Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis

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

[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]Regeneration Next, Duke University, Durham, NC 27710, USA

[4]Translational Genomics Research Institute, Phoenix, AZ 85004, USA

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

[6]Department of Veterans Affairs Medical Center, Nashville, TN 37212, USA

[7]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37212, USA

[8]Duke Cancer Institute, Duke University School of Medicine, Durham, NC 27710, USA

Abstract

Stem cells undergo dynamic changes in response to injury to regenerate lost cells. However, the identity of transitional states and the mechanisms that drive their trajectories remain understudied. Using lung organoids, multiple *in vivo* repair models, single-cell transcriptomics, and lineage tracing, we find that alveolar type-2 epithelial cells undergoing differentiation into type-1 cells acquire pre-alveolar type-1 transitional cell state (PATS) en route to terminal maturation. Transitional cells undergo extensive stretching during differentiation, making them vulnerable to DNA damage. PATS show an enrichment of TP53, TGFβ, DNA-damage response signalling, and cellular senescence. Gain and loss of function and genomic binding assays revealed a direct transcriptional control of PATS by TP53 signalling. Notably, PATS-like cells were accumulated in human fibrotic lungs, suggesting persistence of the transitional state in fibrosis. Our study thus

implicates a transient state associated with senescence in normal epithelial tissue repair and its abnormal persistence in disease conditions.

Introduction

Adult stem cells undergo dynamic changes in response to tissue damage[1,2]. These changes include resurgence from a quiescent or poised state, onset of proliferation, activation of new gene expression and a return to homeostasis. 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. Studies of stem cells 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 including DNA repair and senescence. Here, we explore this question in relation to epithelial repair in gas-exchange alveolar region of the 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[1,3–10]. Recent studies have identified a subset of AEC2s enriched for active Wnt signalling and have higher "stemness" compared to Wnt-inactive AEC2s. Such differences in alveolar progenitor cell subsets have been attributed to differences in microenvironmental signals[3,6]; in this case, to the vicinity of PDGFRα-expressing fibroblasts which produce ligands to activate Wnt signalling in AEC2s. Recent studies have also implicated other signalling 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[11–13]. 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. Additionally, 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 previously unknown, distinct cell states encompassing the transition between AEC2s and AEC1s. Moreover, murine lineage tracing, coupled with injury-repair models, have revealed the existence of similar transition states *in vivo*. Our study reveals signalling pathways that control these transition states. We also discovered that these transitional states exhibit DNA damage responses, and express senescence-related genes en route to AEC1. Mechanistically, use of genetic loss of function, pharmacological gain of function, and genomic binding assays revealed a direct transcriptional control of PATS by TP53 signalling. Importantly, these transitional states correlate with abnormal epithelial cells associated with defective fibrotic foci in human lungs with progressive pulmonary fibrosis.

Results

Single-cell transcriptomics reveals previously unknown 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[5]. Moreover, AEC2s spontaneously generate AEC1s in alveolar organoids[4]. However, the molecular mechanisms and transitional cell states underlying the differentiation of AEC2 into AEC1 are poorly understood. To address these questions, we purified AEC2s and PDGFRα+ fibroblasts to set up alveolar organoids (Extended Data 1a). Single-cell transcriptome analysis was performed on cells isolated from day-10 organoid culture (Fig. 1a). Uniform Manifold Approximation and Projection (UMAP) identified two major clusters, consisting of *Epcam*+ epithelial cells and *Vim+Pdgfra+* fibroblasts (Extended Data 1b). Next, we further deconvoluted and visualized epithelial cell populations and compared them with publicly available scRNA-seq dataset from lipopolysaccharide (LPS) injured alveolar epithelial cells[14]. Organoid-derived epithelial cells overlapped with their *in vivo* counterparts, revealing transcriptional and cell state similarities (Extended Data 1c). Within the organoid-derived epithelial cells, we observed multiple sub-clusters: cells expressing *Sftpc,* a marker for AEC2s, cells expressing *Ager*, a marker for AEC1s, and *Sftpc+Mki67+* proliferating AEC2s (Fig. 1b). Additionally, we identified a population of alveolar epithelial cells expressing *Cldn4*, *Krt19*, and *Sfn* (Fig. 1b,c; Extended Data 1d). The marker genes unique to this cluster showed two distinct patterns when visualized in UMAP and volcano plots (Fig.1 b–d). One subset (*Ctgf+* cells) is enriched for *Ctgf, Clu, Sox4 and Actn1* while the other (*Lgals3+* cells) is enriched for *Lgals3, Csrp1, S100a14, and Cldn18* (Fig.1 b–d; Extended Data 1d). Additional transcripts enriched in the *Lgals3+* sub-cluster including *Ager, Emp2,* and *Hopx,* markers of AEC1, suggest resemblances and a potential lineage hierarchy between *Lgals3+* cells and AEC1 (Fig. 1d). These data suggest that the *Cldn4+Krt19+Sfn+* population is an intermediate between AEC2 and AEC1. We therefore termed this population "pre-alveolar type-1 transitional cell state" (PATS). To validate our single cell data, we performed immunofluorescence for PATS markers on alveolar organoids (Fig. 1e). This confirmed the presence of cells expressing CLDN4, LGALS3, and SOX4 in alveolospheres (Fig. 1f). Taken together, these data identified unique 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, we rendered scRNA-seq data from LPS-treated and control mouse lungs in UMAP plots and found a population that is unique to LPS injury. Significantly, this population is enriched for genes expressed in PATS, including *Cldn4, Sox4, Lgals3,* and *Fn1* (Extended Data 1e). To further characterize PATS, we utilized *Ctgf-GFP* transgenic mouse line[15]. We exposed these mice to bleomycin, a drug causing transient fibrosis and collected lungs on day-12 post injury (Fig. 2a). Immunofluorescence for GFP in uninjured *Ctgf-GFP* mice revealed GFP signal specifically in fibroblasts and not in alveolar epithelial cells (Fig. 2b,c; Extended Data 2a,b). By contrast, in bleomycin injured lungs, we found GFP expression in epithelial cells co-labeled by PATS markers including CLDN4, LGALS3 and SFN (Fig. 2b,c). We also observed a small fraction of GFP+ cells co-

expressing low levels of AEC2 marker, likely due to perdurance of SFTPC protein as AEC2s transition to PATS (Extended Data 2b,c). A recent study found elevated levels of KRT8 in alveolar epithelial cells after bleomycin-induced injury[16]. Therefore, we tested the expression of KRT8 in bleomycin injured lungs. We found a significant increase in overlap between KRT8 and *Ctgf*-GFP-expressing cells (Extended Data 2b). These data further corroborate with elevated levels of *Krt8* expression in scRNA-seq data from organoids and LPS injury (Extended Data 2d). Taken together, our data reveal that PATS cells 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 followed by diphtheria toxin (DT), resulting in selective ablation of AEC1 cells. We collected samples at day-6 and day-28 after DT injection and performed immunostaining (Fig. 2d, Extended Data 3a). Interestingly, we found cells expressing CLDN4, SFN, KRT19, and LGALS3 with elongated morphology in AEC1 ablated lungs on day-6 but not in controls (Fig. 2e,f; Extended Data 3b–e). We also observed low levels of SFTPC expression in elongated cells that co-expressed PATS markers, likely due to perdurance of SFTPC during AEC2s transition to PATS (Extended Data 3c). Of note, we did not find any PATS markers in AEC1 ablated lungs on day-28, suggesting that PATS had matured into fully differentiated AEC1 (Extended Data 3b–e). These data suggest emergence of PATS is a general mechanism in alveolar regeneration.

Lineage tracing reveals that PATS originate from AEC2

Our above data from organoids and injury models suggested that PATS cells originate from AEC2s. To empirically test this hypothesis, we used a *Sftpc-creER;R26R-tdTomato* mouse model, in which tamoxifen administration permanently induces tdTomato expression specifically in AEC2 and their descendants (Fig. 2g)[17]. On day-10 post bleomycin administration, we observed lineage-labeled (tdt+) cells co-expressing LGALS3 (Fig. 2h), SFN (Fig. 2i) and CLDN4 (Fig. 2j) in damaged alveolar regions but not in control lungs (Fig. 2h–j). As expected, we observed tdt+ cells co-expressing AGER that showed flat, thin, and elongated morphology at late times, confirming that AEC2s transition to AEC1. We detected similar PATS marker expression in AEC2-lineage labeled mice during pneumonectomy induced "physiological" regeneration (Extended Data 3f–h). Interestingly, lineage-labeled cells co-expressing PATS markers, including SFN+, CLDN4+, and LGALS3+, had distinct morphological features (rounded, "stretched" and elongated) during the repair process (Fig. 2i–k). To quantify this, we scored lineage positive cells for expression of CLDN4+ and/or LGALS3+ and according to their shape (Fig. 2k,l). We observed three distinct patterns of marker expression: CLDN4+, CLDN4+LGALS3+, and LGALS3+ and found that these correlated with round, stretched, and elongated morphologies, respectively (Fig. 2l,m). Moreover, we observed significant enrichment of CLDN4+ cells on day-8, CLDN4+LGALS3+ on day-10, and LGALS3+ on day-12 following bleomycin-induced injury (Fig. 2l). We noticed CLDN4+ and CLDN4+LGALS3+ cell states in our scRNA-seq dataset but not a distinct LGALS3+ population. We speculate that the discrepancy between the expression of LGALS3 in scRNA-seq versus immunostaining is

due to perdurance of LGALS3 protein in the terminal AEC1s. Taken together, our analysis revealed that AEC2s differentiation into AEC1 goes through two distinct states (Fig. 2k–n).

Lineage tracing analysis reveals PATS cells generate AEC1

Multiple lines of data from the above experiments suggest that PATS are en route to AEC1. To test this, we first used an algorithm tool called Velocyto[18], which allows the prediction of cell differentiation trajectories based on ratios between spliced and unspliced mRNA. We observed strong RNA velocities (trajectory) originating from the *Ctgf*⁺ population to AEC1 through the *Lgals3*⁺ population (Fig. 3a–c). We then used a CreER allele of *Krt19*, a gene enriched in PATS, combined wit*h R26R-tdTomato* (*Krt19-tdt*) to carry out lineage tracing after bleomycin injury to test whether the same trajectory occurs in regenerating alveoli *in vivo*. First, mice were injured with bleomycin, then 7 days later tamoxifen was administered to label *Krt19*-expressing cells and their progeny (Fig. 3d). We then performed immunofluorescence for AEC2, AEC1, and PATS markers (Fig. 3e–g; Extended Data 4a–d). We found numerous *Krt19*-tdt⁺ cells co-expressing AGER, a marker for AEC1 on day-12 post bleomycin administration (Fig. 3e,f; Extended Data 4c,d). In uninjured lungs, we observed tdt expression in some LGALS3⁺ macrophages and rarely in AEC1s (Fig. 3e–g). We did not find *Krt19*-tdt⁺ cells co-expressing SFTPC in uninjured regions of bleomycin treated mice, indicating that *Krt19* expression is specifically activated in response to injury in the damaged regions and the conditions we tested are well suited for labeling and tracing PATS. Co-immunostaining revealed stretched tdTomato-labeled cells co-expressing PATS markers including SFN, CLDN4, LGALS3 in bleomycin treated lungs but not in controls (Fig. 3e–g; Extended Data 4a,b). 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 cell shape and structure, from cuboidal to flat and thin. Such a transition is typically accompanied by many changes in the expression of signalling and structural proteins. We therefore hypothesized that these changes occur in PATS. To test our hypothesis, we analyzed scRNA-seq data from organoid cultures and LPS-induced injury models. We found numerous genes conserved among PATS (Fig. 4a). Pathway enrichment analysis revealed that TP53, TNF/NF-κB, ErbB, HIF1, Hippo/YAP, cell cycle arrest, cytoskeletal dynamics, tight junction and TGFβ signalling are activated in PATS compared to other populations (Fig. 4b–d, Extended Data 5). Unexpectedly, we also found significant enrichment for genes representative of senescence and DNA damage response pathways (Fig. 4c–f, Extended Data 5).

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

The above data indicated that DNA damage and senescence associated genes are highly enriched in PATS. To validate these findings, we assayed lung from bleomycin or PBS treated mice for *β-galactosidase* activity, that serves as a biomarker for senescent cells and γH2AX, a marker for DNA damage. Interestingly, we detected numerous *β-galactosidase* active cells, based on X-gal derived blue color deposition, in bleomycin injured but not control alveoli (Fig. 5a,b). This is further supported by expression of senescence marker,

CDKN1A (p21) specifically in bleomycin treated lungs (Fig. 5c). We observed CDKN1A expression in stretched tdTomato-labeled cells but not in AGER$^+$ cells, indicating that senescence is not observed in AEC1 (Fig. 5c). Similarly, we also detected accumulation of γH2AX puncta in the nuclei of stretched *Sftpc*-tdt$^+$ cells but not in AGER$^+$ cells in bleomycin treated lungs (Fig. 5d). Further analysis revealed a significant number of γH2AX $^+$*Sftpc*-tdt$^+$ cells co-expressing LGALS3 in bleomycin injured lungs but not in controls, indicating that PATS undergo DNA damage during alveolar regeneration (Fig. 5e,f).

Bleomycin is known to induce DNA damage in cells. However, both scRNA-seq analysis and immunostaining revealed a strong enrichment of DNA damage-repair signalling in PATS (Fig. 4 and Fig. 5e,f). To avoid any potential effects from bleomycin on senescence and DNA damage, we performed marker analysis on lungs from the AEC1-specific ablation model (Fig. 5g). We detected numerous X-gal$^+$ cells and found expression of CDKN1A or γH2AX in LGALS3$^+$ cells in AEC1 ablated lungs but not in control alveoli (Fig. 5h–k).

PATS cells are vulnerable to mechanical stretch-induced DNA damage

Previous studies have shown that ionizing radiation, oxidative stress, and the mechanical forces that occur during cell migration and stretching can all cause DNA damage[19]. Pathway analysis revealed an enrichment for genes involved in cytoskeletal changes but not oxidative stress. Since AEC2 differentiation into AEC1 requires extensive cell spreading and cytoskeletal dynamics we hypothesized that PATS cells experience mechanical stretching that can lead to DNA damage. To test this, we purified AEC2s from *SFTPC-GFP* mice and cultured them on a plastic surface (2D culture), conditions under which AEC2s spread and differentiate into AEC1s[20]. Within 5 days after plating, most AEC2s stretched and either lost (GFP$^-$) or downregulated (GFP^{lo}) GFP expression. These cells increased their surface area through stretching and began to express AEC1 marker. Interestingly, AGER$^+$ cells are also positive for the DNA damage marker, γH2AX (Fig. 5l,m). Of note, we did not see DNA damage markers in cells on day-2 after plating (Fig. 5m). Taken together, our data show that cuboidal AEC2s that differentiate into large and thin AEC1s naturally experience DNA damage-repair and undergo transient senescence.

Genetic loss of function and pharmacological activation of TP53 signalling dysregulates AEC2 to AEC1 differentiation via PATS during alveolar injury-repair

Our data suggest that activation of TP53 signalling is associated with PATS in injured lungs. Previous studies in other tissues have suggested that TP53 signalling promotes differentiation and suppresses stem cell self-renewal[21]. This led us to hypothesize that enhanced activation of TP53 signalling increases AEC2 differentiation into AEC1 involving PATS after injury. To test this, we performed bleomycin-induced injury in *Sftpc-CreER;R26-tdTomato* mice followed by Nutlin-3a (TP53 activator) or DMSO (control) administration starting on day-8 after injury and tissues collection on day-20 (Fig. 6a). Immunostaining for AGER revealed that Nultin-3a treatment led to significantly greater differentiation of *Sftpc*-tdt$^+$ cells into AEC1 compared to DMSO (Fig. 6b,c). We did not observe any AGER$^+$*Sftpc*-tdt$^+$ cells in uninjured lungs that received Nutlin-3a, suggesting that Nutlin3a-induced TP53 activation alone is not sufficient to induce AEC2 differentiation into AEC1 (Fig. 6b,c). Additionally, we tested the effect of Nutlin-3a on AEC2s in alveolar organoids. Organoids

were treated with Nutlin-3a or DMSO starting on day-7 until harvest on day-15 (Extended Data 6a,b). Nutlin-3a treated organoids showed an increase in the number and intensity of AGER+ cells and decrease in the number of Ki67+ proliferating cells compared to controls (Extended Data 6a,b). These data further support our hypothesis that ectopic activation of TP53 signalling enhances AEC2 differentiation *in vivo* and *ex vivo*.

The above data indicated that activation of TP53 signalling enhances AEC2 differentiation. To test whether TP53 is necessary for differentiation of AEC2 into AEC1, we utilized the *Sftpc-CreER;R26-tdTomato;Trp53fl/fl* mouse model where tamoxifen administration allows deletion of TP53 and expression of tdTomato specifically in AEC2s. We then performed bleomycin injury followed by tissue collection at day-12 (Fig. 6d). *Sftpc-CreER;R26-tdTomato;Trp53+/+* mice were used as controls. Immunostaining revealed numerous stretched lineage-labeled cells co-expressing AGER in bleomycin treated control mice (Fig. 6e,f). In contrast, in TP53 deficient lungs tdt+ cells showed no AGER expression after bleomycin treatment and had abnormal morphology (Fig. 6e,f). Of note, in regions that did not have bleomycin-induced injury the TP53 deficient AEC2s appeared normal (Extended Data 6c,d). Immunostaining and quantification revealed a significant increase in the number of CLDN4+Sftpc-tdt+ and fewer LGALS3+Sftpc-tdt+ in TP53 mutant lungs compared to controls following injury (Fig. 6g).

Expression analysis revealed a significant decrease of CDKN1A (a known target of TP53 in TP53fl/fl lungs compared to controls (Fig. 6h, i). Immunostaining for cell death marker revealed no difference between TP53 mutant and controls, indicating that loss of TP53 did not affect cell death (Extended Data 6e). Taken together, our data revealed that TP53 is essential for differentiation of AEC2s into AEC1 and TP53 deficient AEC2s are stuck in PATS and unable to progress to AEC1 (Fig. 6a–i; Extended Data 6).

TP53 binds to PATS associated gene loci and controls their expression

To identify the transcription factor modules that are potentially active in PATS, we performed chromatin immunoprecipitation analysis for histones: H3K4me3, H3K36me3 and H3K27ac, to identify active promoters, transcribing genes, and enhancers, respectively. PATS and AEC2s were isolated from bleomycin treated *Ctgf-GFP* or control mice, respectively, and used them for histone marks analysis (Fig. 6f). As expected, we found H3K4me3 enrichment in genes corresponding to cell type specific promoters (*Sftpc* in AEC2s and *Fn1* in PATS) (Extended Data 7a–c). Further analysis revealed enrichment of numerous H3K4me3 peaks overlapping with transcriptional start sites (TSS) and promoters of transcripts specific to PATS from bleomycin-induced lungs compared to AEC2s from controls (Extended Data 7d). Motif analysis in H3K4me3 peaks predicted enrichment of binding sites for transcription factors (TFs) including TP53, ETS1, NF1, ATF3, and SOX4, all of which have been implicated in PATS enriched pathways (Extended Data 7e)[22–25]. For example, we found predicted binding sites for TP53 in the *Mdm2* enhancer and *Fas* promoter, two well-known direct targets of TP53 (Extended Data 7f,g)[26]. Taken together, our data implicate a direct role for TP53 in transcriptional control of PATS-specific genes.

To empirically test whether TP53 directly regulates transcriptional programs associated with PATS, we performed chromatin immunoprecipitation analysis for TP53 in PATS. Since our

scRNA-seq analysis revealed that TP53 signalling is highly enriched in the *Ctgf*⁺ subset of PATS, we generated *Sftpc-creER;R26R-tdTomato;Ctgf-GFP* mouse model to purify AEC2-lineage derived PATS for ChIP analysis. We purified PATS (tdt⁺GFP⁺) from *Sftpc-creER;R26R-tdTomato;Ctgf-GFP* lungs on day-8 post bleomycin administration and performed Mint-ChIP analysis for TP53 (Fig. 6j; Extended Data 7h). Data visualization on IGV (integrative genome viewer) tracks revealed strong enrichment of TP53 on previously described targets including *Mdm2, Trp53, and Basp1* (Extended Data 7i). More importantly, we observed strong enrichment of TP53 on many (22 of the 28) PATS target genes that we analyzed, including, *Fn1, Sfn, Actn1, Nupr1, Myl12a, Ctgf, Cdkn1a, and Krt19)* (Fig. 6k, Extended Data 7j). Of note, we did not find TP53 enrichment on AEC1 or AEC2 gene loci (Extended Data 7k,l). Additionally, we also found overlap between TP53 enrichment regions and H3K4me3 or H3K27ac marks, consistent with TP53 binding on promoters and enhancers of PATS associated genes (Fig. 6k). Taken together, our data revealed that TP53 directly binds on PATS enriched gene promoters and enhancers.

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

Recent studies suggested that alveolar epithelial cells that line fibrotic foci in idiopathic pulmonary fibrosis (IPF) show features of senescence, growth arrest, and differentiation blockade[27,28]. Thus, we hypothesized that in response to a non-permissive pathologic microenvironment, alveolar progenitors stall their differentiation process in a PATS-like state. To test this, we analyzed a recently reported scRNA-seq dataset from human IPF lungs[29]. First, we excluded all non-epithelial 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⁺[29]. 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 murine organoids and injury model (Fig. 1, Fig. 4, Fig. 7a,b, Extended Data 8a). We therefore named human IPF-specific cluster as 'PATS-like' cells (Fig. 7a–c). Other transcripts highly enriched in the PATS-like population are *CALD1, PRSS2, MMP7,* and *S100A2* (Extended Data 8a,b). To validate scRNA-seq data, we performed immunofluorescence analysis on healthy human and fibrotic and non-fibrotic regions of IPF lungs (Extended Data 8c). Co-immunofluorescence analysis for PATS-like (SFN, CLDN4, KRT17), AEC2 (SFTPC, HTII-280), AEC1 (AGER) and myofibroblasts (ACTA2) revealed the expression of PATS-like markers specifically in fibrotic regions of IPF lungs but not in healthy controls and healthy appearing regions in IPF lungs (Fig. 7d–i; Extended Data 8d–g). Interestingly, KRT17 has been shown recently to be expressed in basal cells of the normal lung and basaloid-like cells in IPF lungs [29,30].We also found a similar expression pattern of TP63, another marker for basal cells specifically in fibrotic but not in non-fibrotic regions of the IPF lungs (Fig. 7h,i; Extended Data 8e,f). We find that PATS-like markers are highly enriched in severe-fibrotic regions (Fig. 7d–i) coinciding with high levels of COL1A1 and accumulation of myofibroblasts (Fig. 7d; Extended Data 8h). Quantification of PATS markers (SFN, CLDN4, KRT17 and TP63) further supported these findings (Fig. 7e,g,i).

Moreover, we observed that PATS acquired an elongated morphology in IPF lungs as opposed to cuboidal AEC2s in healthy lungs (Fig. 7d,f,h and Extended Data 8d,g). Of note, AGER expression is absent in regions with many PATS-like cells in IPF lungs (Fig. 7f; Extended Data 8g). Gene ontology and pathway analysis revealed an enrichment for components of p53 signalling (*CDKN1A, CDKN2A, MDM2*), DNA-damage checkpoint (*RPS27L, PLK3*) and cellular senescence (*TGFB2, HIPK2*) (Fig. 7j; Extended Data 9a–e). 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 (Fig. 4; Extended Data 9a–e). We also found a significant enrichment for *AREG, TGFB1, TGFB2* and *TIMP1*, known regulators of fibrosis, in PATS-like cells (Fig. 7k). 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. 7l–o; Extended Data 9f–h). Taken together, our analysis revealed similarities between PATS from regenerating tissues and PATS-like cells that are specific to pathological fibrotic lungs.

Discussion

This study uncovered a previously unknown transitional state (PATS) that traverses between AEC2 and AEC1 and has a specific gene expression signature. Fate mapping studies demonstrate a clear lineage relationship between AEC2 to PATS to AEC1. Therefore, with the addition of these transitional cell states, our work revises the alveolar epithelial cell hierarchy. Specific gene expression signatures indicate that PATS are not merely undergoing a gradual loss of AEC2 characteristics but represent a unique transitional population (Extended Data 10). These cells are enriched for pathways, including, TP53, NF-κB, YAP, TGFβ, and HIF1, previously shown to be important for lung regeneration[12,14,31,32]. We observed enrichment for transcripts associated with cell cycle arrest, senescence, and SOX4, a known regulator of epithelial-mesenchymal transition and cell adhesion in other tissues[33]. Through genetic and pharmacological modulation, our data demonstrate that TP53 signalling is necessary and sufficient to promote AEC2s differentiation into AEC1 involving PATS in regenerating tissues. ChIP analysis revealing a direct control of PATS genes by TP53. Although binding of TP53 to numerous cytoskeletal genes had previously been postulated[34,35] our study is the first, to our knowledge, to provide direct evidence for this activity. Moreover, we also observed TP53 binding on genes that are implicated in DNA damage pathways (*Nupr1, Sox4*).

Our analysis of human lungs has identified PATS-like cells that are specifically present in the fibrotic regions of IPF lungs. Similar to murine, PATS-like cells in human lungs are characterized by enrichment for genes associated with cellular senescence, TP53 signalling, and TGFβ regulated genes, all known to be involved in fibrosis in multiple organs, including the lung[36–38]. In contrast, we also found some differences in gene expression signatures between murine and human IPF-specific PATS-like cells. Some of them include, TP63, KRT17, and COL1A1, which are found only in human PATS-like cells. Our RNA velocity projections from human scRNA-seq data suggest that these KRT17+/TP63+ cells originate from AEC2. Indeed, immunofluorescence for KRT17 and TP63 further suggested that these

cells are surrounded by PATS-like cells within the same alveoli. Alternatively, it is possible that these cells may arise from airways as previously suggested in influenza-induced mouse injury models that develop so-called basal cell "pods"[39]. However, the morphology of these "pod" cells appear to differ from the basal cells we observed in our study.

Our findings showed that PATS cells undergo extensive stretching during AEC2 differentiation into thin and large AEC1, which makes them vulnerable to DNA damage, a feature associated with most degenerative lung diseases, notably pulmonary fibrosis and cancers[40–42]. Previous studies revealed that cell stretching causes DNA damage when they migrate or squeeze through narrow spaces[43]. Indeed, our injury models and 2D culture suggest that AEC2 undergo extensive stretching during differentiation into AEC1. Therefore, the transitional cell state has clinical implications, notably lung diseases associated with DNA damage[41,44]. Interestingly, the PATS population is enriched for SOX4, known to regulate cytoskeletal genes, which is induced following DNA damage and is critical in TP53 stabilization and function[25]. Altogether, these data support a model in which cells evolved co-transcriptional programs to combat DNA damage that can occur when cells undergo stretching. Additionally, genome-wide studies have identified mutations in DNA damage-repair components, such as *XRCC* family genes, *LIG4*, *TERC*, PARP, and *RTEL1* with emphysema and pulmonary fibrosis[40]. Thus, the newly identified transitional state implicates cell shape changes and associated vulnerabilities accompanying alveolar stem cell differentiation in the lung pathogenesis.

Senescence is often seen as an age-associated pathological state in which cells acquire an abnormal and irreversible state[28,42,45,46]. We find that alveolar stem cell differentiation involves a transitional state which exhibits cardinal features of senescence in normal tissue regeneration. Indeed, prior studies have found senescent cells in developing limb bud tissues[47]. Therefore, we propose that senescence may not necessarily occur exclusively in aged tissues but can be a reversible transient state accompanying tissue 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 and *in vivo* injury-repair models, we have identified a pre-AEC1 transitional state in lung regeneration. This unique state is associated with cellular senescence and enrichment for defective alveolar regeneration pathways (Extended Data 10). 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

Both male and female mice between 8–16 weeks were used for experiments. All of the mice were C57BL/6 unless otherwise indicated. The following mice were used for experiments: *Sftpc^tm1(cre/ERT2)Blh* (*Sftpc-CreER*)[17], *Krt19^tm1(cre/ERT)Ggu/J* (*Krt19-CreER*)[48], *Rosa26R-CAG-lsl-tdTomato*[49] (crossed with *Sftpc-CreER*), *B6.Cg-Gt(ROSA)26Sor^tm14(CAG-tdTomato)Hze/J* (*R26R-tdTomato*)[50] (crossed with *Krt19-CreER*), *Tg(SFTPC-GFP)#Heat* (*Sftpc-GFP*)[51], *B6.Cg-Ager^tm2.1(cre/ERT2)Blh/J* (*Ager-CreER*)[52], *B6-*

Gt(ROSA)26Sor^tm1(HBEGF)Awai/J (*R26R-DTR*)[53], *Mki67^tm1.1Cle/J* (*Mki67RFP*)[54] and *Tg(Ctgf-EGFP)FX156Gsat* (*Ctgf-GFP*)[15] and TP53^fl/fl (mixed background)[55]. For lineage tracing with *Sftpc-CreER;R26R-tdTomato*, 3–5 doses of 2mg/20gm body weight Tamoxifen (Tmx, Sigma-Aldrich) was given via oral gavage or intraperitoneal injection. For lineage tracing using *Krt19-CreER;R26R-tdTomato*, one dose of 1mg/20gm body weight tamoxifen was given via intraperitoneal injection seven days after bleomycin injury or PBS administration. Animal experiments were approved by the Duke University Institutional Animal Care and Use Committee in accordance with US National Institutes of Health guidelines.

Bleomycin injury

For bleomycin-induced lung injury, 2.5U/Kg bleomycin was administered intranasally at two weeks after tamoxifen injection and mice were monitored daily. PBS administered mice served as controls. Mice were sacrificed at different times after bleomycin injury.

Diphtheria toxin (DT) administration

Two weeks before DT administration, *Ager-CreER;R26R-DTR* mice received tamoxifen via IP injection. One dose of 3μg diphtheria toxin (Millipore #322326) was administer via intraperitoneal injection and mice were sacrificed six days later for tissue collection and analysis.

Mouse lung dissociation and fluorescence assisted cell sorting

Lung dissociation and FACS were performed as described previously[11]. Briefly, lungs were intratracheally inflated with 1ml of enzyme solution (Dispase (5U/ml, Corning #354235), DNase I (0.33U/ml) and collagenase type I (450U/ml, Gibco #17100–017)) in DMEM/F12. Separated lung lobes were diced and incubated with 3ml enzyme solution for 25min at 37°C with rotation. The reaction was quenched with an equal amount of 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 (155mM NH_4Cl, 12mM $NaHCO_3$, 0.1mM EDTA) and incubated for 2min, then filtered through a 40μm 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[56]. 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[4]. Briefly, lineage labeled AEC2s ($1–3 \times 10^3$) from *SFTPC-GFP* or *Sftpc-CreER;R26R-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* mice 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 a concentration of 20mg/Kg/day for ten consecutive days and samples were collected 20 days after bleomycin administration. For *ex vivo* studies, alveolar organoids were grown for seven days followed by Nutlin-3a (2μM) 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 20min followed by incubation with 0.25% Trypsin-EDTA at 37°C for 10min. 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 3min, 15–17 cycles of 98°C for 15sec, 65°C for 30sec, 68°C for 4min 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 [57]. 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 [58]. UMI count matrix of murine lungs treated with LPS (GSE130148) [14] 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 [59] 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)[29]. 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 signalling pathways and gene ontology through Enrichr [60]. Z-scores were calculated based on combined score in Kyoto encyclopedia of genes and genomes (KEGG) to compare enrichment of signalling 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 signalling pathways. The gene member lists of utilized pathways were obtained from KEGG pathways [61] and AmiGO [62]. 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.

Mint-ChIP (Multiplexed indexed T7 ChIP-seq)

For histone marks ChIP analysis, PATS (CD31⁻/CD45⁻/CD140a⁻/CD326⁺/*CTGF*-GFP⁺ cells) were sorted from *Ctgf-GFP* mice on day-12 post Bleomycin-induced lung injury. AEC2s (CD31⁻/CD45⁻/CD326⁺/Lysotracker⁺/*Mki67*-RFP⁻ cells) were sorted from *Mki67-RFP* homeostatic mice. *Mki67* mice were used to eliminate any cells that are undergoing cell division by negatively gating out RFP⁺ cells. Mint-ChIP protocol was described previously[63]. For TP53 ChIP-seq, PATS (CD31⁻/CD45⁻/*Sftpc*-tdTomato⁺/*CTGF*-GFP⁺ cells) were sorted from *Sftpc-CreER;R26R-tdTomato;Ctgf-GFP* mice on day-8 post bleomycin administration. In this case, we used a modified Mint-ChIP (version-3) protocol as described previously (https://tinyurl.com/udqksct). Following cell lysis, chromatin was digested with 300 units of MNase (#M0247S, New England Biolabs) at 37°C for 10 min. T7 adapter ligation was performed for 2 hrs and then the samples were split to give ~7,000 cells per antibody for histone modification ChIP-seq. For TP53, we used ~40,000 cells for each replicate sample. The samples were incubated with Histone H3 (H3) antibody (1 ul, #39763, Active Motif), Histone H3 lysine 36 trimethylation (H3K36me3) antibody (1 ul, #61101, Active Motif), Histone H3 lysine 4 trimethylation (H3K4me3) antibody (1 ul, #ab8580, Abcam), Histone H3 lysine 27 acetylation (H3K27ac) antibody (1 ul, #39133, Active Motif) or TP53 antibody (5 ul, #2524T, Cell Signaling Technology) at 4°C overnight.

DNA was purified followed by T7-RNA polymerase mediated *in vitro* transcription at 37°C for 3 hrs. Reverse transcription was performed as described in original protocol followed by library preparation using Terra Direct PCR polymerase (#639271, TaKaRa). Two experimental replicates were performed for each cell population. Libraries were sequenced (at least 5M reads of 150-bp paired end per sample) using Hiseq X or NovaSeq 6000.

Computational analyses for Mint-ChIP

FASTQ files were generated using Bcl2fastq. Additional demultiplexing for Mint-ChIP FASTQ files were performed using Je [64]. Low quality reads were trimmed out from FASTQ files using trimmomatic v0.38 [65]. Reads were mapped on mm10 genome reference using BWA [66]. The packages were run through a pipeline called MintChIP (https://github.com/jianhong/MintChIP). HOMER[67] was used for generating bedGraph files to visualize peaks in Integrative Genomics Viewer (IGV)[68]. Peak calling for H3K4me3 was performed using HOMER's *getDifferentialPeaksReplicates.pl -region -size 1000 -minDist 2000 -C 0 -L 50* with normalization by H3. Motif analysis was performed using HOMER's *findMotifsGenome.pl.* deepTools[69] was used for generating a chart of called peaks of H3K4me3. Called peaks for each genomic loci from different cell populations were prepared in Affinity Designer.

Human lung tissue

Excised subtransplant-quality human lung tissues from donors without preexisting chronic lung diseases and explanted fibrotic human lungs were procured through the BioRepository and Precision Pathology Center at Duke University in accordance with institutional procedures (Duke University Pro00082379 – "Human Lung Stem Cells"; exempt research as described in 45 CFR 46.102(f), 21 CFR 56.102(e) and 21 CFR 812.3(p) which satisfies the Privacy Rule as described in 45CFR164.514). IPF tissue samples used for scRNA-seq were obtained from lungs removed at the time of lung transplantation at two lung transplant centers (VUMC and NTI). Nonfibrotic control tissue samples were obtained from lungs declined for organ donation. For PF lungs, diagnoses were determined according to ATS/ERS consensus criteria (1). All studies were approved by the local Institutional Review Boards (Vanderbilt IRB #'s 060165, 171657, Western IRB # 20181836). The diagnosis of idiopathic pulmonary fibrosis (IPF) was evaluated by a surgical pathology team. Specimens were washed thoroughly in PBS prior to inflation and immersion in 4% PFA overnight at 4°C. Specimens were subsequently washed in PBS until the appearance of blood was minimal followed by incubation in 30% sucrose at 4°C. Samples were then incubated with 1:1 mixture of OCT for 1hour at 4°C before embedding in OCT. 7–9μm 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 30min then embedded in OCT or Paraffin. 10μm sectioned samples were utilized for staining following incubation at 95°C for 10–15min 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; Proteintech, 16165–1-AR, 1:500), GFP (Novus Biologicals, NB100–1770, 1:500), LGALS3 (Cedarlane, CL8942AP, 1:500); SOX4 (Invitrogen, MA5–31424, 1:250), SFN (Invitrogen, PA5–95056, 1:250, Proteintech, 66251– 1-Ig, 1:500, Abcam, ab77187, 1:200), ACTA2 (Sigma, C6198, 1:500), gamma-H2AX (R&D, 4418-APC, 1:500 and Novus Biologicals, NB100–74435, 1:250), CDKN1A (p21) (Sigma, ZRB1141, 1:200, BD Bioscience, 556430, 1:200), COL1A1 (Proteintech, 67288-Ig, 1:1000), YAP (Cell Signaling Technology, 4912S, 1:250).

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

PFA-fixed frozen sections were incubated with X-gal staining buffer containing 1mg/ml of X-gal (Thermo, R0941), 5mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6$, 2mM $MgCl_2$, 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, 10μm 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.

Proximity ligation in situ hybridization (PLISH)

PLISH was performed as described previously[70]. Briefly, fixed-frozen mouse lung sections were fixed with 4.0% paraformaldehyde for 20 min, treated with 20 μg/mL proteinase K for 9 min at 37°C, and dehydrated with up-series of ethanol. The sections were incubated with gene-specific oligos (Supplementary Table 2) in hybridization buffer (1 M sodium trichloroacetate, 50 mM Tris [pH 7.4], 5 mM EDTA, 0.2 mg/mL heparin) for 2 h at 37°C. Common bridge and circle probes were added to the section and incubated for 1 h followed by T4 DNA ligase reaction for 2 h. Rolling circle amplification was performed by using phi29 polymerase (#30221, Lucigen) for 12 hours at 37°C. Fluorophore-conjugated detection probe was applied and incubated for 30 min at 37°C followed by mounting in medium containing DAPI.

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 over the single channel immunofluorescence stain of interest[11]. Images were processed using Olympus CellSens application or ImageJ and figures were prepared using Affinity Designer.

Statistics and Reproducibility

Experiments were performed on at least three biological replicates (except organoid scRNA-seq and ChIP-seq). 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 and un-paired student's t-test were 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.

Data availability

All NGS sequencing data in this manuscript are available at NCBI GEO under accession numbers GSE141634 (organoid scRNA-seq); GSE141635 (ChIP-seq); and GSE135893 (scRNA-seq data from normal and IPF human lungs). Previously published sequencing data that were re-analyzed here are available under accession code GSE130148. All other data supporting the findings from this study are available from the corresponding author on reasonable request.
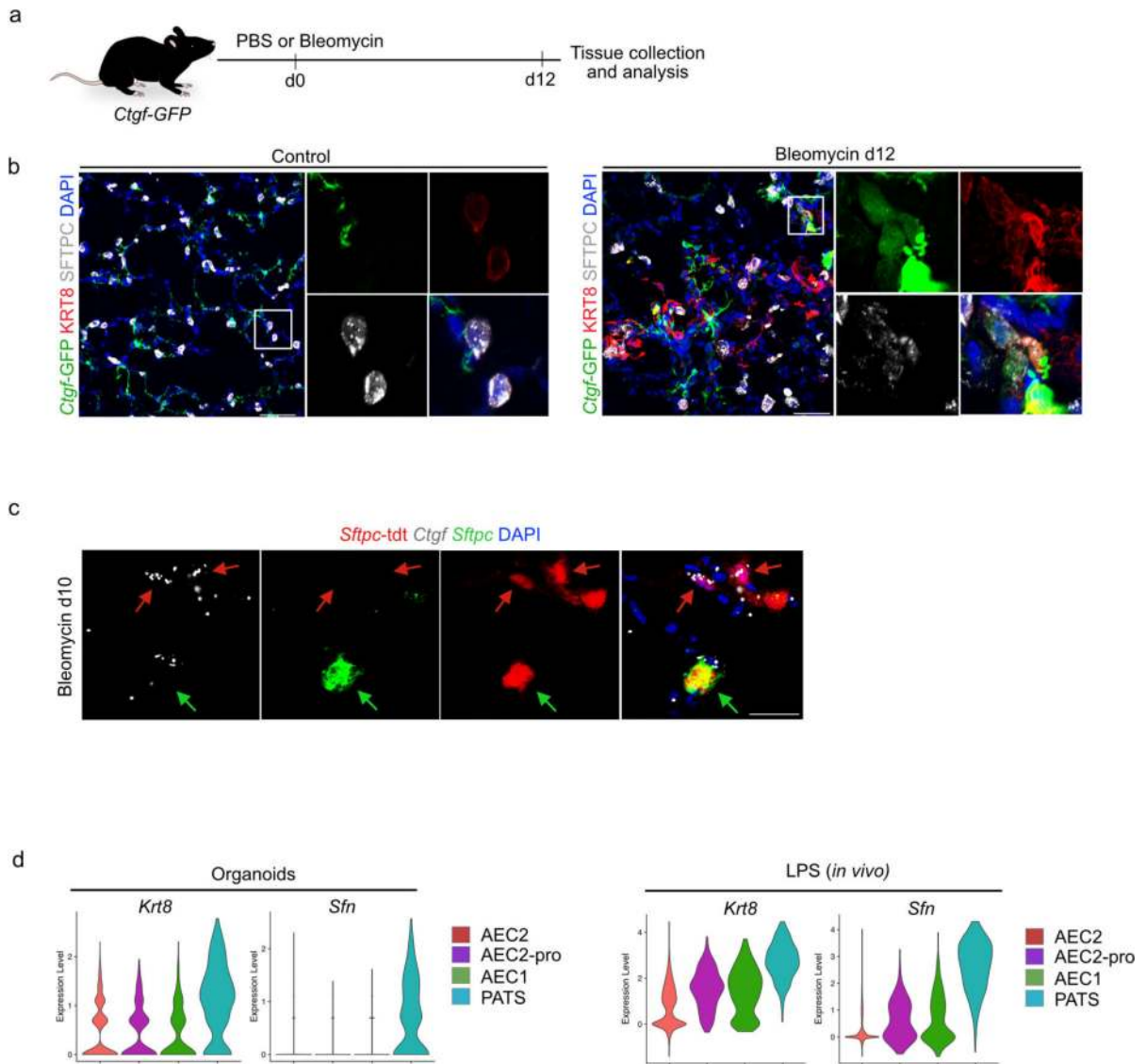
Extended Data
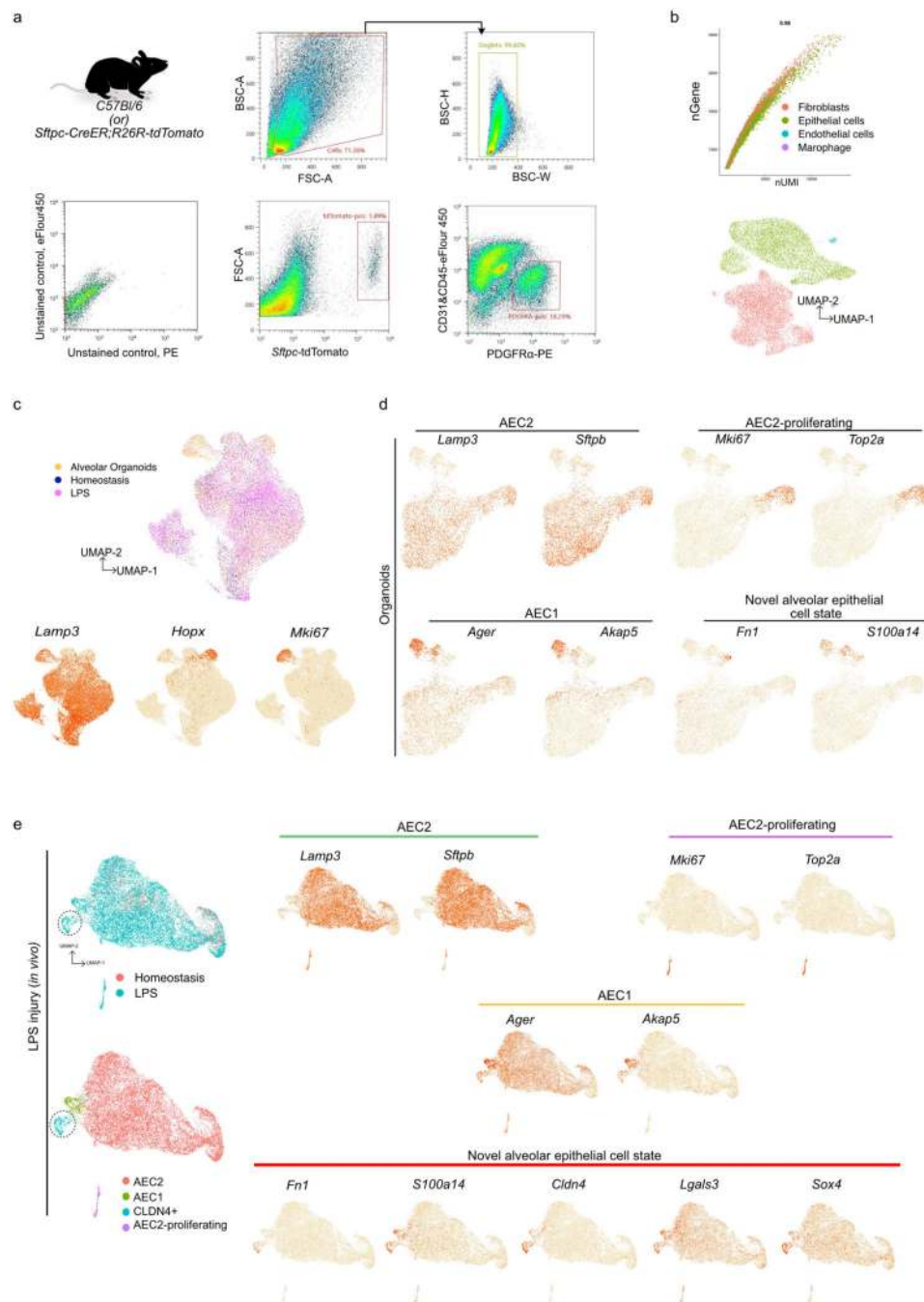
Extended Data 1. scRNA-seq identifies distinct alveolar cell populations in organoid cultures and LPS treated lungs *in vivo*.
a, Representative gating for FACS sorting of *Sftpc*-tdTomato⁺ AEC2s and PDGFRα⁺ fibroblasts utilized for organoid cultures. Singlet cells were used for further gating based on antibody staining (CD31, CD45 and PDGFRα) or tdTomato expression. Unstained control is shown in left bottom. b, Pearson correlation plot visualizes the number of genes per cell (nGene) and unique molecular identifier (nUMI) in total cells derived from alveolar

organoids (left panel, 10,948 cells). UMAP shows major cell populations including epithelial cells (green, 5,163 cells), fibroblasts (red, 5,686 cells), and some minor populations such as endothelial cells (blue, 66 cells) and macrophages (purple, 33 cells) in alveolar organoids (right panel). c, Integrated UMAP showing cells derived from alveolar organoids (orange, 4,787 cells), *in vivo* homeostatic mouse lung (blue, 1,878 cells) and LPS-treated mouse lung (magenta, 14,323 cells) (left). Expression of indicated genes in the integrated UMAP (right). d and e, UMAP plots show the expression of indicated genes in AEC2, AEC2-proliferating, AEC1, and transitional alveolar epithelial cell state in alveolar organoid scRNA-seq dataset (d, 4,573 cells) and LPS treated or control lungs (13,204 cells). In panel e, UMAP shows homeostatic (red) and LPS-treated (blue) lung.

Extended Data 2. Expression pattern of transitional alveolar epithelial cell state enriched genes in organoids and LPS or bleomycin treated lungs *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

treated lungs on day-12 (right). n=3 mice. Magnified single channel images are shown on the right. White line box indicates magnified region. c, Proximity ligation *in situ* hybridization analysis for *Ctgf* (grey) and *Sftpc* (green) on sections derived from *Sftpc-tdTomato* mice administered with bleomycin. Red arrows indicate *Ctgf$^+$tdTomato$^+$Sftpc$^-$* cells. Green arrows indicate *Ctgf$^+$tdTomato$^+$Sftpc$^+$* cells. d, Violin plots show the expression of *Krt8* and *Sfn* in epithelial cell populations derived from alveolar organoid (left, 4,573 cells) and LPS -injured lung scRNA-seq datasets (right, 13,204 cells). Violin plots indicate distribution of the cells. Scale bars: 30 μm.
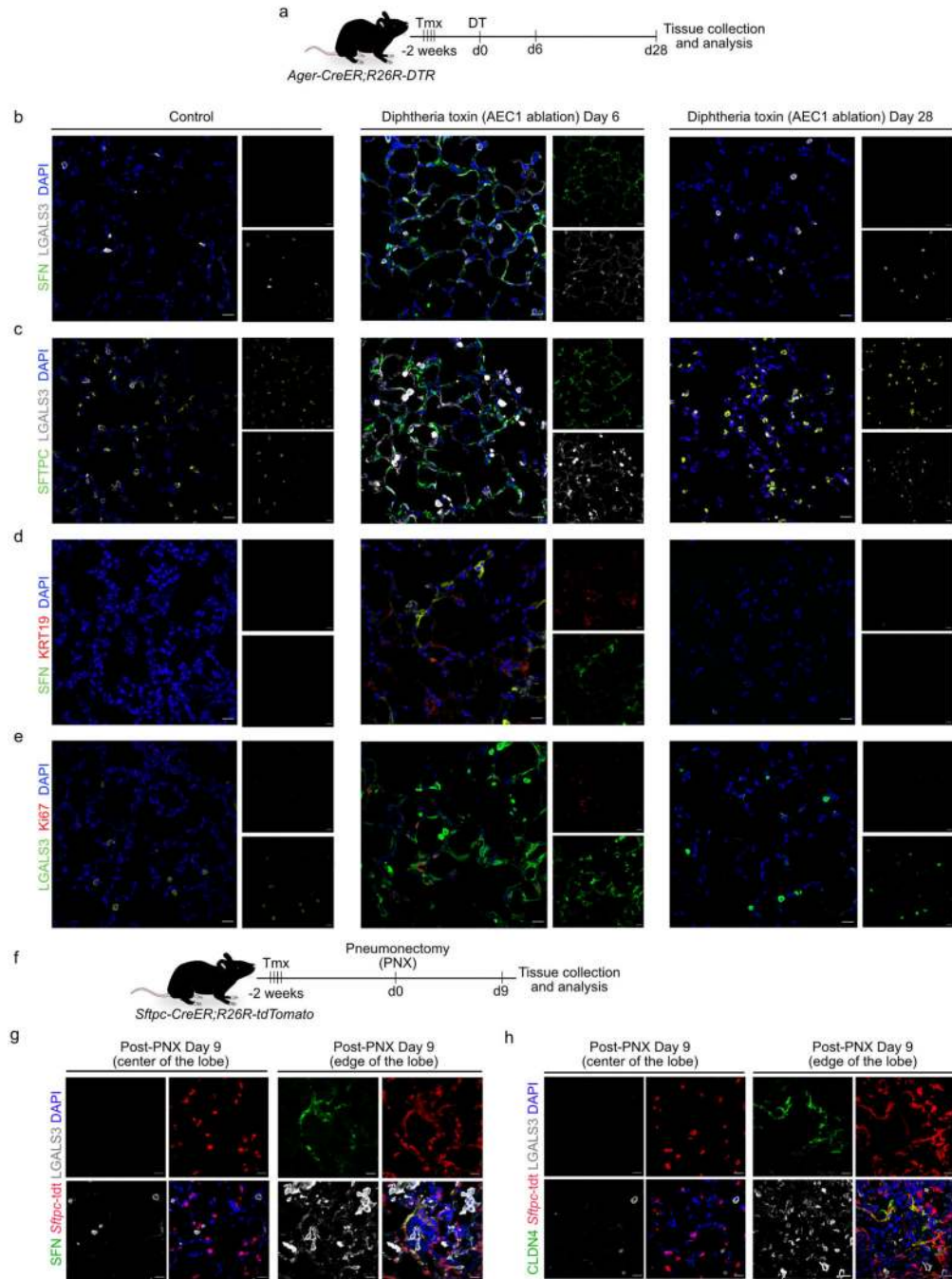
Extended Data 3. Expression of markers specific to the transitional alveolar population in AEC1-specific ablation mouse model and pneumonectomy-induced alveolar regeneration.

a, Experimental design to ablate AEC1 cells using *Ager-CreER;R26R-DTR* mouse model.

b-e, Immunostaining for b) SFN (green) and LGALS3 (grey) or c) SFTPC (green) and LGALS3 (grey) or d) SFN (green) and KRT19 (red) or e) LGALS3 (green) and Ki67 (red) in control (left panel) and AEC1-ablated lungs on day 6 (middle panel) and day 28 (right panel). f, Schematic of AEC2 lineage tracing using *Sftpc-CreER;R26R-tdTomato* mice follow by pneumonectomy (PNX) and tissue collection on day 9. g, Immunostaining for SFN (green), *Sftpc*-tdt (red) and LGALS3 (grey) in center (left panel) and edge (right panel)

of the lungs after pneumonectomy. h, immunostaining for CLDN4 (green), *Sftpc*-tdt (red) and LGALS3 (grey) (right panel) in center (left panel) and edge (right panel) of the lungs after pneumonectomy. DAPI stains nuclei (blue). Scale bars indicate 20 μm. Images in b-e, g and h are representative from three mice repeated independently with similar results.
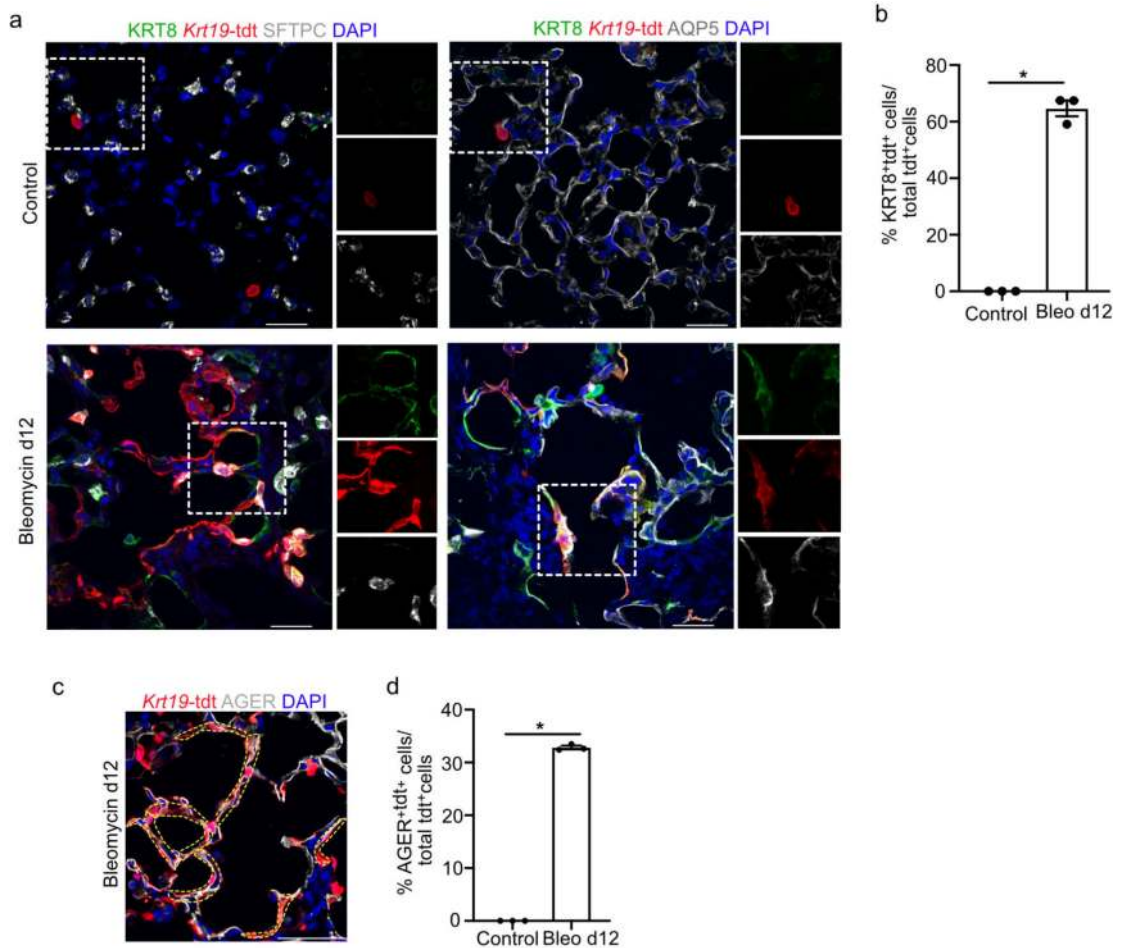
Extended Data 4. Expression pattern of PATS and AEC1 markers in Krt19-lineage traced lungs during alveolar regeneration.

a, Immunostaining for KRT8 (green), *Krt19*-tdt (red) and SFTPC (grey) (left panel) or AQP5 (grey) (right panel) in control (upper panel) and bleomycin-treated lungs (lower panel). Scale bars indicate 30 μm. n=3 mice. b. Quantification of $KRT8^+Krt19$-tdt^+ cells in total $Krt19$-tdt^+ cells. $p=0.0318$, (one-tailed, Mann-Whitney). n=3 mice. c, Dotted lines indicates the methodology used for quantification of $AGER^+$ cells. n=3 mice. d. Quantification of $AGER^+Krt19$-tdt^+ cells in total $Krt19$-tdt^+ cells $p=0.0318$ (one-tailed, Mann-Whitney). n=3 mice. Data are from three independent experiments and are presented as mean ± s.e.m. White boxed inset indicates individual color channels shown on the right. DAPI stains nuclei (blue).
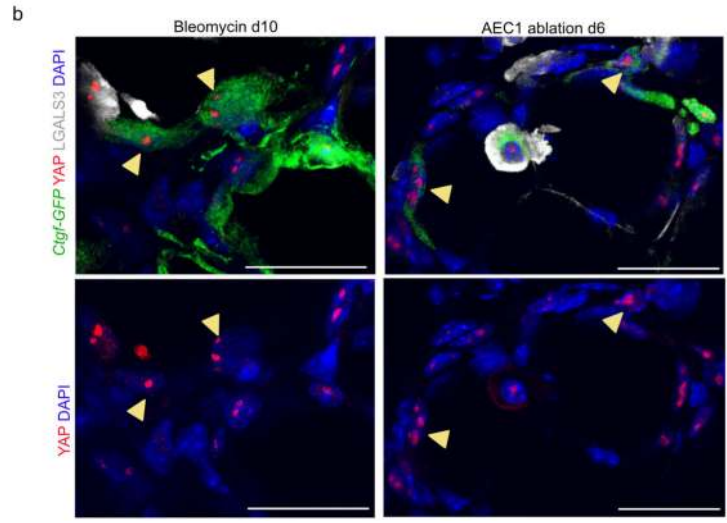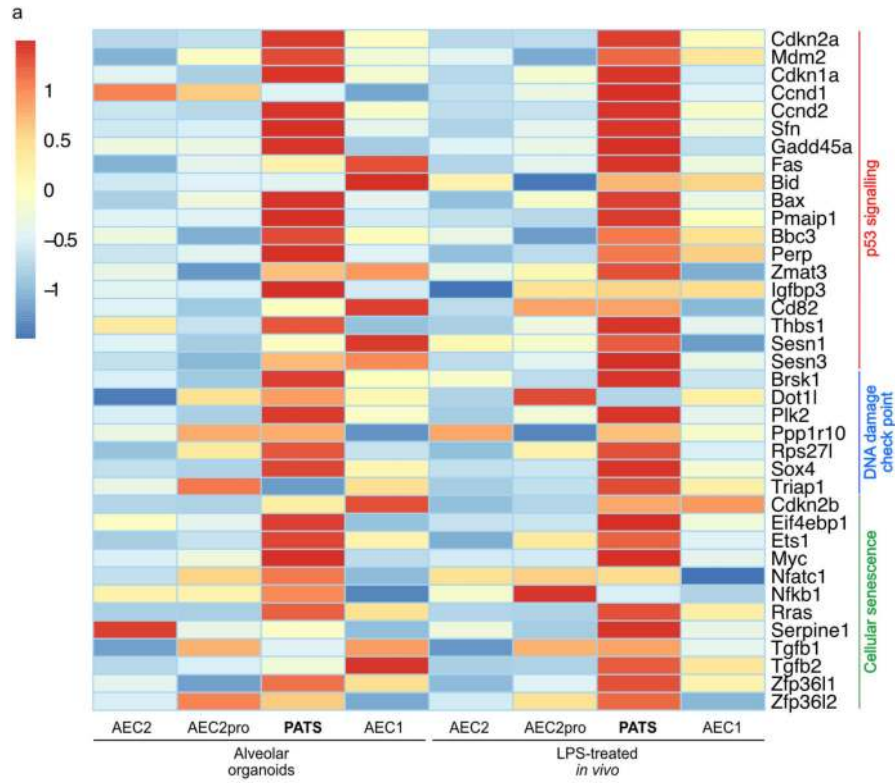
Extended Data 5. Signalling pathways enriched in murine PATS.

a, Heatmap shows expression of known target genes of indicated signalling pathways in AEC2, proliferating AEC2 (AEC2pro), PATS and AEC1 in alveolar organoids (4,573 cells) and LPS-treated murine lung (13,204 cells). Scale indicates z-score where red is high, and blue is low. b, Immunostaining for YAP (red), *Ctgf*-GFP (green) and LGALS3 (grey) in bleomycin-treated mouse lung (left) and AEC1-ablated lung (right). Arrowheads indicate YAP expression in *Ctgf*-GFP+ cells. DAPI stains nuclei (blue). Scale bars indicate 25μm. n=3 mice.
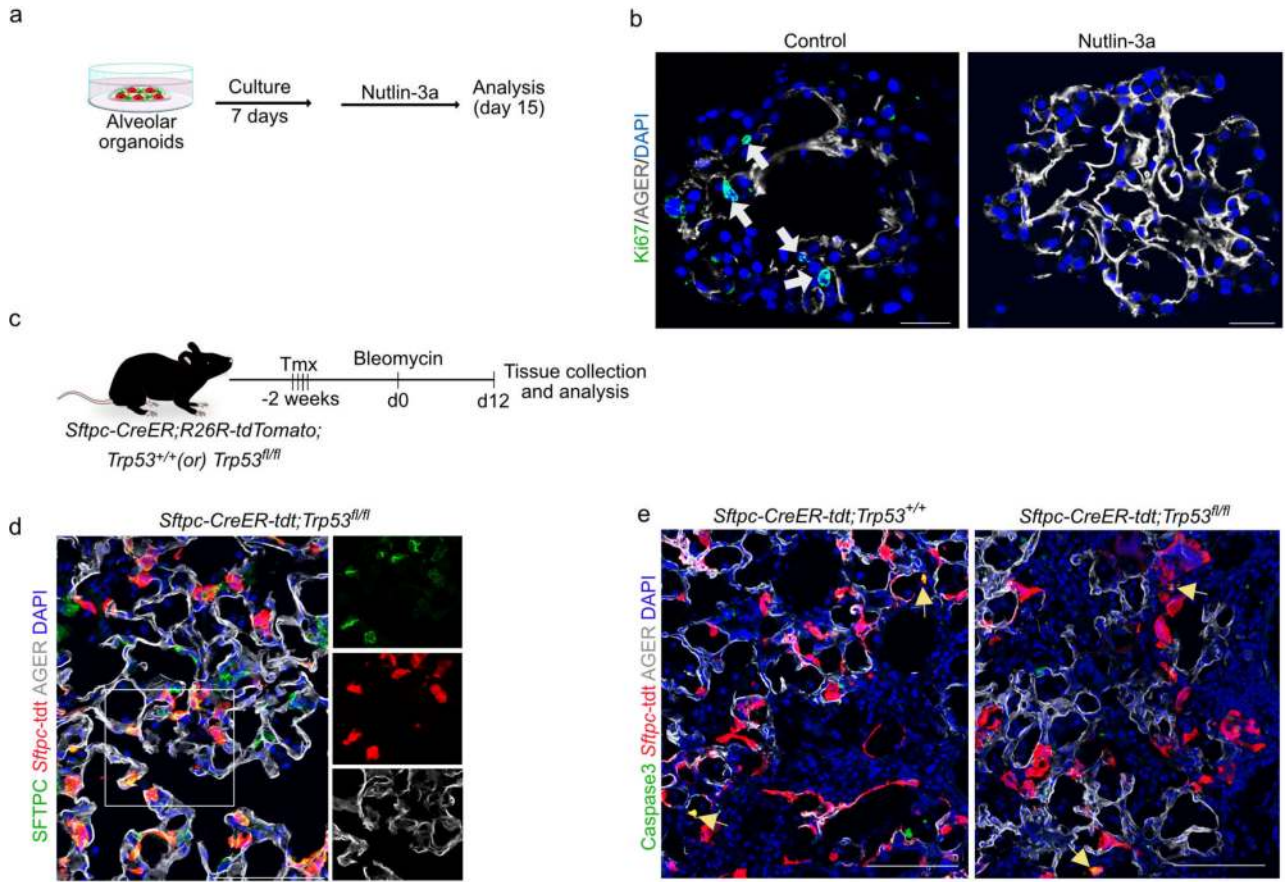
Extended Data 6. Genetic and pharmacological modulation of TP53 signalling in organoid cultures and during alveolar regeneration *in vivo*.

a, Schematic of alveolar organoid culture treated with Nutlin-3a. b, Immunostaining for Ki67 (green) and AGER (grey) in control or Nutlin-3a treated alveolar organoids. Scale bar: 30 μm. c, Schematic representation of experimental design to delete *Trp*53 in AEC2s followed by bleomycin-induced lung injury in Sftpc-tdt;Trp53$^{+/+}$ and Sftpc-tdt;Trp53$^{fl/fl}$. d, Immunostaining for SFTPC (green), *Sftpc*-tdt (red) and AGER (grey) in lungs that show normal appearing regions in bleomycin treated TP53 deleted (*Sftpc*-tdt;Trp53$^{fl/f}$) mice. e, Immunostaining for active Caspase 3 (green), *Sftpc*-tdt (red) and AGER (grey) in control (*Sftpc*-tdt;Trp53$^{+/+}$) and TP53 deleted (*Sftpc*-tdt;Trp53$^{fl/fl}$) mice. Scale bars: 100 μm. DAPI stains nuclei (blue). Images from b, d and e are representative from three mice repeated independently with similar results.
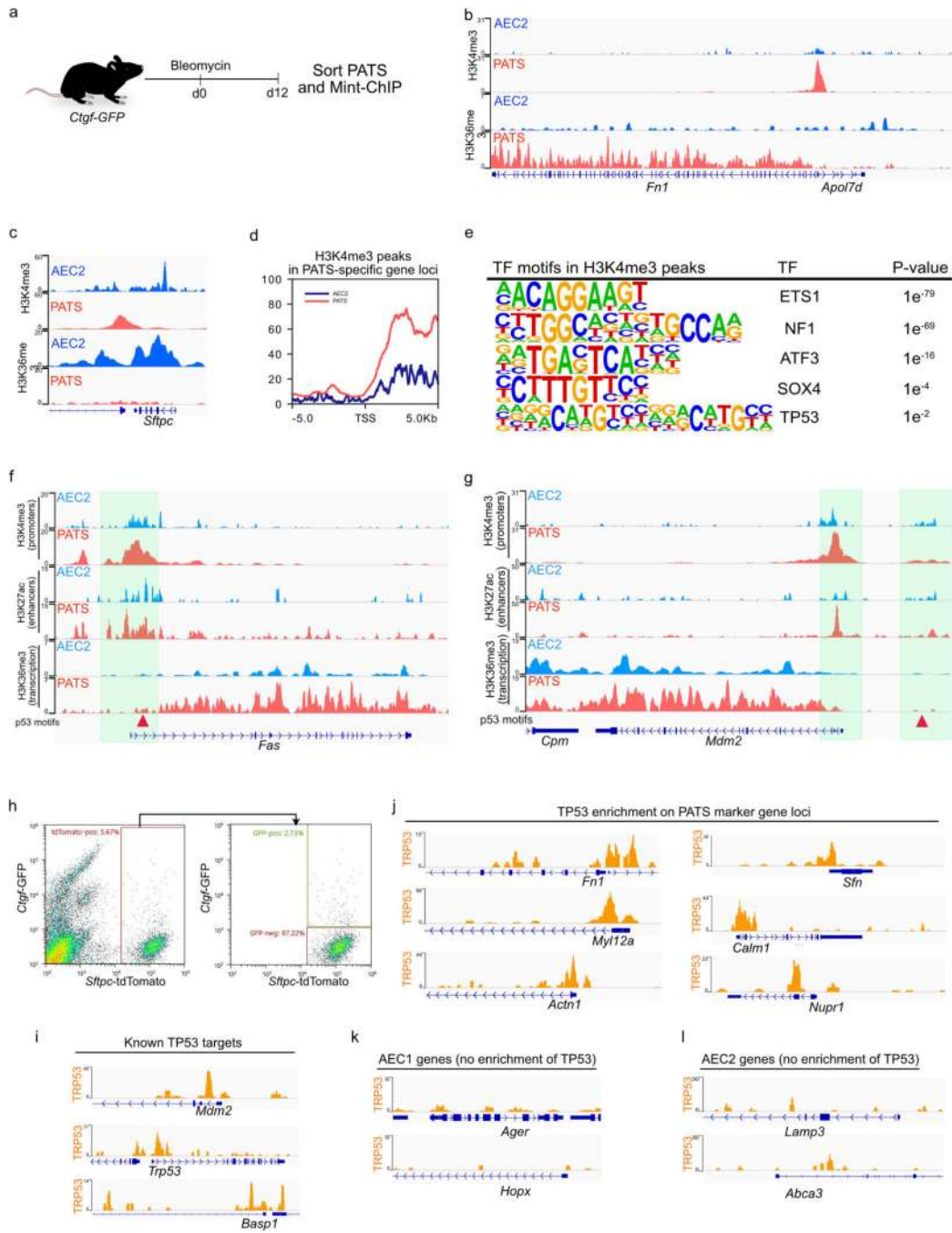
Extended Data 7. Transcriptional control of PATS by TP53 signalling.
a, Schematic of bleomycin-induced lung injury in *Ctgf-GFP* mice. b,c, IGV tracks show
presence or absence of H3K4me3 and H3K36me3 marks in *Fn1* (b) and *Sftpc* gene loci (c)
in AEC2 and PATS. d, Distribution of H3K4me3 peaks in PATS marker gene loci in PATS
(red line) and homeostatic AEC2s (blue line). e, Enriched TF motifs in H3K4me3 called
peaks in PATS specific gene loci (n=2, enrichment was detected using HOMER's
findMotifsGenome.pl). f,g, Enrichment for H3K4me3, H3K27ac and H3K36me3 in known
TP53 target gene loci (*Fas* and *Mdm2*) in PATS compared to AEC2s. Arrowhead indicates

location of predicted TP53 binding motifs. Green-shade regions are promoter or enhancer. h, Representative gating for FACS sorting of PATS utilized for ChIP-seq. *Ctgf*-GFP+/*Sftpc*-tdTomato[+] cells are sorted from Bleomycin-treated mouse lung as PATS. i-l, IGV tracks show significant enrichment for TP53 binding in genomic loci corresponding to known targets of TP53 (i), PATS enriched genes (j) but not on AEC1 (k) and AEC2 (l) gene loci.
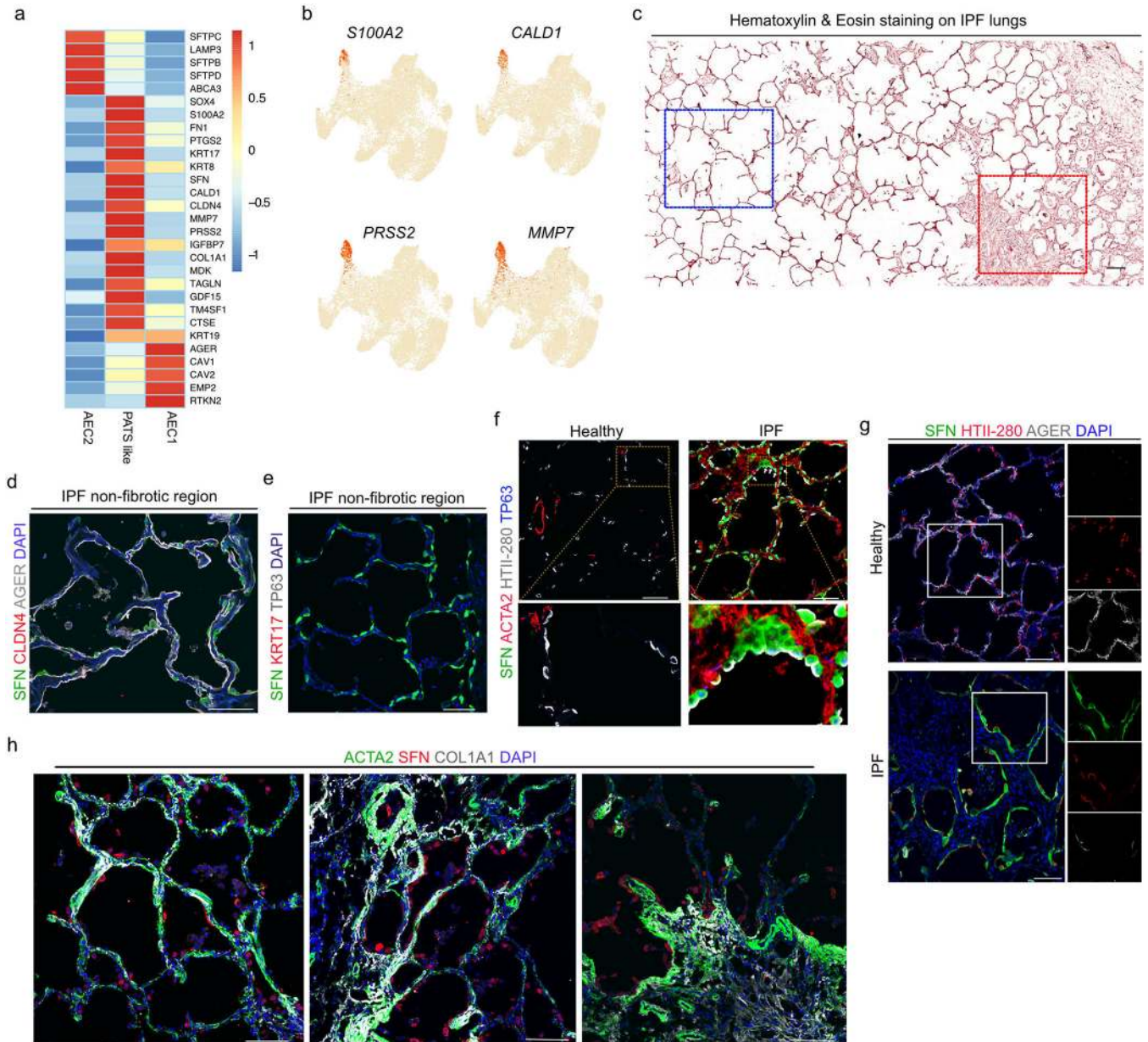
Extended Data 8. Markers specific to PATS-like cells are highly enriched in human fibrotic lungs compared to healthy lungs.

a, Heatmap shows expression of marker genes of each cell population in human lungs (scale shows z-score, 11,725 cells). b, UMAP plots show the expression of indicated genes in alveolar epithelial populations in heathy controls and fibrotic human lungs (11,725 cells). c, Hematoxylin and Eosin staining on IPF lung tissue sections. Representative image depicting

fibrotic (red square box) and non-fibrotic (blue square box) regions in IPF lung. Scale bar indicates 200 μm. d-g, Immunostaining for PATS-like markers in non-fibrotic regions of IPF lungs. d, Immunostaining for SFN (green), CLDN4 (red) and AGER (grey). e, Immunostaining for SFN (green), KRT17 (red) and TP63 (grey). f, Immunostaining for SFN (green), TP63 (blue), HTII-280 (grey) and ACTA2 (red) in healthy (left panel) and IPF lungs (right panel). g, Immunostaining for SFN (green), HTII-280 (red) and AGER (grey) in healthy (left panel) and IPF (right panel) lung. h, Immunostaining for ACTA2 (green), SFN (red) and COL1A1 (grey) in IPF lung. White line box in merged images indicate region of single channel images shown on right. DAPI (blue) stains nuclei. Scale bars in d-h indicate 100 μm. Images from c, d, e, f, g and h are representative from four humans repeated independently with similar results.
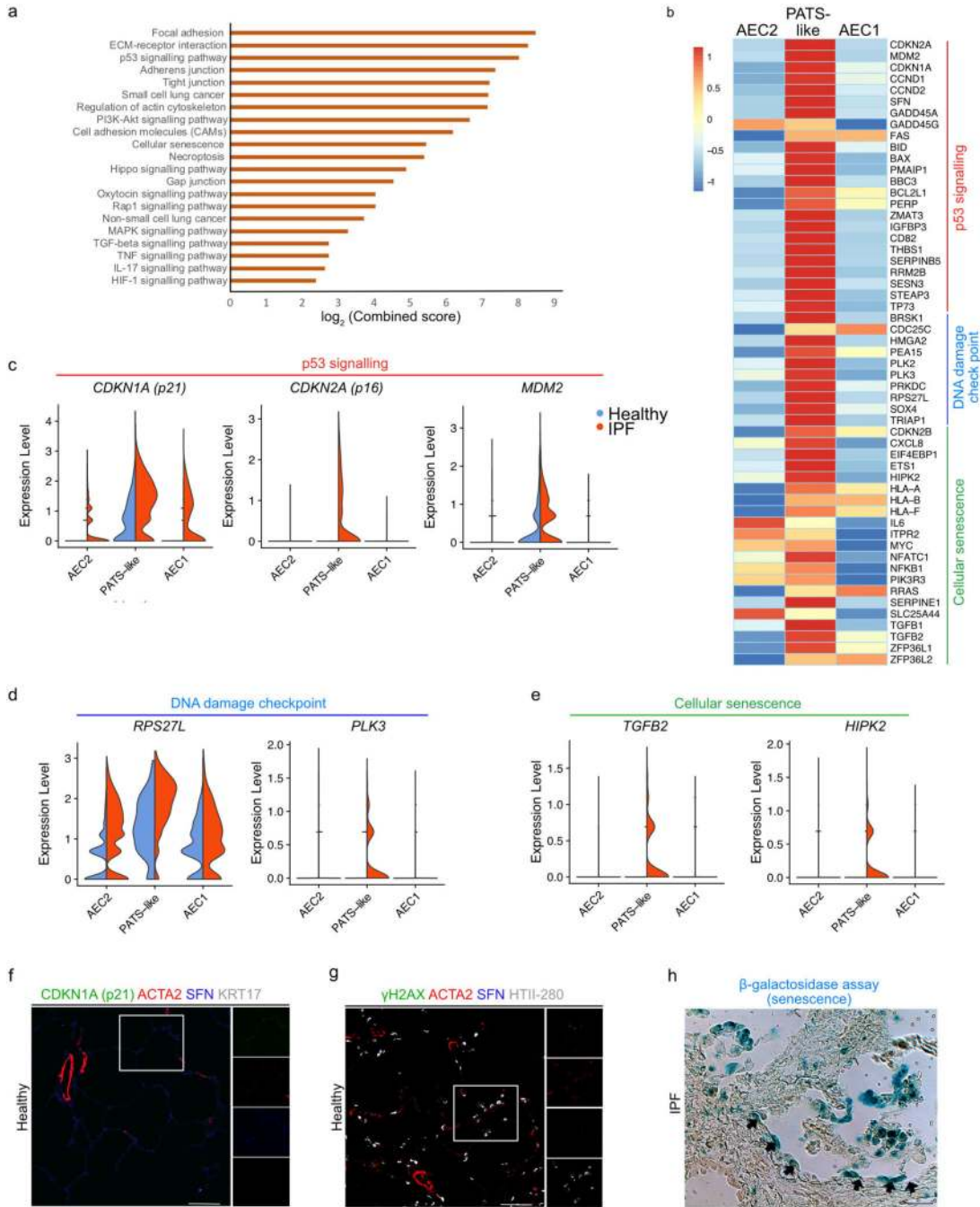
Extended Data 9. scRNA-seq analysis revealed enrichment of signalling pathways associated with PATS-like cells in human fibrotic lungs.

a, KEGG pathway enrichment analysis shows signalling pathways highly represented in PATS-like cells in human fibrotic lungs. Scale shows log$_2$ (combined score) obtained from Enrichr (see methods section for details). b, Heatmap shows expression of known target genes of indicated signalling pathways in AEC1, AEC2, and PATS-like state (11,725 cells). Scale indicates z-score where red is high, and blue is low. c-e Violin plots show the relative gene expression levels of indicated pathways/cellular processes: p53 signalling (c), DNA damage checkpoint (d), and cellular senescence (e) in different cell types in control and IPF

lungs (11,725 cells). Violin bodies indicate distribution of the cells in Healthy (blue) and IPF (red) lungs. f, Immunostaining for p21 (green), ACTA2 (red), SFN (blue) and KRT17 (grey). n=3 repeated independently with similar results. g, Immunostaining for γH2AX (green), SFN (blue), ACTA2 (red) and HTII-280 (grey) in healthy human lung. n=3 repeated independently with similar results. h, *β-galactosidase* staining on IPF lung section. Black arrows indicate X-gal staining in epithelial cells. n=4 repeated independently with similar results. White line box in merged images indicate region of single channel images shown on right. DAPI (blue) stains nuclei. Scale bars in b-i indicate 100 μm.
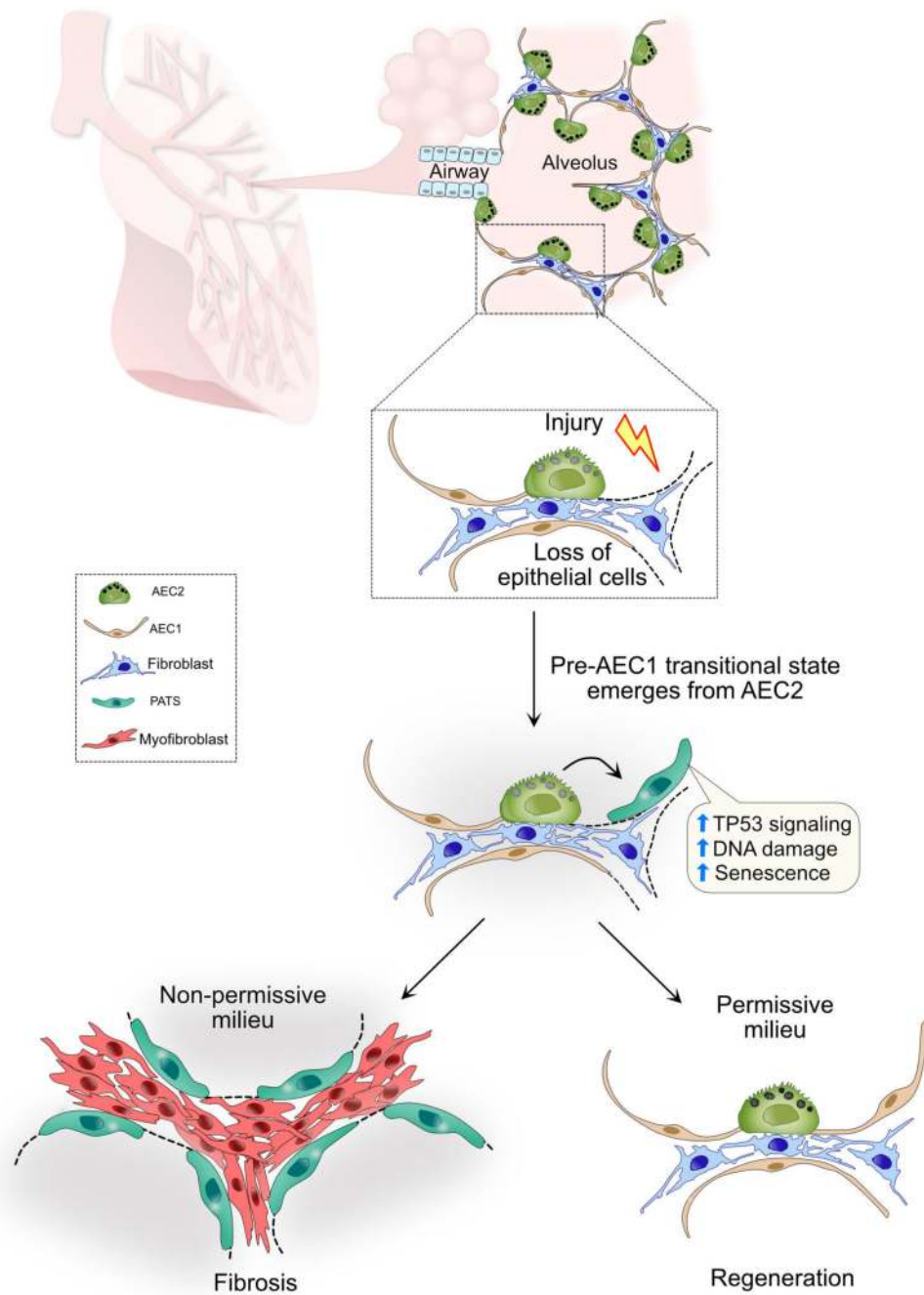
Extended Data 10. Schematic describing emergence of a transitional cell state in alveolar stem cell-mediated epithelial regeneration and its persistence in disease pathogenesis.
Alveolar stem cells replicate in response to damage and generate a transitional cell state which normally matures into functional alveolar type-1 epithelial cells. The newly identified transitional state is directly regulated by TP53 signalling, vulnerable to DNA damage and undergoes a transient senescent state. This transitional state is enriched in human fibrotic lungs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Hogan BLM 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). [PubMed: 25105578]

2. Basil MC et al. The Cellular and Physiological Basis for Lung Repair and Regeneration: Past, Present, and Future. Cell Stem Cell 26, 482–502 (2020). [PubMed: 32243808]

3. Nabhan A, Brownfield DG, Harbury PB, Krasnow MA & Desai TJ Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. Science (2018) doi:10.1126/science.aam6603.

4. Barkauskas CE et al. Type 2 alveolar cells are stem cells in adult lung. J. Clin. Invest 123, 3025–3036 (2013). [PubMed: 23921127]

5. Desai TJ, Brownfield DG & Krasnow MA Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature 507, 190–194 (2014). [PubMed: 24499815]

6. Zacharias WJ et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. Nature 555, 251–255 (2018). [PubMed: 29489752]

7. Zepp JA et al. Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. Cell 170, 1134–1148.e10 (2017). [PubMed: 28886382]

8. Tata PR & Rajagopal J Plasticity in the lung: making and breaking cell identity. Development 144, 755–766 (2017). [PubMed: 28246210]

9. Weibel ER Lung morphometry: the link between structure and function. Cell Tissue Res. 367, 413–426 (2017). [PubMed: 27981379]

10. Hogan B & Tata PR Cellular organization and biology of the respiratory system. Nat. Cell Biol (2019) doi:10.1038/s41556-019-0357-7.

11. Chung M-I, Bujnis M, Barkauskas CE, Kobayashi Y & Hogan BLM Niche-mediated BMP/SMAD signaling regulates lung alveolar stem cell proliferation and differentiation. Development 145, dev163014 (2018). [PubMed: 29752282]

12. LaCanna R et al. Yap/Taz regulate alveolar regeneration and resolution of lung inflammation. J. Clin. Invest 129, 2107–2122 (2019). [PubMed: 30985294]

13. Finn J et al. Dlk1-Mediated Temporal Regulation of Notch Signaling Is Required for Differentiation of Alveolar Type II to Type I Cells during Repair. Cell Rep 26, 2942–2954.e5 (2019). [PubMed: 30865885]

14. Riemondy KA et al. Single cell RNA sequencing identifies TGFβ as a key regenerative cue following LPS-induced lung injury. JCI Insight 5, (2019).

15. Hall-Glenn F et al. CCN2/connective tissue growth factor is essential for pericyte adhesion and endothelial basement membrane formation during angiogenesis. PLoS ONE 7, e30562 (2012). [PubMed: 22363445]

16. Strunz M et al. Longitudinal single cell transcriptomics reveals Krt8+ alveolar epithelial progenitors in lung regeneration. bioRxiv 705244 (2019) doi:10.1101/705244.

17. Rock JR et al. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. Proc. Natl. Acad. Sci. U.S.A 108, E1475–1483 (2011). [PubMed: 22123957]

18. La Manno G et al. RNA velocity of single cells. Nature 560, 494–498 (2018). [PubMed: 30089906]

19. Raab M et al. ESCRT III repairs nuclear envelope ruptures during cell migration to limit DNA damage and cell death. Science 352, 359–362 (2016). [PubMed: 27013426]

20. Li J et al. The Strength of Mechanical Forces Determines the Differentiation of Alveolar Epithelial Cells. Dev. Cell 44, 297–312.e5 (2018). [PubMed: 29408236]

21. Spike BT & Wahl GM p53, Stem Cells, and Reprogramming: Tumor Suppression beyond Guarding the Genome. Genes Cancer 2, 404–419 (2011). [PubMed: 21779509]

22. Soldatenkov VA et al. Regulation of the human poly(ADP-ribose) polymerase promoter by the ETS transcription factor. Oncogene 18, 3954–3962 (1999). [PubMed: 10435618]

23. Zhu Y et al. Early inactivation of p53 tumor suppressor gene cooperating with NF1 loss induces malignant astrocytoma. Cancer Cell 8, 119–130 (2005). [PubMed: 16098465]

24. Fan F et al. ATF3 induction following DNA damage is regulated by distinct signaling pathways and over-expression of ATF3 protein suppresses cells growth. Oncogene 21, 7488–7496 (2002). [PubMed: 12386811]

25. 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).

26. Nag S, Qin J, Srivenugopal KS, Wang M & Zhang R The MDM2-p53 pathway revisited. J Biomed Res 27, 254–271 (2013). [PubMed: 23885265]

27. Fernandez IE & Eickelberg O New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis. Lancet 380, 680–688 (2012). [PubMed: 22901889]

28. Gulati S & Thannickal VJ The Aging Lung and Idiopathic Pulmonary Fibrosis. Am. J. Med. Sci 357, 384–389 (2019). [PubMed: 31010465]

29. Habermann AC 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.

30. Adams TS 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.

31. Cheng D et al. Airway epithelium controls lung inflammation and injury through the NF-kappa B pathway. J. Immunol 178, 6504–6513 (2007). [PubMed: 17475880]

32. McConnell AM et al. p53 Regulates Progenitor Cell Quiescence and Differentiation in the Airway. Cell Reports 17, 2173–2182 (2016). [PubMed: 27880895]

33. 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). [PubMed: 30258935]

34. Araki K et al. p53 regulates cytoskeleton remodeling to suppress tumor progression. Cell. Mol. Life Sci 72, 4077–4094 (2015). [PubMed: 26206378]

35. Gadéa G, Lapasset L, Gauthier-Rouvière C & Roux P Regulation of Cdc42-mediated morphological effects: a novel function for p53. EMBO J. 21, 2373–2382 (2002). [PubMed: 12006490]

36. 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). [PubMed: 26247732]

37. Lipson KE, 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). [PubMed: 23259531]

38. Reyfman PA et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med 199, 1517–1536 (2019). [PubMed: 30554520]

39. Vaughan AE et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. Nature 517, 621–625 (2015). [PubMed: 25533958]

40. Kropski JA, Lawson WE, Young LR & Blackwell TS Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. Disease Models & Mechanisms 6, 9–17 (2013). [PubMed: 23268535]

41. da Silva ALG et al. Evaluation of DNA damage in COPD patients and its correlation with polymorphisms in repair genes. BMC Med. Genet 14, 93 (2013). [PubMed: 24053728]

42. Muñoz-Espín D & Serrano M Cellular senescence: from physiology to pathology. Nat. Rev. Mol. Cell Biol 15, 482–496 (2014). [PubMed: 24954210]

43. McGregor AL, 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). [PubMed: 26895141]

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

45. Mercado N, Ito K & Barnes PJ Accelerated ageing of the lung in COPD: new concepts. Thorax 70, 482–489 (2015). [PubMed: 25739910]

46. Gorgoulis V et al. Cellular Senescence: Defining a Path Forward. Cell 179, 813–827 (2019). [PubMed: 31675495]

47. Storer M et al. Senescence Is a Developmental Mechanism that Contributes to Embryonic Growth and Patterning. Cell 155, 1119–1130 (2013). [PubMed: 24238961]

48. Means AL, Xu Y, Zhao A, Ray KC & Gu G A CK19CreERT knockin mouse line allows for conditional DNA recombination in epithelial cells in multiple endodermal organs. genesis 46, 318–323 (2008). [PubMed: 18543299]

49. Arenkiel BR et al. Activity-induced remodeling of olfactory bulb microcircuits revealed by monosynaptic tracing. PLoS ONE 6, e29423 (2011). [PubMed: 22216277]

50. 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). [PubMed: 20023653]

51. Lo B, Hansen S, Evans K, Heath JK & Wright JR Alveolar Epithelial Type II Cells Induce T Cell Tolerance to Specific Antigen. The Journal of Immunology 180, 881–888 (2008). [PubMed: 18178827]

52. Chung M-I & Hogan BLM *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). [PubMed: 30011373]

53. Buch T et al. A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. Nat Methods 2, 419–426 (2005). [PubMed: 15908920]

54. 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). [PubMed: 25092767]

55. Marino S, Vooijs M, Gulden H van der, Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. Genes Dev. 14, 994–1004 (2000). [PubMed: 10783170]

56. Katsura H, Kobayashi Y, Tata PR & Hogan BLM IL-1 and TNFα Contribute to the Inflammatory Niche to Enhance Alveolar Regeneration. Stem Cell Reports 12, 657–666 (2019). [PubMed: 30930244]

57. Macosko EZ et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214 (2015). [PubMed: 26000488]

58. Stuart T et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21 (2019). [PubMed: 31178118]

59. Manno GL et al. RNA velocity of single cells. Nature 560, 494 (2018). [PubMed: 30089906]

60. Kuleshov MV et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44, W90–W97 (2016). [PubMed: 27141961]

61. Kanehisa M & Goto S KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28, 27–30 (2000). [PubMed: 10592173]

62. Carbon S et al. AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288–289 (2009). [PubMed: 19033274]

63. van Galen P et al. A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. Molecular Cell 61, 170–180 (2016). [PubMed: 26687680]

64. Girardot C, Scholtalbers J, Sauer S, Su S-Y & Furlong EEM Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. BMC Bioinformatics 17, 419 (2016). [PubMed: 27717304]

65. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]

66. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]

67. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010). [PubMed: 20513432]

68. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A & Mesirov JP Variant Review with the Integrative Genomics Viewer. Cancer Res 77, e31–e34 (2017). [PubMed: 29092934]

69. Ramírez F et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160–W165 (2016). [PubMed: 27079975]

70. Nagendran M, Riordan DP, Harbury PB & Desai TJ Automated cell type classification in intact tissues by single-cell molecular profiling. eLife https://elifesciences.org/articles/30510 (2018) doi:10.7554/eLife.30510.
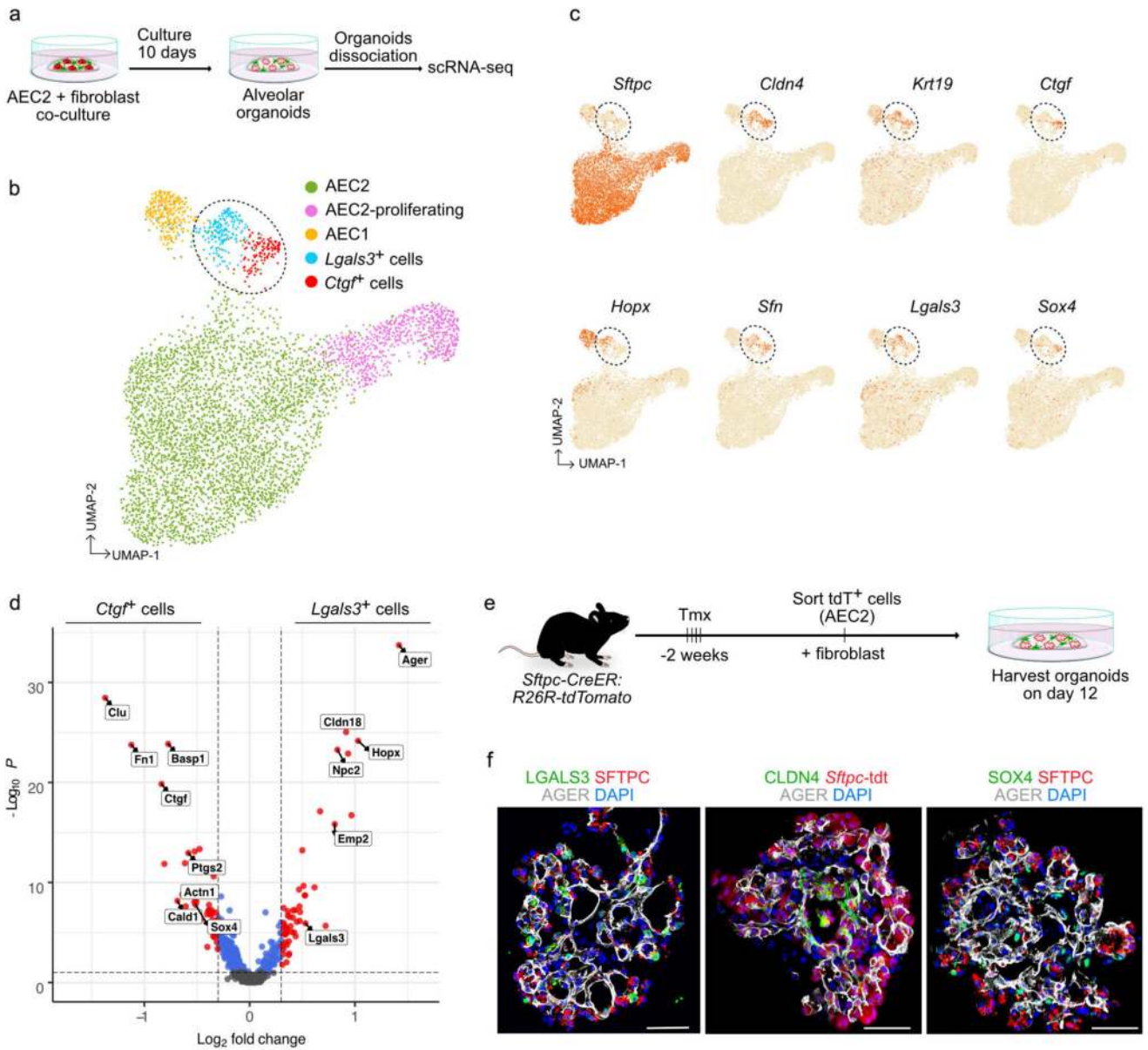
Figure 1. scRNA-seq reveals previously unknown alveolar epithelial cell states in *ex vivo* organoids.

a, Schematic of alveolar organoid culture utilized for single-cell RNA-seq. b, Uniform manifold approximation and projection (UMAP) visualization of epithelial populations in cultured alveolar organoids (4,573 cells). AEC2 (green, 3,303 cells) – alveolar epithelial type-2 cells, AEC2-proliferating (purple, 696 cells) – proliferating alveolar epithelial type-2 cells, AEC1 (yellow, 262 cells) – alveolar epithelial type-1 cells, new cell states : *Lgals3+* cells (blue, 184 cells) and *Ctgf+* cells (red, 128 cells) c, UMAP plots show the expression of indicated genes in epithelial populations in cultured alveolar organoids. Dotted circles in b and c indicate the transitional cell states. d, Volcano plot shows specific genes enriched in *Ctgf+* (n=128 cells) and *Lgals3+* (n=184 cells) transitional cell states. Wilcoxon rank sum test was used for the statistical analysis. e, Schematic of alveolar organoid culture using

fibroblasts and AEC2 cells. f, Immunostaining for PATS markers in alveolar organoids. Left panel - LGALS3 (green), SFTPC (red) and AGER (grey); middle panel - CLDN4 (green), *Sftpc*-tdt (red) and AGER (grey); right panel - SOX4 (green), SFTPC (red) and AGER (grey). Image is representative of 30 organoids from three mice. DAPI stains nuclei (blue). Scale bars indicate 20 μm.
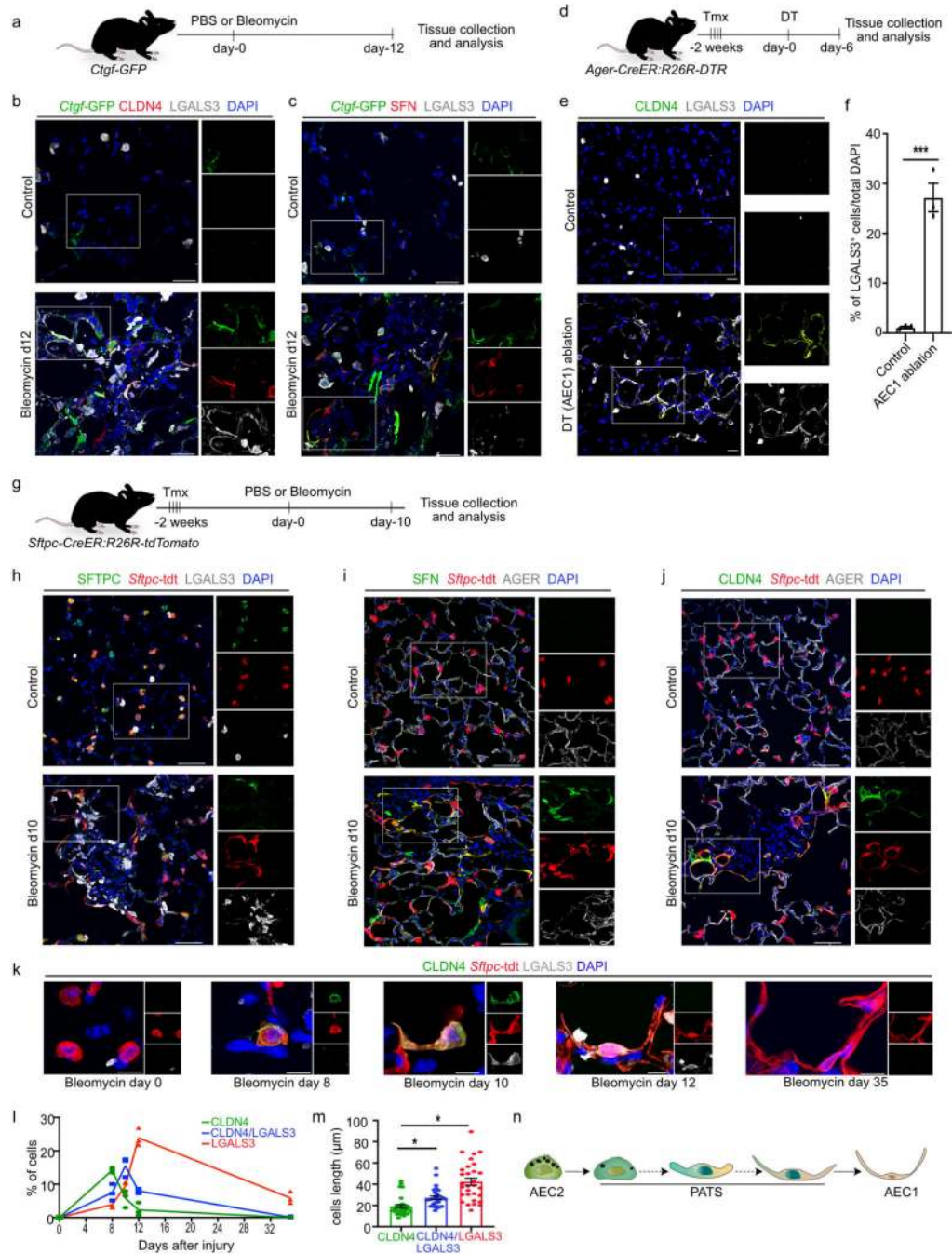
Figure 2. PATS emerge transiently and originate from alveolar stem cells after lung injury *in vivo*.

a, Schematic of bleomycin-induced lung injury in *Ctgf-GFP* mice. b-c, Immunostaining for PATS markers. b, *Ctgf*-GFP (green), CLDN4 (red) and LGALS3 (grey), c, *Ctgf*-GFP (green), SFN (red) and LGALS3 (grey) in control lung (upper panel) and bleomycin-treated lungs collected on day 12 after injury (lower panel). Scale bars indicate 30 μm. d, Experimental design to ablate AEC1s in *Ager-CreER;R26R-DTR* mouse model. Mice were administered tamoxifen (Tmx) followed by diphtheria toxin (DT) and tissues collected on day 6. e, Immunostaining for CLDN4 (green) and LGALS3 (grey) in control (upper panel)

and AEC1-ablated lungs (lower panel). Scale bars indicate 20 μm. f, Quantification of elongated LGALS3$^+$ cells in control and AEC1-ablated lungs. ***p<0.0008 (two-tailed, un-paired student's t-test). n=3 mice. g, Experimental workflow for sequential administration of tamoxifen followed by bleomycin injury and tissue collection for analysis using *Sftpc-CreER;R26R-tdTomato* mice. h-j, Immunostaining for PATS markers in *Sftpc*-lineage labeled cell in control (upper panel) and bleomycin injured lungs (lower panel). h, SFTPC (green), *Sftpc*-tdt (red) and LGALS3 (grey), i, SFN (green), *Sftpc*-tdt (red) and AGER (grey), and j, CLDN4 (green), *Sftpc*-tdt (red) and AGER (grey). DAPI stains nuclei (blue). Scale bars indicate 30 μm. White box in merged image indicates region of single channel images. k, Time course of immunostaining for CLDN4 (green), *Sftpc*-tdt (red) and LGALS3 after bleomycin injury. Scale bars indicate 10 μm. l, Quantification of CLDN4$^+$ cells (green), CLDN4$^+$/LGALS3$^+$ cells (blue) and LGALS3$^+$ cells (red) at different times after injury. n=3 mice. m, Quantification of cell length of CLDN4$^+$ cells (green), CLDN4$^+$/LGALS3$^+$ cells (blue) and LGALS3$^+$ cells (red). *p =0.049 for green vs blue; *p =0.0235 for green vs red (two-tailed un-paired student's t-test). n=30 cells from three mice. n, Schematic showing transition from AEC2 to AEC1 through different PATS subtypes. Data are from three independent experiments. In f, l, m, data are presented as mean ± s.e.m. Images from b, c, e, h, i, j, k are representative from three mice. h-k experiment was repeated three times independently with similar results.
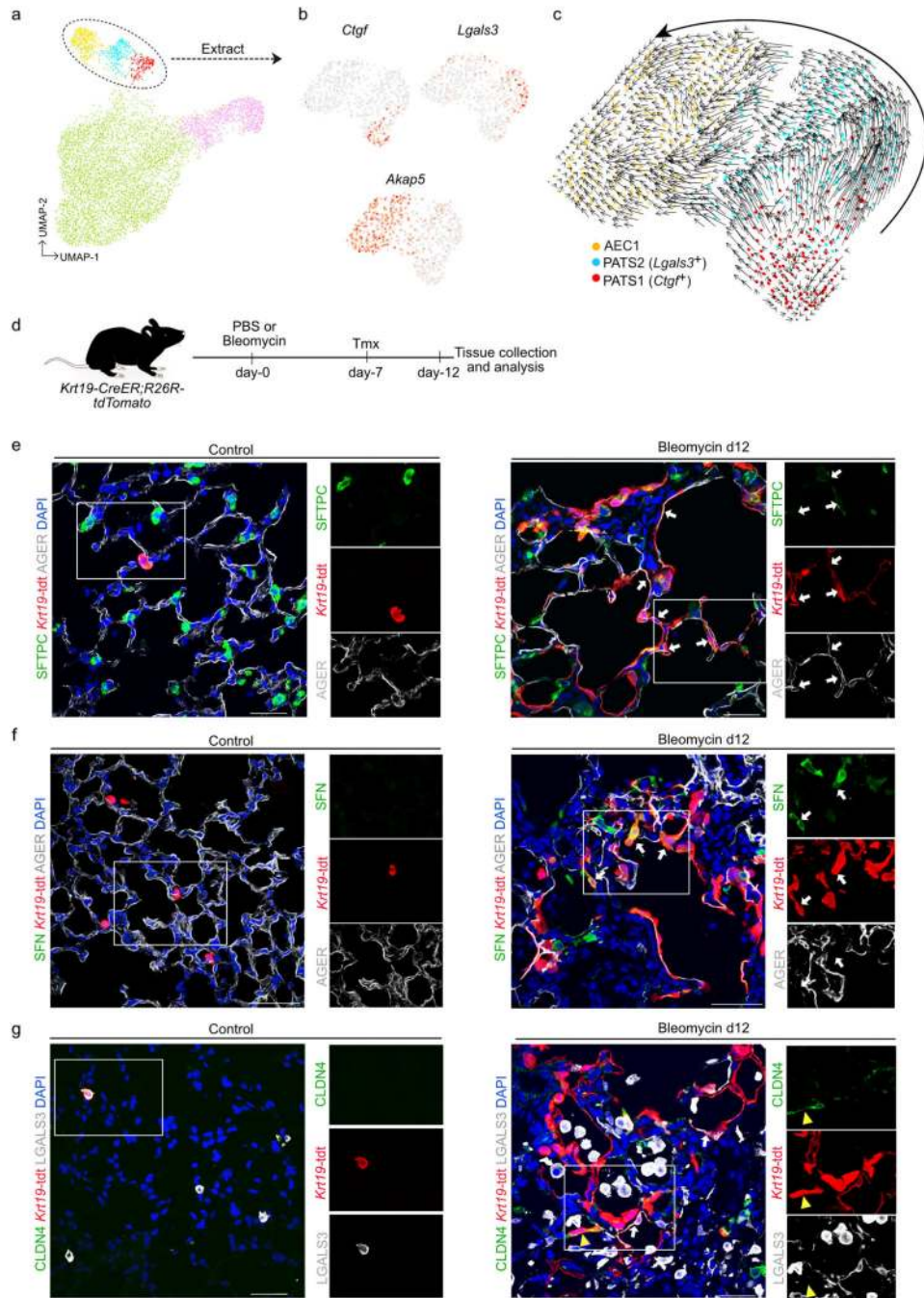
Figure 3. Lineage tracing revealed that PATS generate AEC1.
a, UMAP of epithelial populations in cultured alveolar organoids (4,573 cells). Arrow
indicates selected cell populations in the oval-shaped circle are shown in panel b. b, UMAP
plots show the expression of indicated genes in the selected populations (oval-shaped circle
in panel a, 574 cells). 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 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$^+$ cells. (Scale bar: 50 μm). g. Immunostaining for CLDN4 (green), *Krt19*-tdt (red) and LGALS3 (grey). White arrows indicate CLDN4$^+$*Krt19*-tdt$^+$ cells. Yellow arrowhead indicates LGALS3$^+$*Krt19*-tdt$^+$ cell. (Scale bar: 30 μm). e-g Control lungs are shown in left panels and bleomycin day 12 injured lungs are shown in right panels. DAPI stains nuclei (blue). White box in merged image indicates region of single channel images shown in left side. Images from e-g are representative from three mice repeated independently with similar results.
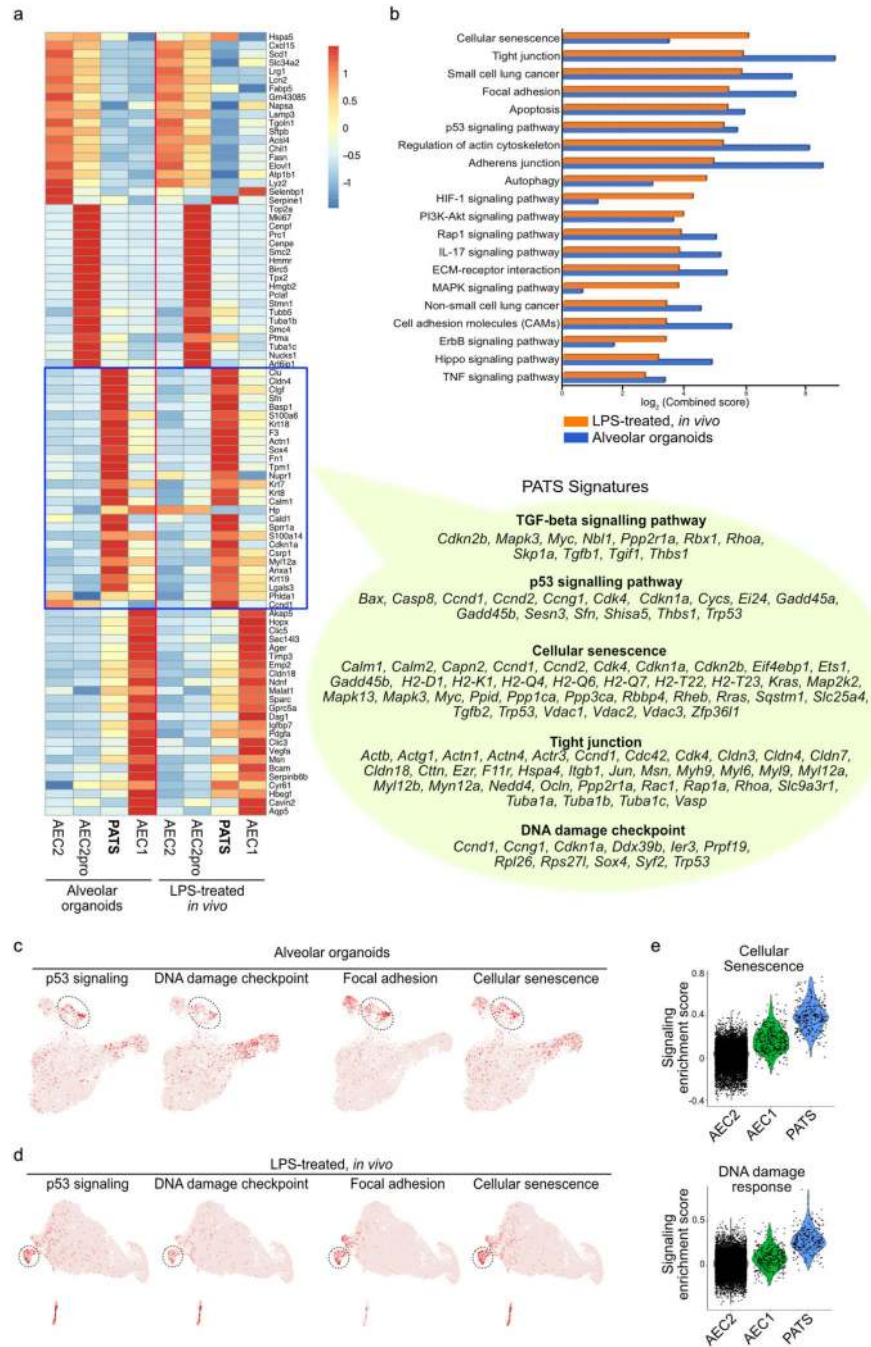
Figure 4. Gene expression signatures and signalling pathways enriched in PATS

a, Heatmap shows marker gene expression of each cell population in alveolar organoids (left, n=4,573 cells) and in LPS-treated murine lungs (right, n=13,204 cells) (scale shows z-score). Table on the right indicates genes enriched in the indicated pathways and cellular processes specifically in PATS. b, KEGG analysis reveals pathways enriched in PATS. Scale shows $\log_2$ (combined score) obtained from web-based tool - Enrichr. c and d, UMAP rendering of PATS enriched signalling pathways and cellular processes in alveolar organoids (c, 4,573 cells) and LPS-treated murine lungs (d, 13,204 cells). e, Violin plots for senescence

(upper) panel and DNA damage response pathways (lower panel) in indicated cell populations from LPS-injury model *in vivo* (n=13,039 cells). Black dots represent cells. Violin bodies indicate distribution of the cells.
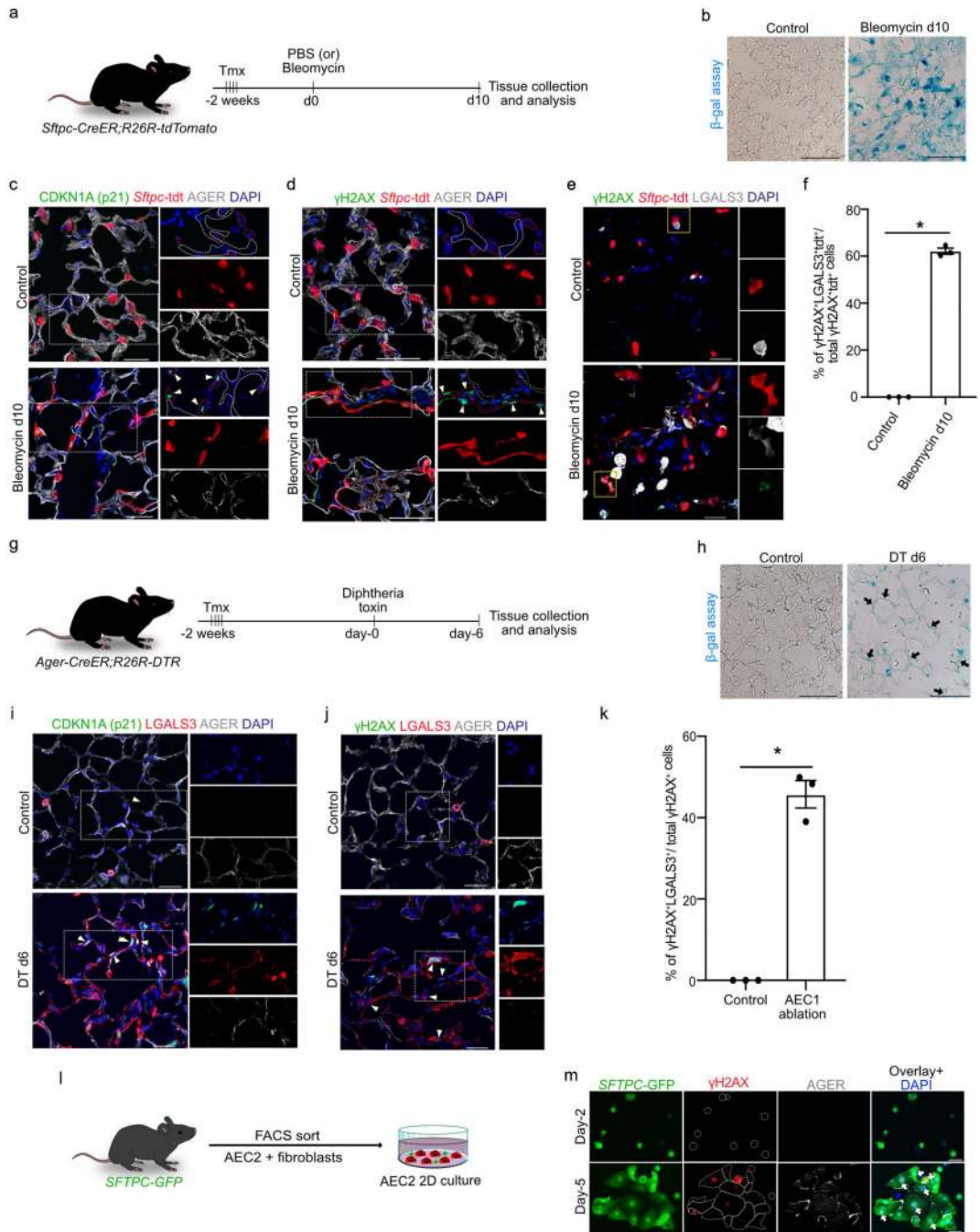
Figure 5. PATS undergo stretch-induced DNA damage and senescence *in vivo* and *ex vivo*
a, Schematic of bleomycin-induced lung injury in *Sftpc-CreER;R26RtdTomato* mice. b, *β-galactosidase* staining in bleomycin injured and control lungs. c, Immunostaining for CDKN1A (p21) (green), *Sftpc*-tdt (red) and AGER (grey). d, Immunostaining for γH2AX (green), *Sftpc*-tdt (red) and AGER (grey) in control and bleomycin injured lungs. Yellow arrowheads indicate γH2AX$^+$ cells. White dotted lines indicate AEC1 cells and red dotted lines outline tdt$^+$ cells. e, Immunostaining for γH2AX (green), *Sftpc*-tdt (red) and LGALS3 (grey) in control (upper) and bleomycin-treated lung (lower). f, Quantification of LGALS3$^+$

γH2AX$^+$Sftpc-tdt$^+$ in total γH2AX$^+$Sftpc-tdt$^+$ cells in control and bleomycin injured mice on day-10. Asterisks indicate *p<0.0318 (one-tailed, Mann-Whitney). n=3 mice. g, Schematic of AEC1 ablation using diphtheria toxin (DT). h, β-galactosidase staining in DT-treated and control lungs. i, Immunostaining for CDKN1A (p21) (green), LGALS3 (red) and AGER (grey). j, Immunostaining for γH2AX (green), LGALS3 (red) and AGER (white) in control (upper panel) and DT-treated lungs collected on day-6 after injury (lower panel). k, Quantification of LGALS3$^+$$\gamma$H2AX$^+$ cells in total γH2AX$^+$ cells in control and DT-treated lung. *p=0.0318. (one-tailed, Mann-Whitney). n=3 mice. l, Schematic of 2D culture of AEC2. m, Immunostaining for *SFTPC*-GFP (green), γH2AX (red) and AGER (grey) in 2D culture of AEC2. White arrows indicate cells with DNA damage marker. Insets on the right side in all panels show individual fluorescence channels of region indicated by dotted line boxes. DAPI stains nuclei (blue). Data are from three independent experiments and are presented as mean ± s.e.m. Images from b, c, d, e, h, i, j, m are representative from three mice repeated independently with similar results. Scale bar: 100μm in b and h; 50μm in c and d; 30μm in e and m; 20μm in i and j.
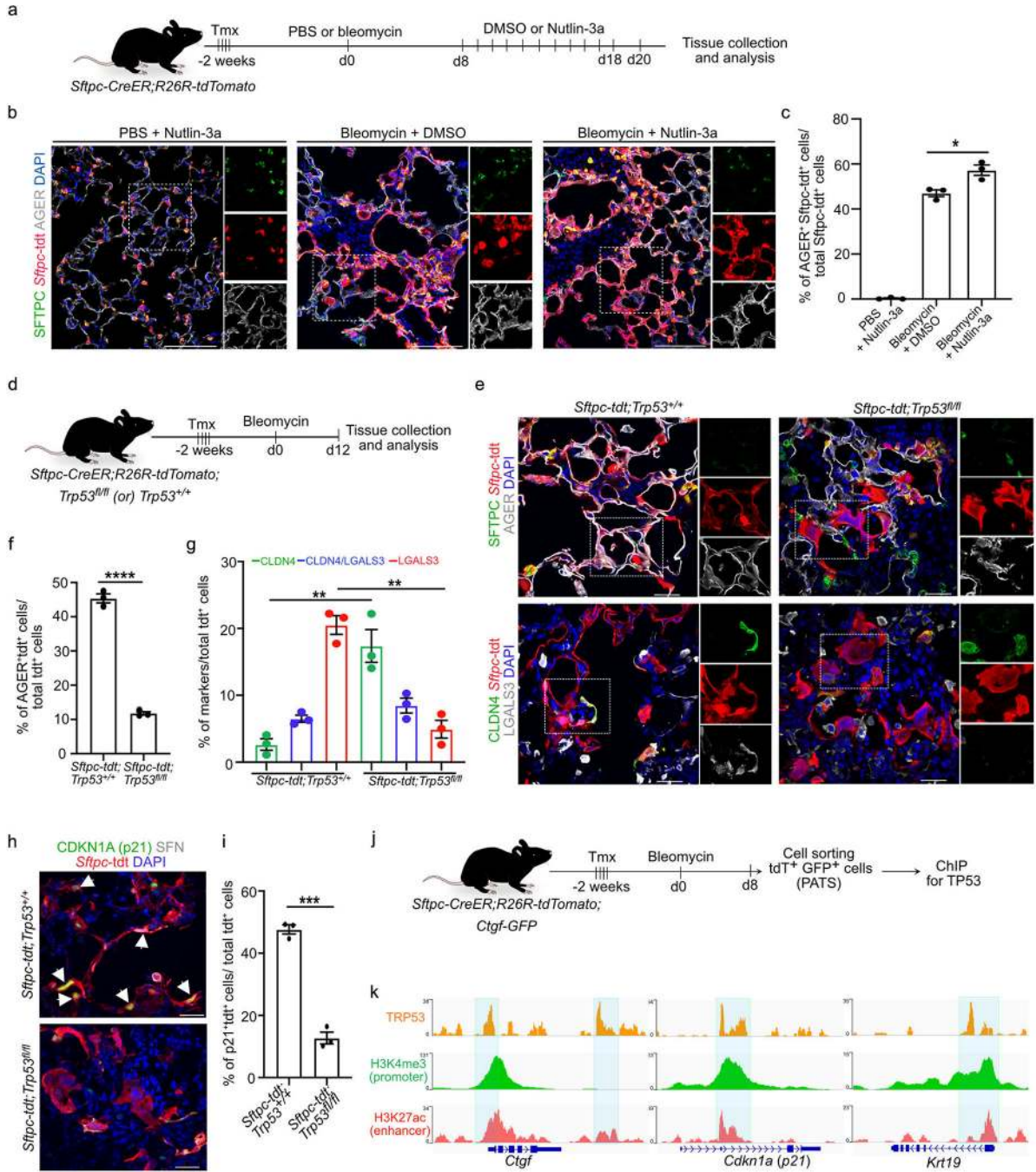
Figure 6. Genetic and pharmacological modulation and genomic binding assays reveal transcriptional control of PATS by TP53 signalling

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, Immunostaining for 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. Scale bar: 100 μm. c, Quantification of AGER$^+$*Sftpc*-tdt$^+$ cells in total *Sftpc*-tdt$^+$ Cells. *p = 0.0201; (two-tailed, unpaired t-test). n=3 mice. d. Experimental workflow for sequential

administration of tamoxifen to delete TP53 in AEC2 cells followed by bleomycin injury (d0) in *Sftpc-CreER;R26R-tdTomato;Trp53$^{fl/fl}$ or* control mice (*Sftpc-CreER;R26R-tdTomato;Trp53$^{+/+}$*). e, Immunostaining for SFTPC (green), *Sftpc*-tdt (red) and AGER (grey) (upper panel) and CLDN4 (green), *Sftpc*-tdt (red) and LGALS3 (grey) (lower panel) in control (left panel) and Trp53 knockout (right panel) mice. Scale bar: 30 μm. f, Quantification of AGER$^+$*Sftpc*-tdt$^+$ cells in total *Sftpc*-tdt$^+$. ****p<0.0001 (two-tailed, un-paired student's t-test). n=3 mice. g, Quantification of CLDN4$^+$ cells (green), CLDN4$^+$/LGALS3$^+$ cells (blue) and LGALS3$^+$ cells (red) in bleomycin-treated lungs. **p=0.048 (green bars) and **p=0.0013 (red bars) (two-tailed, un-paired student's t-test). n=3 mice. h, Immunostaining for CDKN1A (p21) (green), *Sftpc*-tdt (red) and SFN (grey) in control (upper panel) and Trp53 knockout mice treated with bleomycin (lower panel). White arrowheads indicate p21$^+$*Sftpc*-tdt$^+$ cells. Scale bar: 30 μm. i, Quantification of p21$^+$*Sftpc*-tdt$^+$ cells in total *Sftpc*-tdt$^+$ cells. ***p<0.0001 (two-tailed, un-paired student's t-test). n=3 mice. j, Schematic of bleomycin-induced lung injury in *Sftpc-CreER;R26R-tdTomato;Ctgf-GFP* mice followed by PATS sorting and ChIP analysis. k, ChIP enrichment for TRP53 (orange), H3K4me3 (active promoter; green) and H3K27ac (active enhancer; red) shown in IGV tracks in *Ctgf,* (left) *Cdkn1a* (middle) and *Krt19* (right). Blue-shaded regions indicate promoter/enhancer. DAPI stains nuclei (blue). Insets on the right side in all panels show individual fluorescence channels of region indicated by dotted line boxes. Data are from three independent experiments and are presented as mean ± s.e.m. Images from b, e, h are representative from three mice repeated independently with similar results.
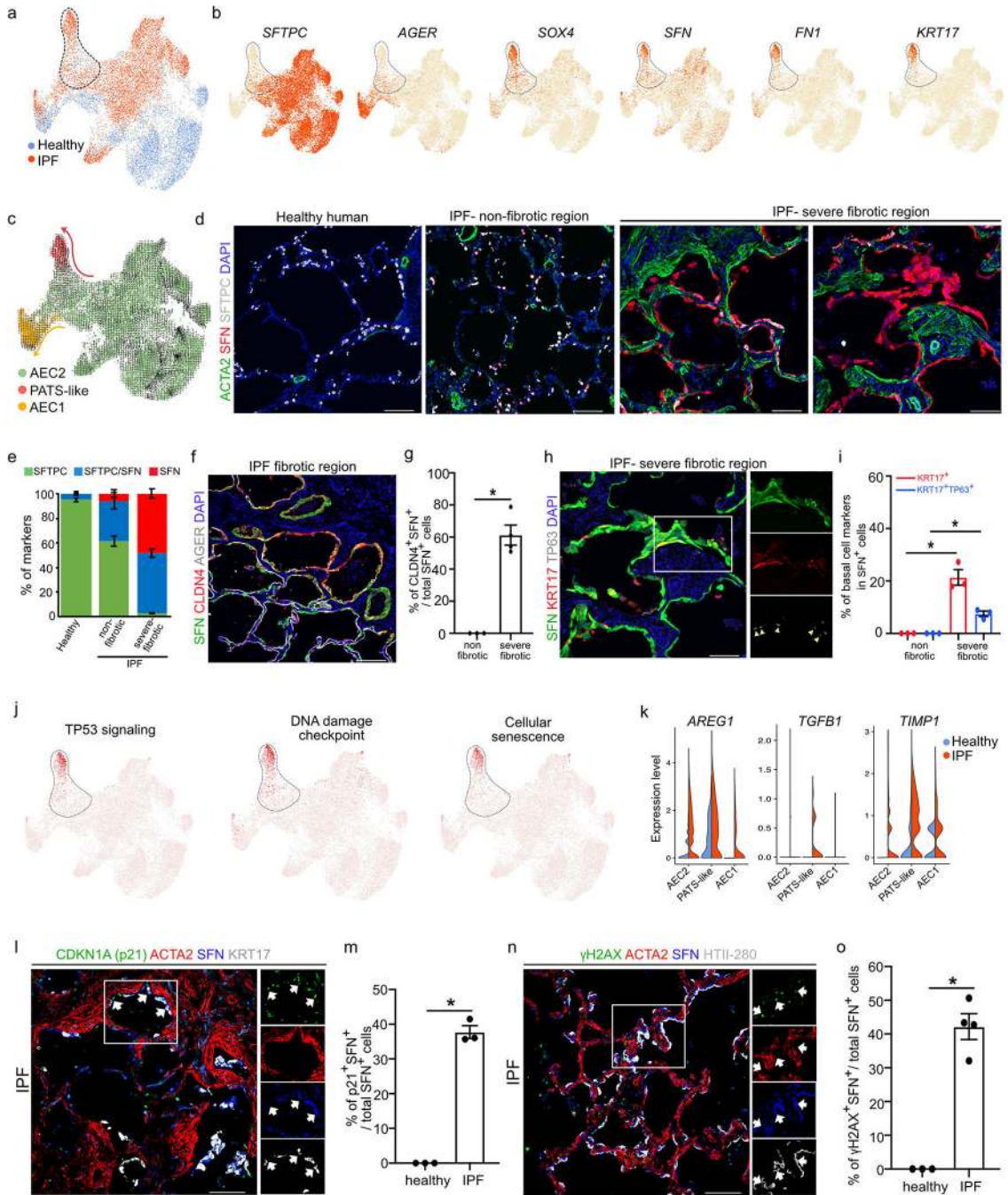
Figure 7. Enrichment of PATS-like states in IPF suggests persistence of this state in pathological milieu

a, UMAP shows scRNA-seq data from AECs in healthy (blue) and IPF (red) lungs (11,725 cells). b, UMAP plots indicate the expression of genes in healthy and IPF lung scRNA-seq data. c, RNA velocity analysis predicts lineage trajectories in AEC populations. Arrows indicates strong RNA velocities. d, Immunostaining for ACTA2 (green), SFN (red) and SFTPC (grey) in healthy, non-fibrotic and severe-fibrotic regions of IPF lungs. e, Quantification of SFTPC$^+$ (green), SFTPC$^+$SFN$^+$ (blue) and SFN$^+$(red) cells in healthy, and non-fibrotic/fibrotic regions in IPF lungs. n=3 human samples. f, Immunostaining for SFN

(green), CLDN4 (red) and AGER (grey) in IPF lung. g, Quantification of CLDN4$^+$ cells in SFN$^+$ cells in non-fibrotic and severe fibrotic regions in IPF lungs. *p=0.0218 (one-tailed, Mann-Whitney). n=3 human samples. h, Immunostaining for SFN (green), KRT17 (red) and TP63 (grey) in IPF lung. i, Quantification of KRT17$^+$ cells (red) and KRT17$^+$TP63$^+$ cells (blue) in total SFN$^+$ cells in non-fibrotic regions compared to severe fibrotic regions in IPF lungs. *p=0.0318 (one-tailed, Mann-Whitney). n=3 human samples. j, UMAP plots from scRNA-seq data show enrichment of candidate signalling pathways in healthy and IPF lungs (11,725 cells). k, Violin plots show IPF-relevant gene expression in indicated cell types/ states in control and IPF lungs (11,725 cells). Violin plots indicate cells distribution in healthy (blue) and IPF (red) lungs. l, Immunostaining for CDKN1A (p21) (green), ACTA2 (red), SFN (blue) and KRT17 (grey) in IPF lung. Arrows indicate SFN$^+$KRT17$^+$CDKN1A$^+$ cells. m, Quantification of p21$^+$SFN$^+$ in total SFN$^+$ cells. *p=0.0318 (one-tailed, Mann-Whitney). n=3 human samples. n, Immunostaining for γH2AX (green), SFN (blue), HTII-280 (grey) and ACTA2 (red) in IPF lungs. o, Quantification of γH2AX$^+$SFN$^+$ in total SFN$^+$ cells. *p=0.0218 (one-tailed, Mann-Whitney one). healthy human samples n=3, IPF n=4. Inset indicates single channel images shown in right. Data are from at least three lungs and presented as mean±s.e.m. All scale bars indicate 100μm. Images from d,f,h,l,n are representative of three human samples repeated independently with similar results.