. DNA damage response in adult stem cells:
991 pathways and consequences. *Nat. Rev. Mol. Cell Biol.* **12**, 198–202 (2011).
992 33. Raab, M. *et al.* ESCRT III repairs nuclear envelope ruptures during cell migration to limit
993 DNA damage and cell death. *Science* **352**, 359–362 (2016).
994 34. Li, J. *et al.* The Strength of Mechanical Forces Determines the Differentiation of Alveolar
995 Epithelial Cells. *Dev. Cell* **44**, 297-312.e5 (2018).
996 35. Spike, B. T. & Wahl, G. M. p53, Stem Cells, and Reprogramming: Tumor Suppression
997 beyond Guarding the Genome. *Genes Cancer* **2**, 404–419 (2011).
998 36. van Galen, P. *et al.* A Multiplexed System for Quantitative Comparisons of Chromatin
999 Landscapes. *Molecular Cell* **61**, 170–180 (2016).
1000 37. Soldatenkov, V. A. *et al.* Regulation of the human poly(ADP-ribose) polymerase promoter
1001 by the ETS transcription factor. *Oncogene* **18**, 3954–3962 (1999).
1002 38. Zhu, Y. *et al.* Early inactivation of p53 tumor suppressor gene cooperating with NF1 loss
1003 induces malignant astrocytoma. *Cancer Cell* **8**, 119–130 (2005).
1004 39. Fan, F. *et al.* ATF3 induction following DNA damage is regulated by distinct signaling
1005 pathways and over-expression of ATF3 protein suppresses cells growth. *Oncogene* **21**,
1006 7488–7496 (2002).

40. Pan, X. *et al.* Induction of SOX4 by DNA damage is critical for p53 stabilization and function. *Proceedings of the National Academy of Sciences* **106**, 3788–3793 (2009).

41. Nag, S., Qin, J., Srivenugopal, K. S., Wang, M. & Zhang, R. The MDM2-p53 pathway revisited. *J Biomed Res* **27**, 254–271 (2013).

42. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).

43. Zhou, Y. *et al.* Extracellular matrix in lung development, homeostasis and disease. *Matrix Biol.* **73**, 77–104 (2018).

44. Trappmann, B. *et al.* Extracellular-matrix tethering regulates stem-cell fate. *Nat Mater* **11**, 642–649 (2012).

45. Burgess, J. K., Mauad, T., Tjin, G., Karlsson, J. C. & Westergren-Thorsson, G. The extracellular matrix - the under-recognized element in lung disease? *J. Pathol.* **240**, 397–409 (2016).

46. Wasnick, R. M. *et al.* Restored alveolar epithelial differentiation and reversed human lung fibrosis upon Notch inhibition. *bioRxiv* 580498 (2019) doi:10.1101/580498.

47. Fernandez, I. E. & Eickelberg, O. New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis. *Lancet* **380**, 680–688 (2012).

48. Gulati, S. & Thannickal, V. J. The Aging Lung and Idiopathic Pulmonary Fibrosis. *Am. J. Med. Sci.* **357**, 384–389 (2019).

49. Habermann, A. C. *et al.* Single-cell RNA-sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *bioRxiv* 753806 (2019) doi:10.1101/753806.

50. Adams, T. S. *et al.* Single Cell RNA-seq reveals ectopic and aberrant lung resident cell populations in Idiopathic Pulmonary Fibrosis. *bioRxiv* 759902 (2019) doi:10.1101/759902.

51. Cheng, D. *et al.* Airway epithelium controls lung inflammation and injury through the NF-kappa B pathway. *J. Immunol.* **178**, 6504–6513 (2007).

52. McConnell, A. M. *et al.* p53 Regulates Progenitor Cell Quiescence and Differentiation in the Airway. *Cell Reports* **17**, 2173–2182 (2016).

53. Peng, X. *et al.* SOX4 contributes to TGF-β-induced epithelial-mesenchymal transition and stem cell characteristics of gastric cancer cells. *Genes Dis* **5**, 49–61 (2018).

54. Zhang, J. *et al.* SOX4 induces epithelial-mesenchymal transition and contributes to breast cancer progression. *Cancer Res.* **72**, 4597–4608 (2012).

55. Tiwari, N. *et al.* Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. *Cancer Cell* **23**, 768–783 (2013).

56. Zhang, M. *et al.* Chop deficiency prevents UUO-induced renal fibrosis by attenuating fibrotic signals originated from Hmgb1/TLR4/NFκB/IL-1β signaling. *Cell Death Dis* **6**, e1847 (2015).

57. Lipson, K. E., Wong, C., Teng, Y. & Spong, S. CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis. *Fibrogenesis Tissue Repair* **5**, S24 (2012).

58. Kropski, J. A., Lawson, W. E., Young, L. R. & Blackwell, T. S. Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. *Disease Models & Mechanisms* **6**, 9–17 (2013).

59. da Silva, A. L. G. *et al.* Evaluation of DNA damage in COPD patients and its correlation with polymorphisms in repair genes. *BMC Med. Genet.* **14**, 93 (2013).

60. Sauler, M. *et al.* The DNA repair transcriptome in severe COPD. *Eur. Respir. J.* **52**, (2018).

61. McDonough, J. E. *et al.* A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. *Respir Res* **19**, (2018).

62. Muñoz-Espín, D. & Serrano, M. Cellular senescence: from physiology to pathology. *Nat. Rev. Mol. Cell Biol.* **15**, 482–496 (2014).

63. McGregor, A. L., Hsia, C.-R. & Lammerding, J. Squish and squeeze—the nucleus as a physical barrier during migration in confined environments. *Current Opinion in Cell Biology* **40**, 32–40 (2016).

64. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell* **40**, 179–204 (2010).

65. Mercado, N., Ito, K. & Barnes, P. J. Accelerated ageing of the lung in COPD: new concepts. *Thorax* **70**, 482–489 (2015).

66. Gorgoulis, V. *et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813–827 (2019).

67. Storer, M. *et al.* Senescence Is a Developmental Mechanism that Contributes to Embryonic Growth and Patterning. *Cell* **155**, 1119–1130 (2013).

68. Means, A. L., Xu, Y., Zhao, A., Ray, K. C. & Gu, G. A CK19CreERT knockin mouse line allows for conditional DNA recombination in epithelial cells in multiple endodermal organs. *genesis* **46**, 318–323 (2008).

69. Arenkiel, B. R. *et al.* Activity-induced remodeling of olfactory bulb microcircuits revealed by monosynaptic tracing. *PLoS ONE* **6**, e29423 (2011).

70. 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).

71. Lo, B., Hansen, S., Evans, K., Heath, J. K. & Wright, J. R. Alveolar Epithelial Type II Cells Induce T Cell Tolerance to Specific Antigen. *The Journal of Immunology* **180**, 881–888 (2008).

72. Chung, M.-I. & Hogan, B. L. M. ***Ager-CreER*** ^{T2} : A New Genetic Tool for Studying Lung Alveolar Development, Homeostasis, and Repair. *Am J Respir Cell Mol Biol* **59**, 706–712 (2018).

73. Buch, T. *et al.* A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nat Methods* **2**, 419–426 (2005).

74. Basak, O. *et al.* Mapping early fate determination in Lgr5+ crypt stem cells using a novel Ki67-RFP allele. *The EMBO Journal* **33**, 2057–2068 (2014).

75. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

76. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

77. Manno, G. L. *et al.* RNA velocity of single cells. *Nature* **560**, 494 (2018).

78. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–W97 (2016).

79. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

80. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).

81. Girardot, C., Scholtalbers, J., Sauer, S., Su, S.-Y. & Furlong, E. E. M. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics* **17**, 419 (2016).

82. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

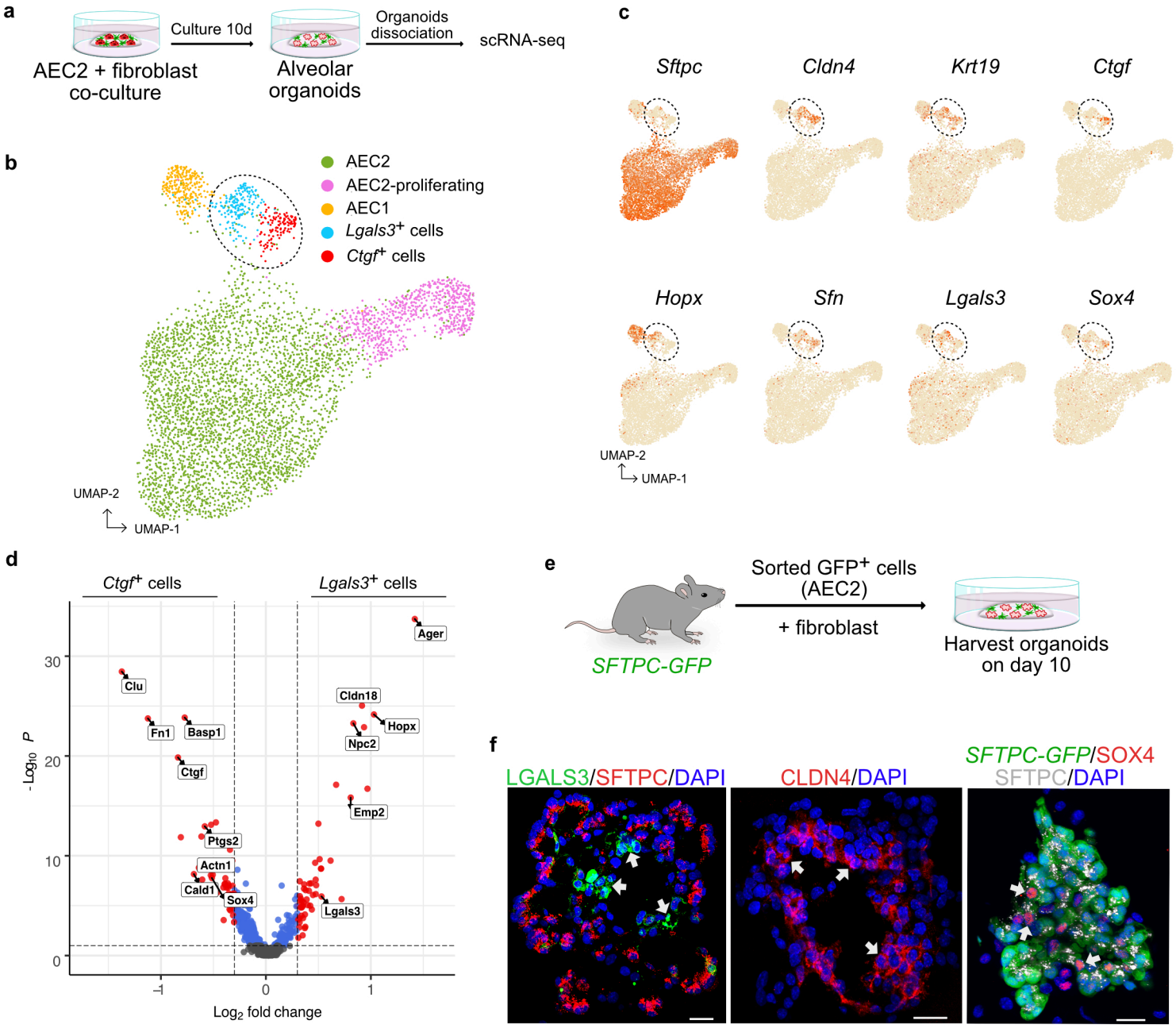
83. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

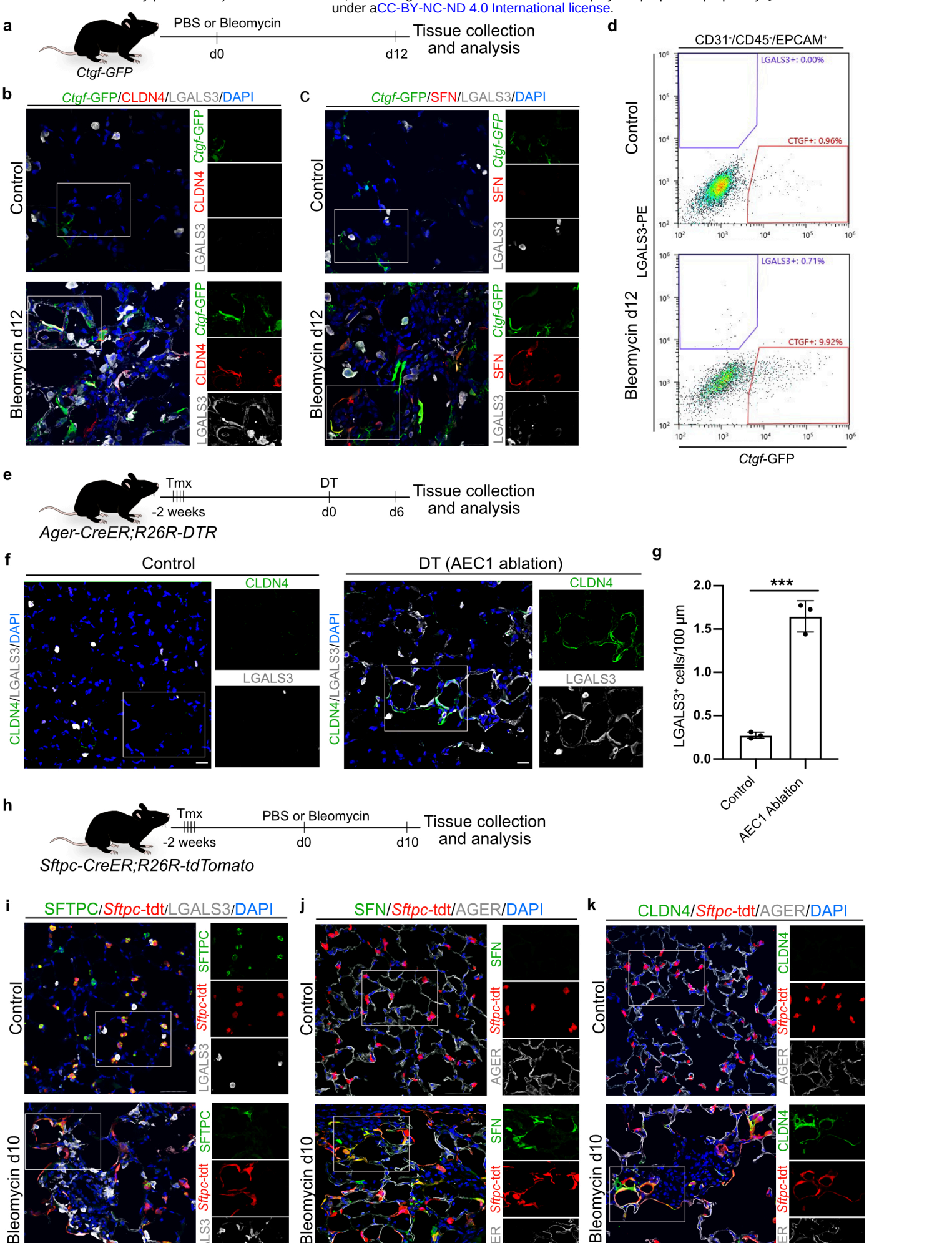
84. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1097276510003667?token=EB605C1A491037738AFFA53843A087D06F83E897B9D534ACF84F30AAC2FDC25A8447DCF9ADA1E016AB8105B69BA43AF9 doi:10.1016/j.molcel.2010.05.004.

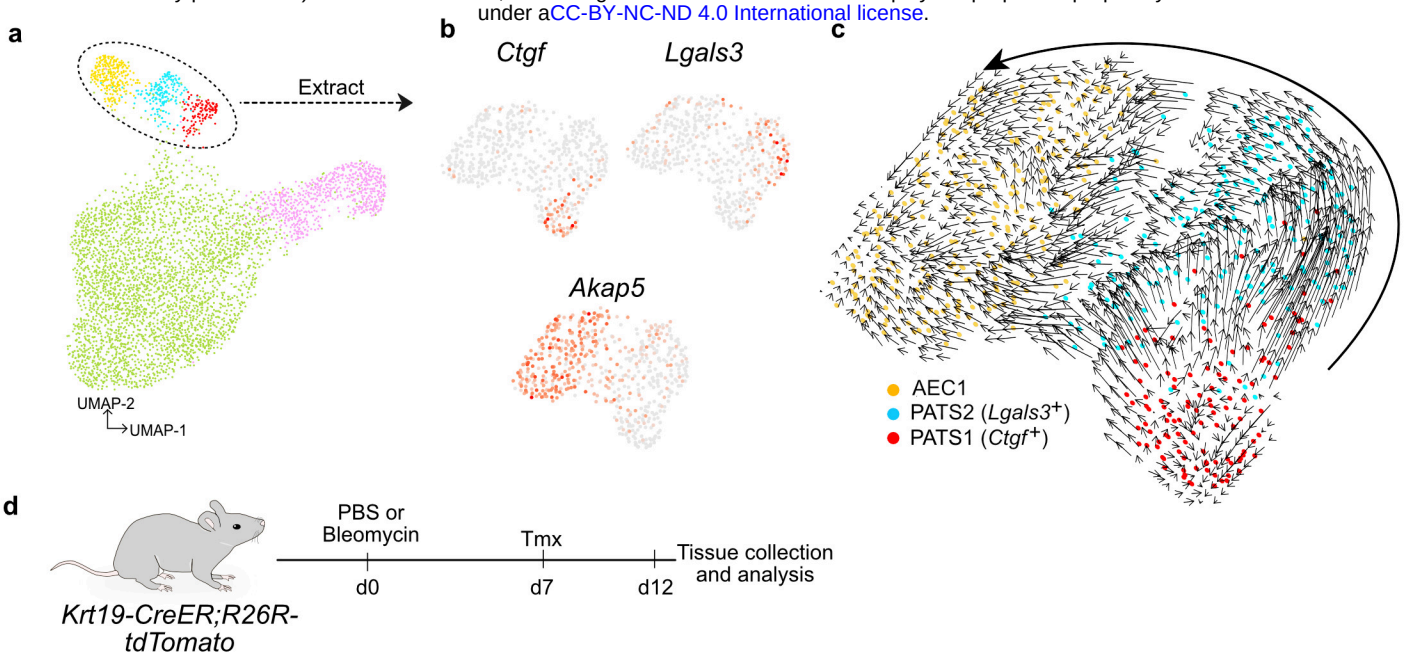
85. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31–e34 (2017).

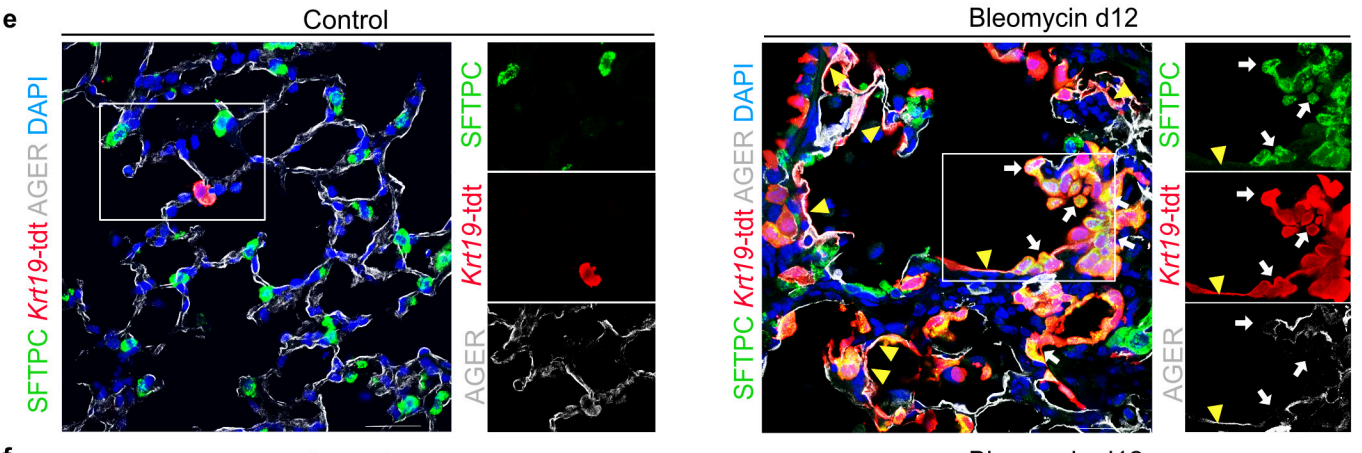
86. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).

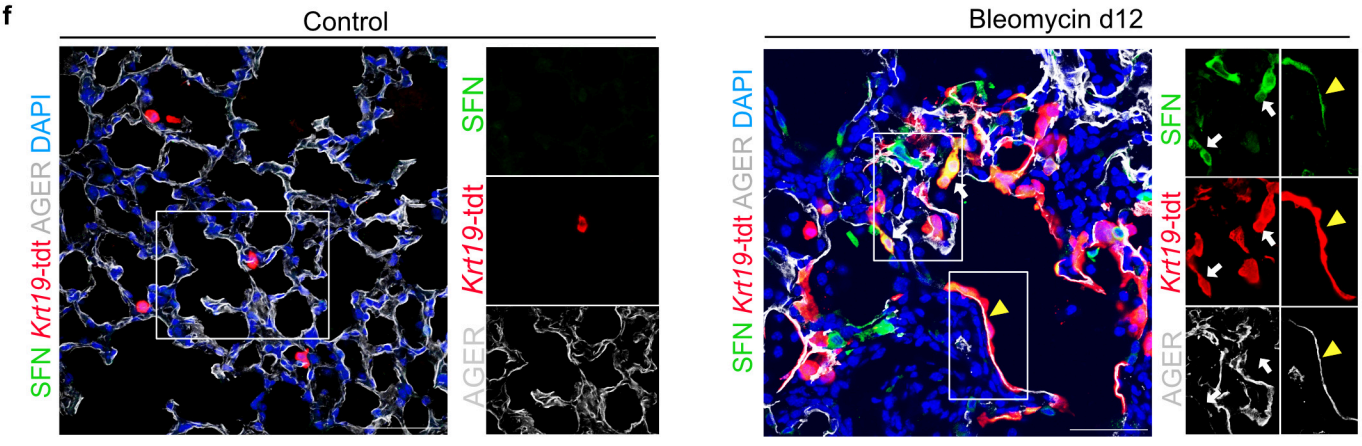
87. Knudsen, L., Weibel, E. R., Gundersen, H. J. G., Weinstein, F. V. & Ochs, M. Assessment of air space size characteristics by intercept (chord) measurement: an accurate and efficient stereological approach. *Journal of Applied Physiology* **108**, 412–421 (2010).











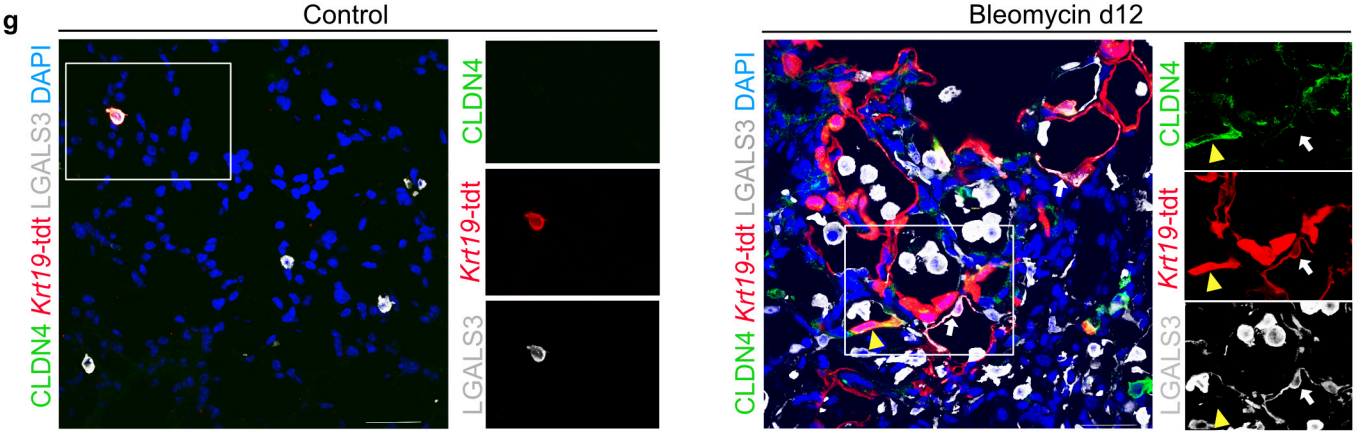
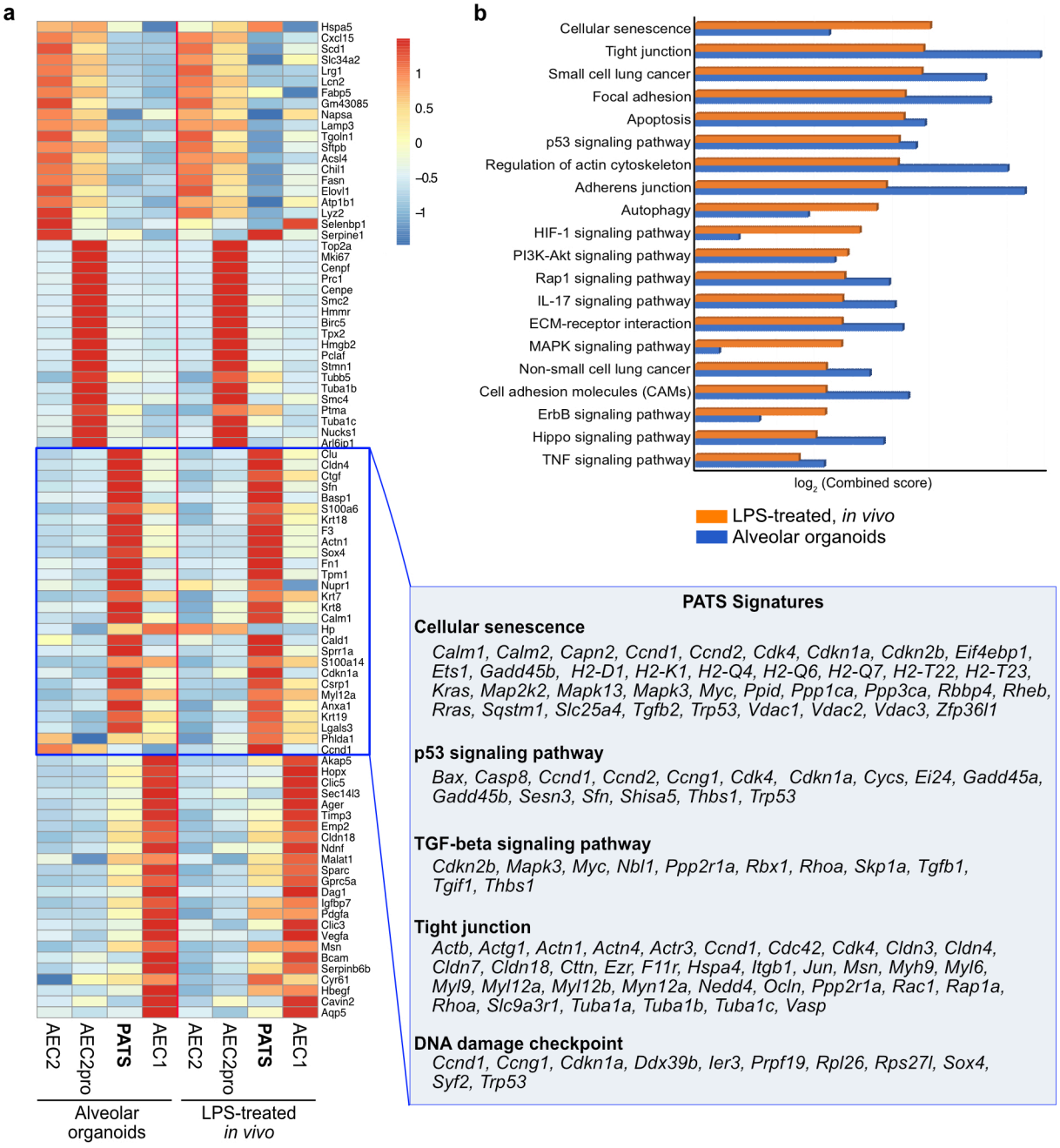
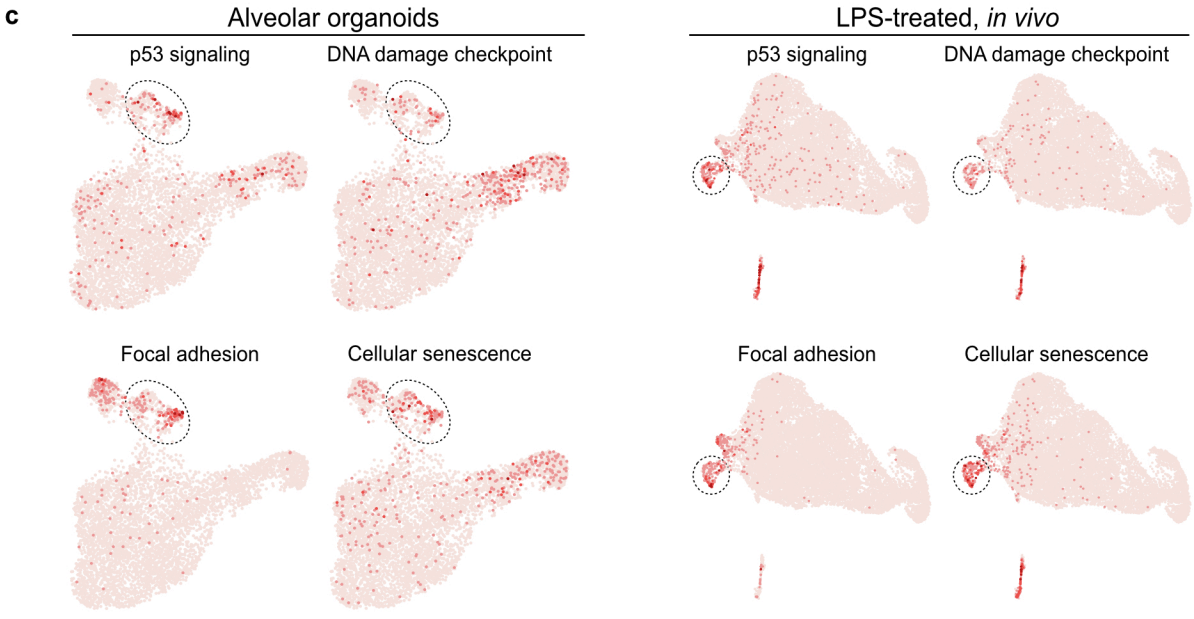
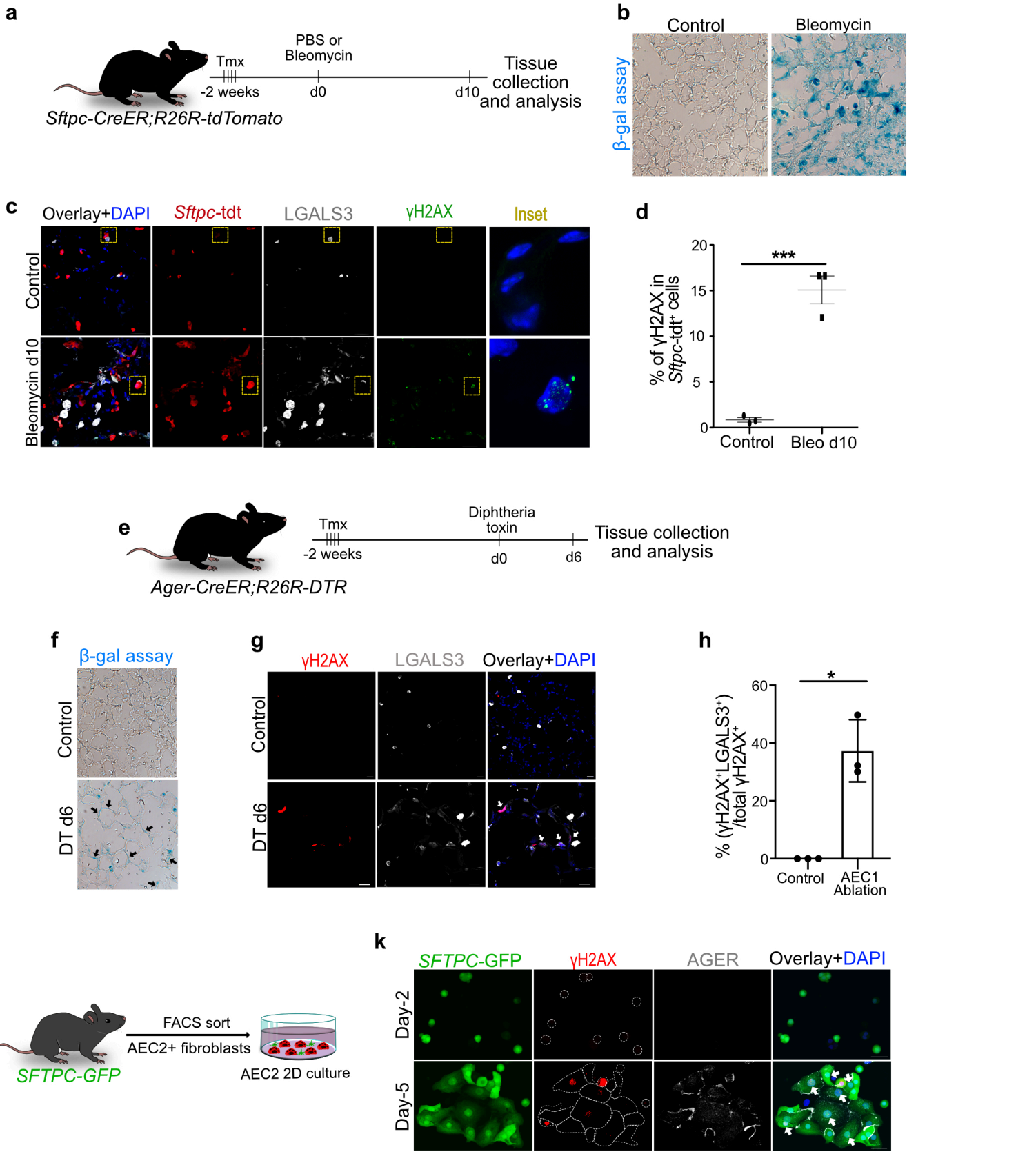
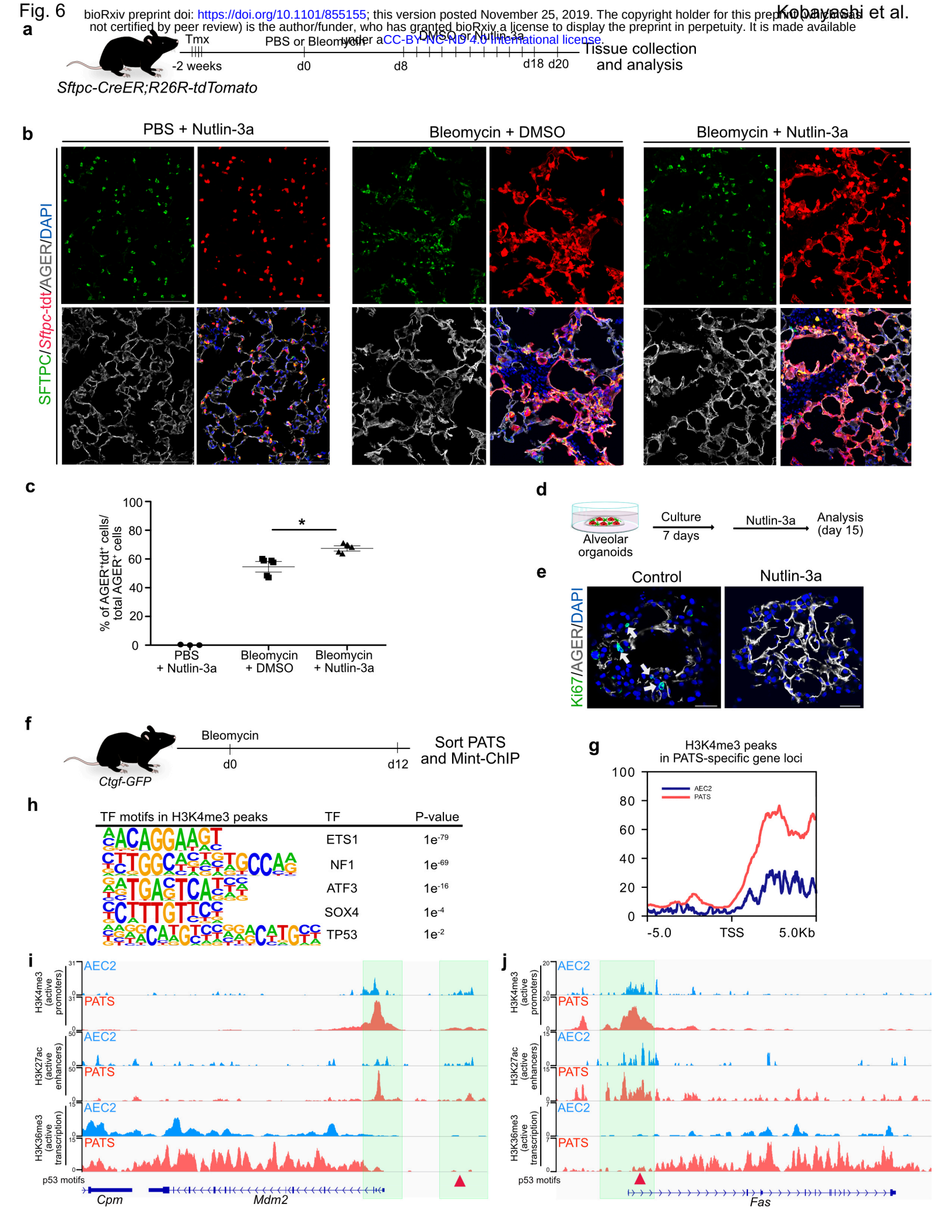


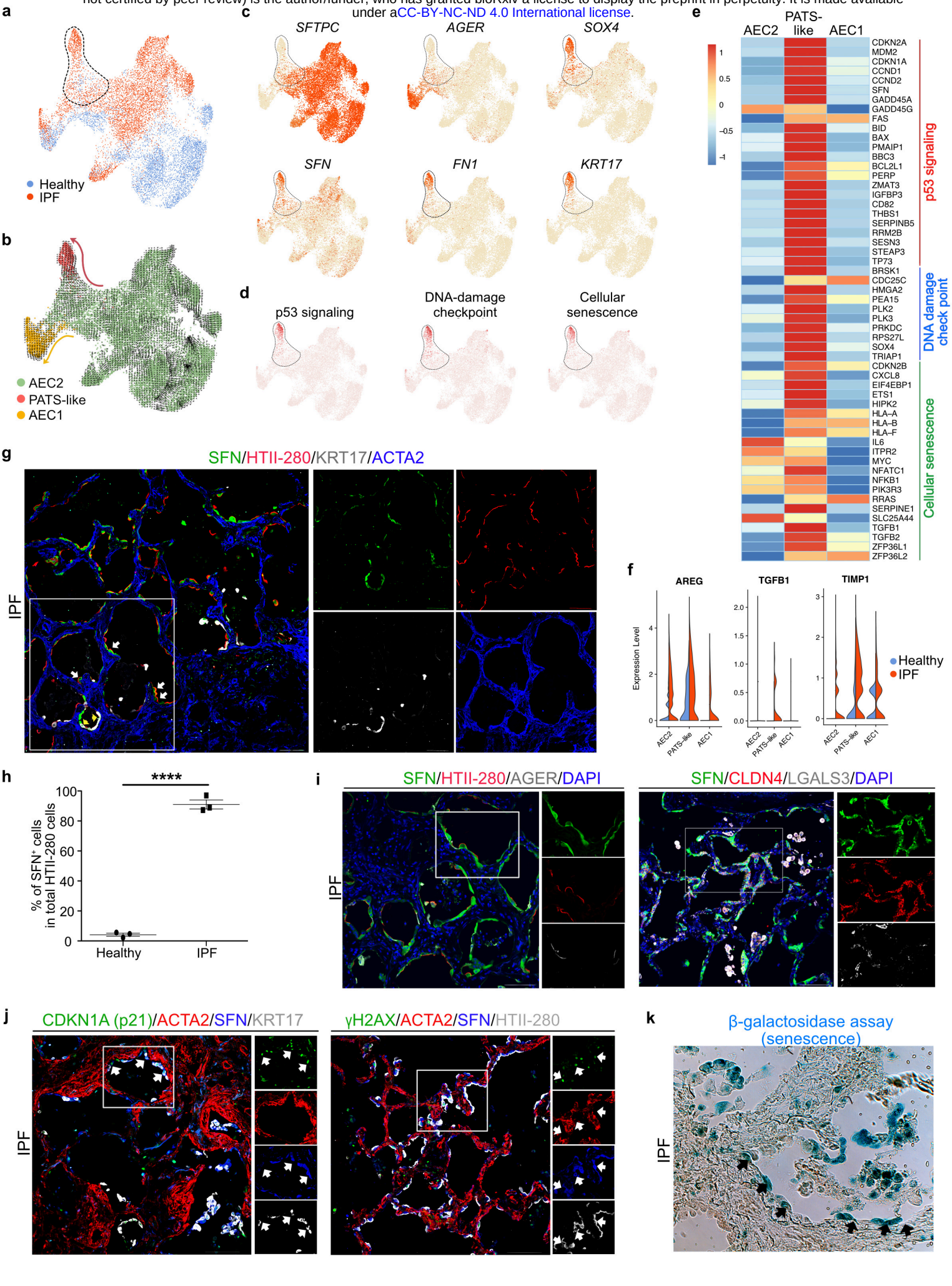
Fig. 4

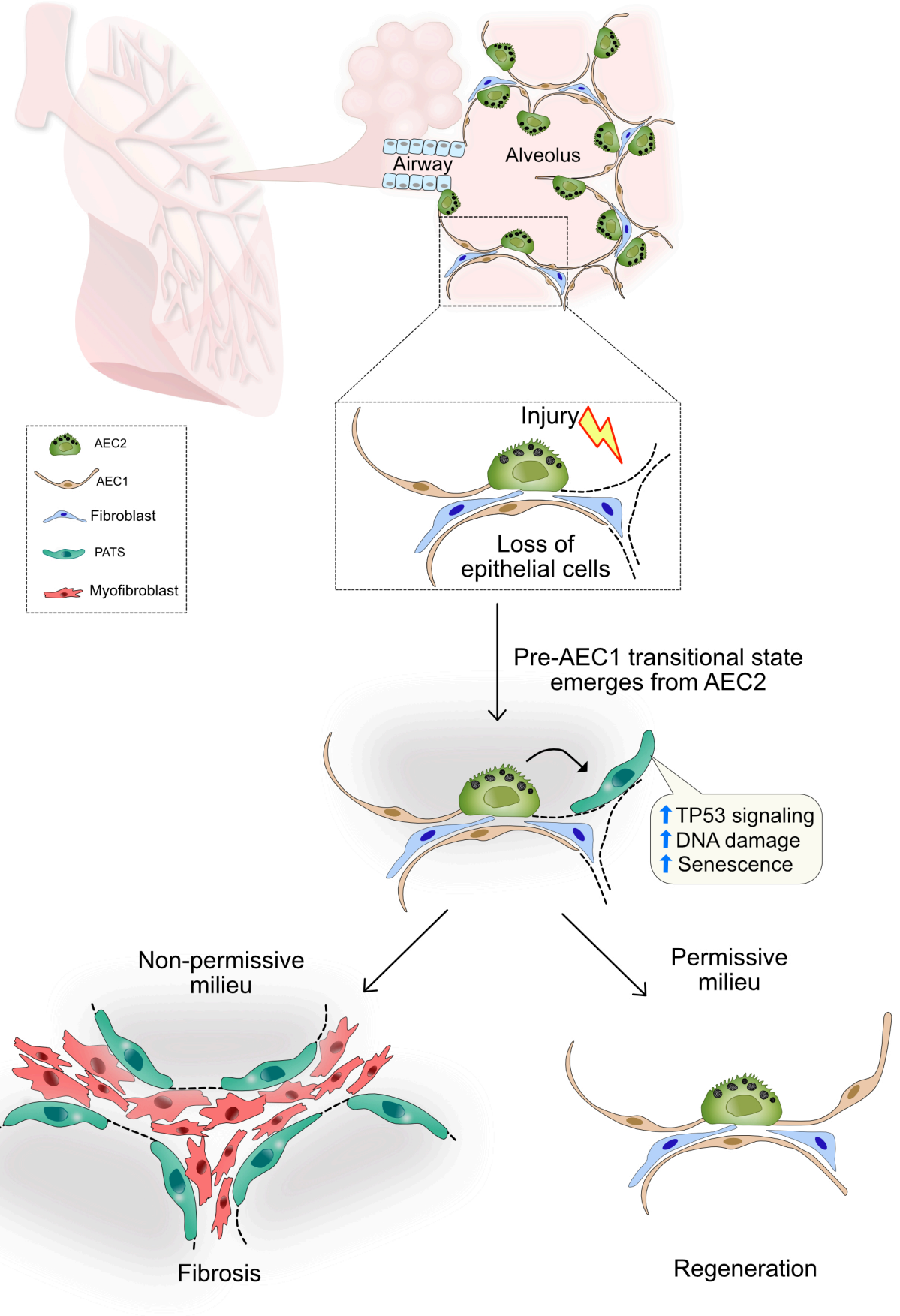


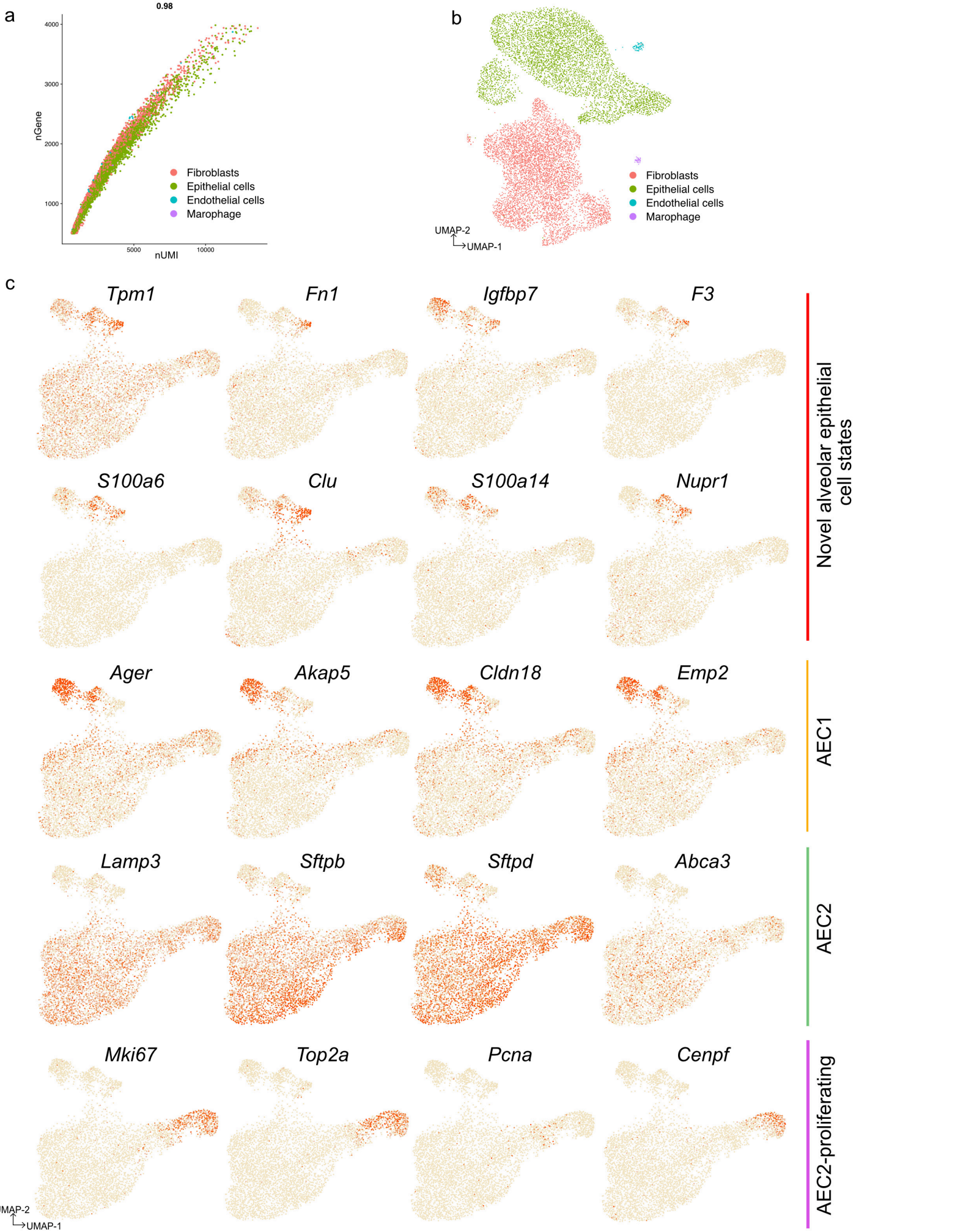


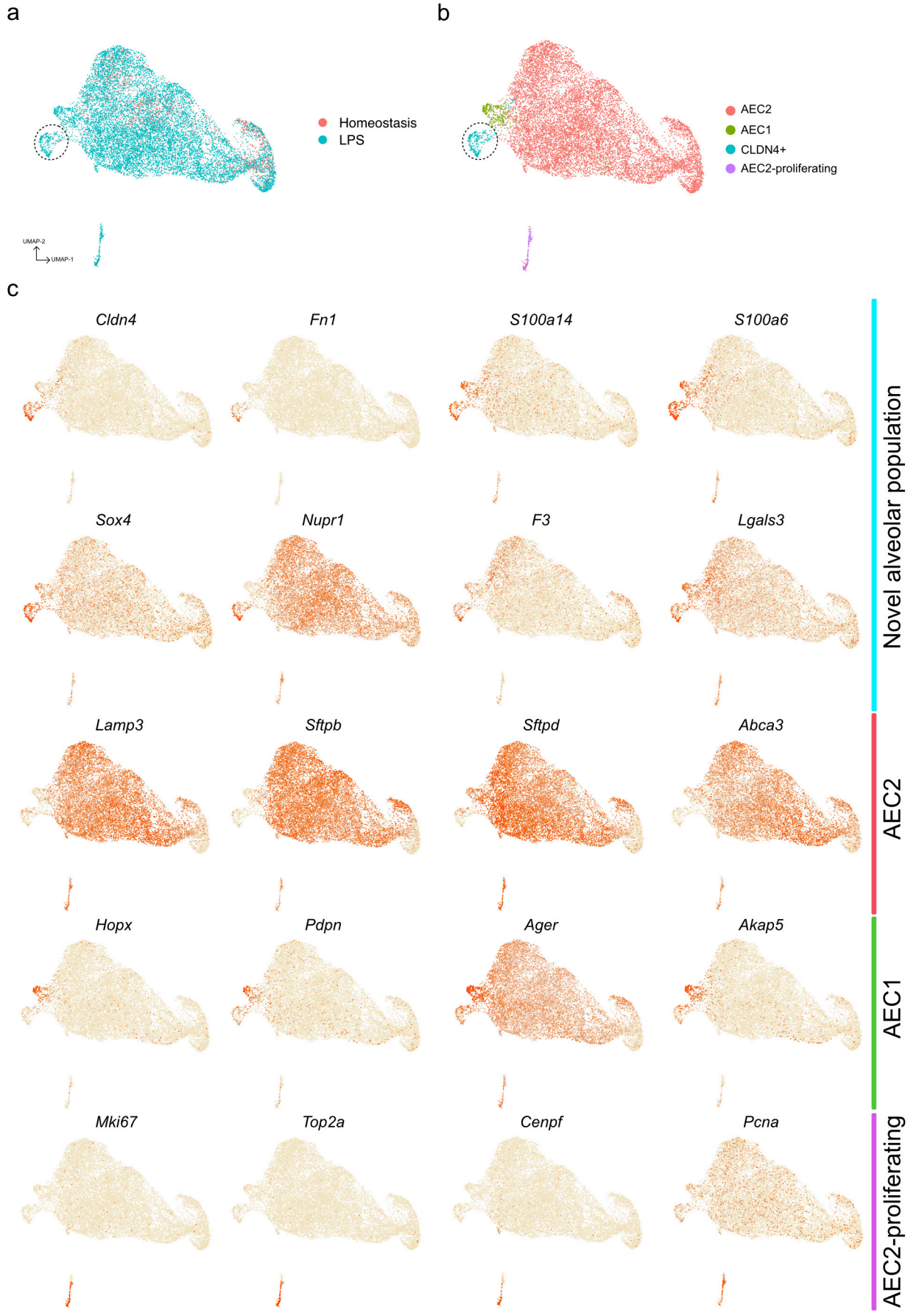


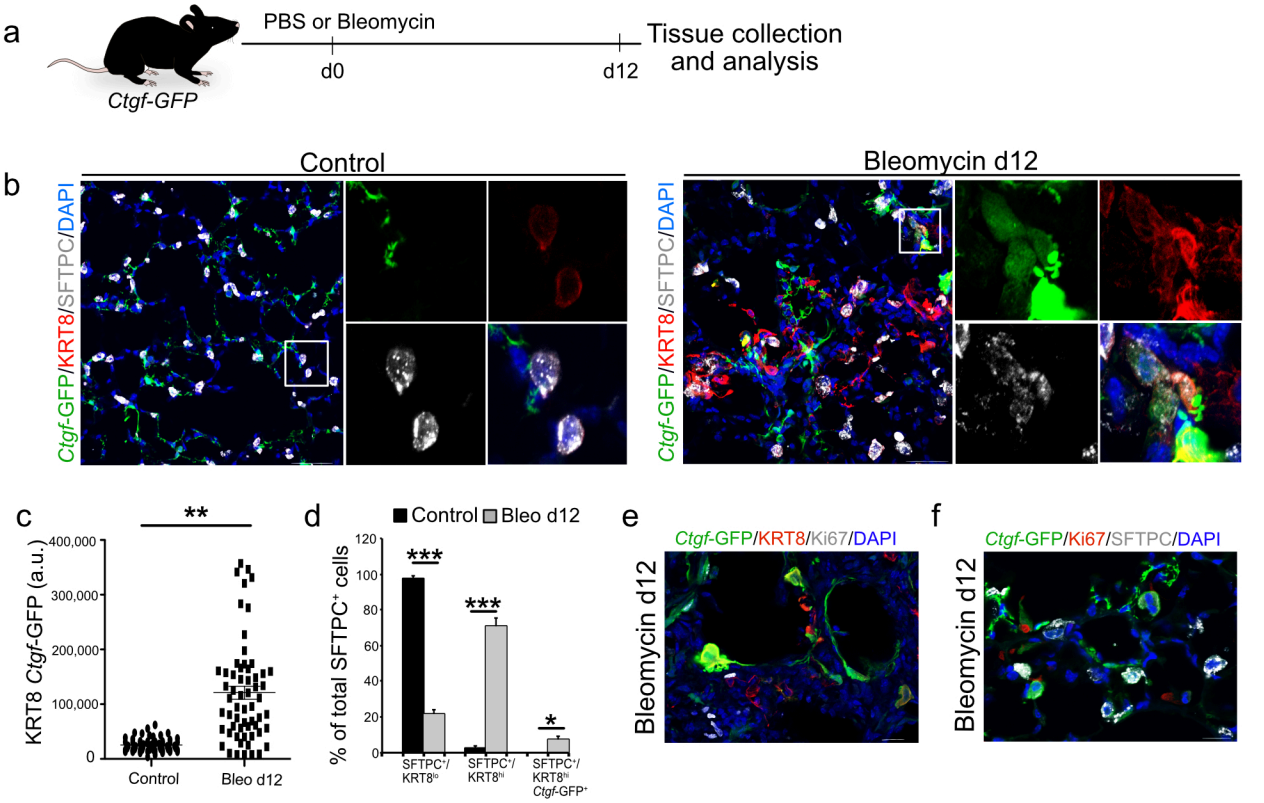





Airway
Alveolus
Injury
Loss of epithelial cells
AEC2
AEC1
Fibroblast
PATS
Myofibroblast
Pre-AEC1 transitional state emerges from AEC2
TP53 signaling
DNA damage
Senescence
Non-permissive milieu
Permissive milieu
Fibrosis
Regeneration






a
Ctgf-GFP
PBS or Bleomycin
d0
d12
Tissue collection and analysis
b
Control
Bleomycin d12
Ctgf-GFP/KRT8/SFTPC/DAPI
c
KRT8 Ctgf-GFP (a.u.)
400,000
300,000
200,000
100,000
0
**
Control
Bleo d12
d
Control
Bleo d12
% of total SFTPC+ cells
120
100
80
60
40
20
0
***
***
*
SFTPC+/KRT8lo
SFTPC+/KRT8hi
SFTPC+/KRT8hi Ctgf-GFP+
e
Ctgf-GFP/KRT8/Ki67/DAPI
Bleomycin d12
f
Ctgf-GFP/Ki67/SFTPC/DAPI
Bleomycin d12